The Journal of

Economic Perspectives

A journal of the
American Economic Association

Fall 2025

The Journal of
Economic Perspectives

A journal of the American Economic Association

The Journal of
Economic Perspectives

Contents Volume 39 • Number 4 • Fall 2025

Symposia

Articles

Features

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.

Japan's Debt Puzzle: Sovereign Wealth Fund from Borrowed Money

Yili Chien, Wenxin Du, and Hanno Lustig

J apan presents a striking puzzle in public finance. Government debt exceeds 200 percent of GDP, budget deficits have persisted for decades, and economic growth has been sluggish. Yet inflation has remained subdued, and no major debt crisis has emerged. Understanding how Japan has managed to defy the standard logic of debt sustainability is the starting point for our analysis.

The key lies in the Japanese public sector's operation of a de facto sovereign wealth fund. Unlike countries such as Norway and Saudi Arabia, which fund such vehicles with national savings from natural resources, Japan finances its investments largely through domestic borrowing at very low floating interest rates. While the risk premia on these investments have generated strong returns over the past two decades and supported debt sustainability, this strategy exposes the government to considerable interest rate and exchange rate risks, which are especially salient in an environment of rising rates. In this paper, we explore these risks and discuss lessons from the Japanese experience with a high debt-to-GDP ratio for other advanced economies, including the United States.

Advanced economies around the world are experiencing a demographic transition and a secular slowdown in long-term economic growth rates. Japan is at the leading edge of this transition (for details, see World Bank Group 1960–2024). As

■ *Yili Chien is Economic Policy Adviser, Federal Reserve Bank of St. Louis, St. Louis, Missouri. Wenxin Du is Sylvan C. Coleman Professor of Financial Management, Harvard Business School, Boston, Massachusetts. Hanno Lustig is Mizuho Financial Group Professor of Finance, Stanford Graduate School of Business, Stanford, California. Their email addresses are yili.chien@stls.frb.org, wdu@hbs.edu, and hlustig@stanford.edu.*

the growth rate of the Japanese economy has slowed and its population has aged more quickly than expected, the Japanese government has run large budget deficits over the past three decades. These deficits are primarily driven by large transfer payments, especially to older Japanese households. The debt-to-GDP ratio in Japan has grown to more than 200 percent in 2023, which is the highest among advanced economies (OECD 2022).

This essay begins with an overview of how Japan accumulated so much debt. We then turn to a consolidated balance sheet of the entire Japanese government sector, including the annual government budget, various government funds including the public pension funds, and the central bank, the Bank of Japan. We show that Japan's government runs, in effect, a sovereign wealth fund based on borrowed money, using a substantial share of its extremely high levels of debt to fund the purchase of risky financial assets. Japan's government keeps its costs of borrowing very low, by methods that include the extraordinary monetary easing policies of the Bank of Japan and favorable regulatory treatment for Japan's financial institutions that hold government debt. Yields on Japanese government bonds have remained among the lowest in the world for decades. The Japanese government has used its cheap domestic funding to take highly-levered long positions in risky assets, including foreign equities and bonds. This strategy involves absorbing a number of risks, including risks of movements in interest rates, exchange rates, and equity prices. But over the past decade, the government has earned an additional 6 percent of GDP per annum above its funding costs by investing in risky assets. Even though Japan's low-interest-rate policies and the risk premia earned from risky investments help debt sustainability, the Japan's government cannot make the risks disappear, and so these risks ultimately have to be borne by its taxpayers, bondholders, and bank depositors.

The United States faces growing fiscal pressures similar to those Japan encountered decades ago, with rising government debt and persistent deficits driven by budgetary reactions to events like the global financial crisis of 2008–2009 and the COVID pandemic, as well as longer-term trends of an aging population and increasing social security and healthcare costs. Could the United States run a sovereign wealth fund and address its government debt burden with borrowed money? We will explain why the United States is unlikely to replicate Japan's model.

How Did Japan Accumulate So Much Public Debt?

As of 2023, Japan's debt-to-GDP ratio surpassed 200 percent, the highest among advanced economies. This substantial public debt stems from decades of persistent fiscal deficits and prolonged economic stagnation, primarily driven by demographic aging. Moreover, Japan's accumulation of debt reflects intricate dynamics between fiscal authorities, the central bank, the social security fund, and other public institutions. This section outlines the historical context and key policy decisions that have shaped Japan's rising debt burden.

Aging Population and Growing Social Security Deficits

Japan's demographic transition started earlier than in other countries. Its fertility rate fell below the replacement level of 2.1 children per woman in the mid-1970s. Over the following decades, Japan's fertility rate continued to decline steadily, while its life expectancy increased. By 2001, the proportion of the population aged 65 and older had reached 18.4 percent, a level comparable to that of the United States in 2024 (based on World Bank Group 2025). Over the past two decades, the aging of Japan's population has accelerated. As of 2024, 30 percent of Japan's population is 65 or older, and projections indicate this share will rise to 36 percent within two decades. Japan's total population peaked in 2010 and has since begun to decline. The combination of an aging society and a shrinking workforce has significantly contributed to Japan's prolonged economic stagnation (Fernández-Villaverde, Ventura, and Yao 2025).

In Japan, "social security" refers to a suite of programs that includes public pensions for the elderly, but also health insurance, services for the elderly and disabled, unemployment insurance, maternity leave and day care, and safety net programs for those with low incomes. The funding for these programs comes from a mixture of earmarked taxes on wages, as well as general revenues and government borrowing. As Japan's population continues to age and economic growth remains subdued, the annual social security deficit has expanded sharply since 2000, rising from 4.1 percent of GDP in 1998 to a peak of 9.4 percent in 2011 before moderating slightly to 7.6 percent in 2023. In response, the Japanese government introduced social security reforms in the early 2010s. These reforms sought to control medical expenditures, to extend social security coverage (especially for nonregular workers),and to boost labor force participation among women and the elderly. While these measures helped slow the growth of the deficit, it remains substantial, hovering around 8 percent of GDP.

In comparison, since 1998, Japan's general government has run an average primary deficit of 5.1 percent of GDP. Notably, this figure is smaller than the social security deficit, implying that, excluding social security expenditures, the Japanese government would have maintained a fiscal surplus. Despite fiscal consolidation efforts, including two consumption tax hikes in 2014 and 2019, the primary deficit remains significant and unable to restore fiscal balance.

Evolution of Public Debt and Growth of the Social Security Fund

Japan's persistent primary fiscal deficits have been financed by a rise in government debt, which has increased by 125 percent of GDP since 1997. The evolution of general government debt since 1997 closely mirrors the cumulative primary fiscal deficits, which amount to 133 percent of GDP. The debt is financed in part by bond issuance, but central and local governments have historically relied on loans and intergovernmental financing. Before 2000, the Japanese government relied on the Fiscal Investment and Loan Fund (FILF), a large state-run lending program, as a major source of funding (Doi and Hoshi 2002). In turn, FILF was funded automatically by deposits from the social security fund and Japan's "postal savings bank."

Most Japanese households traditionally did not participate in financial markets directly, but instead saved in deposits. One of the largest deposit-taking institutions was run by Japan's Postal Service. FILF then used this stable low-cost funding for public investment projects and to provide targeted loans. By the late 1990s, its loan portfolio had surpassed 100 percent of GDP, in part because Japan's postal savings bank had become the world's largest deposit-taking institution. However, as Japan moved to privatize the postal savings bank in the early 2000s, FILF's access to this low-cost funding source declined. Although the program remained sizable, its operations shrank significantly, with its asset position falling from 107 percent of GDP in 1997 to 51 percent in 2012. After the privatization of the postal savings bank in 2007, the Japanese government increasingly turned to bond markets where it had to pay higher interest rates, further driving up public debt.

To finance large social security deficits, the Japanese government chose to issue bonds rather than drawing down social security reserves. Thus, despite persistent and substantial social security deficits, the public pension fund continued to grow over time, with its net asset position rising from 36 percent of GDP in 1997 to approximately 60 percent in 2024. Had the government used these reserves to offset deficits instead of issuing bonds, the level of outstanding public debt would be significantly lower.

This policy decision appears to be linked to shifts in the portfolio strategy of the social security fund. In 1997, over two-thirds of its assets (equivalent to 27 percent of Japan's GDP) were invested in Japan's Fiscal Investment and Loan Fund. However, as FILF transitioned to bond market financing, the social security fund adjusted its holdings, reducing direct lending to FILF from 27 percent of GDP in 1997 to 4.3 percent in 2012 while increasing its holdings of Japanese government bonds from 2.4 percent to 16 percent of GDP. In effect, although the social security fund decreased direct financing to FILF, it continued to provide capital indirectly through Japanese government bond markets.

Since 2012, the social security fund has increased its exposure to riskier assets. In 2013, a government panel recommended a major reallocation of government-run pension funds into higher-yield investments (Hoshi and Yasuda 2015). Following this shift, three-quarters of public pension fund assets were allocated equally across three categories of risky assets: domestic equities, foreign equities, and foreign bonds (each at 25 percent). Over the past decade, these investments have generated strong realized returns, boosting the net asset position of public pension funds from around 40 percent of GDP to over 60 percent of GDP.

Unconventional Monetary Policy of the Bank of Japan

Japan has been stuck in a low-growth regime since the early 1990s. After some years of attempting to stimulate the economy with large budget deficits and low interest rates, in April 1999 the Bank of Japan made the bold commitment to holding short-term interest rates at zero until deflationary concerns were dispelled. In 2001, the Bank of Japan took an additional step and started large-scale asset purchases to stimulate the economy—and also coined the phrase "quantitative

easing" to describe this policy. Its main stated objective was to increase bank reserves, with the assumption that banks would deploy these reserves in lending to firms and households. Just as Japan completed the liberalization of its capital markets, the Bank of Japan intensified its large-scale asset purchases as part of the ambitious government spending program since 2012 referred to as "Abenomics," after the name of then–Prime Minister Abe Shinzo. In effect, the Japanese government replaced the cheap funding it had obtained through the postal savings bank before capital-market liberalization with bank reserves held at the Bank of Japan.

In 2016, the Bank of Japan shifted to a policy of explicit "yield curve control," which refers to managing the yield paid on bonds of different maturities. In this case, the bank announced a willingness to purchase bonds as needed to impose a target for the ten-year yield of zero percent. The hope was that by announcing the target, bond markets would gravitate to this target yield with only modest intervention needed. However, this scenario did not materialize. Over the past decade, the Bank of Japan has purchased more long-dated Japanese government bonds each year than the total annual issuance. By the end of 2023, the Bank of Japan owned ¥581 trillion (nearly $4 trillion in US dollars) in long-dated Japanese government bonds, more than half of outstanding Japanese bonds. In this way, Japan's government removed all interest rate ceilings as part of capital market liberalization, but then it effectively imposed a new cap on interest rates through yield curve control. This large expansion of the Bank of Japan balance sheet blurs the line between the central bank and public financial institutions, making it essential to view these entities as a whole.

A Sovereign Wealth Fund with Borrowed Money

Japan's public sector, taken as a whole, runs a risky sovereign wealth fund funded with money borrowed at very low rates through various channels. To get a complete picture of Japan's fiscal position, we need to take several steps: (1) go beyond central and local governments to include public pension funds, public financial institutions, and the central bank; (2) conduct a comprehensive assessment of all liabilities including other borrowing instruments, such as government loans; and (3) account for the public sector's assets and their portfolio composition. The public sector of Japan holds large positions in risky assets, and their returns fluctuate based on investment strategies. In the case of Japan, evaluating fiscal positions solely through the lens of gross government debt can present a highly distorted picture of the overall fiscal health of the country.

Consolidated Public Sector Balance Sheet

To develop a better understanding of Japan's public finances, we analyze the consolidated balance sheet of the Japanese public sector, following the work by Chien, Cole, and Lustig (2023). The consolidated government includes the Bank of Japan, public financial institutions, and the general government, which consists

Table 1

Consolidated Japanese Government Balance Sheet

Percentage of GDP, quarter end	1997Q4	2012Q4	2024Q2	Δ 1997–2024
Assets				
Deposits	5.9%	8.5%	19.1%	13.2%
Loans	102.8%	63.1%	63.4%	−39.4%
Of which by PFIs	92.2%	50.1%	40.0%	−52.2%
Domestic equities	10.7%	20.9%	41.7%	31.1%
Of which by PPFs	2.4%	4.4%	13.5%	11.1%
Of which by BoJ	0.0%	0.7%	12.6%	12.6%
Foreign securities	7.5%	29.7%	61.8%	54.3%
Of which by PPFs	1.3%	7.0%	27.7%	26.5%
Other assets	8.4%	7.9%	6.4%	−1.9%
Sum (assets)	135.2%	130.1%	192.4%	57.2%
Liabilities				
Currency	10.8%	18.3%	20.3%	9.6%
Bank reserves	0.6%	9.5%	90.8%	90.2%
Bonds & T-bills	41.8%	162.3%	103.8%	62.0%
Of which by BoJ	−9.6%	−23.2%	−93.6%	−84.0%
Loans	55.1%	48.5%	37.8%	−17.4%
Deposits FILF	46.4%	1.1%	1.5%	−44.9%
BoJ external debt	0.0%	0.1%	7.1%	7.1%
Other liabilities	5.2%	8.7%	8.8%	3.6%
Sum (liabilities)	159.9%	248.5%	270.0%	110.1%
Net liabilities	24.7%	118.4%	77.6%	52.9%

Source: Japan Flow of Funds and National Accounts of Japan.
Note: Balance Sheet in Market Values as percentage of GDP. The consolidated Japanese government includes central and local government, public pension funds (PPFs), Bank of Japan (BoJ), and public financial institutions (PFIs). The net liabilities—defined as the sum of liabilities minus the sum of assets—should equal the present discounted value (PDV) of primary surpluses, which represents the implied market value of claims on future primary surpluses.

of the central government, local governments, and the public pension funds. Table 1 provides a snapshot of the consolidated balance sheets in the last quarter of 1997 and 2012, as well as in the second quarter of 2024. The first and last dates were chosen to represent the start and end of our dataset. The end of 2012 marks the beginning of the significant policy shift implemented under Abenomics. All numbers are expressed as a percentage of GDP.

The total size of the public sector's balance sheet has expanded significantly over the past 27 years. In 1997, financial assets were 135 percent of GDP and total liabilities at 160 percent of GDP. By 2024, financial assets and total liabilities rose to 192 percent and 270 percent of GDP, respectively. More than half of these financial assets were invested in risky securities. These risky assets include a long position worth 62 percent of Japan's GDP in foreign securities and another one worth 42 percent of GDP invested in domestic equities. To fund these risky investments, Japan's government issued short-term bills and longer-term bonds (104 percent

of GDP) as well as drawing on bank reserves (91 percent of GDP). Between 2012 and 2024, the Japanese government reduced its outstanding debt in the bond market from 162 percent of GDP to 104 percent of GDP and instead issued bank reserves, thus mostly borrowing from banks at the policy-determined rate instead of borrowing from bond market investors at market-determined yields.

The risky investments in the public sector are mainly undertaken by the Japanese public pension funds, starting in 2012. Between 2012 and 2024, the domestic equity position of these funds increased from 4.4 percent of GDP to 13.5 percent of GDP and their foreign investment rose from 7 percent of GDP to 27.7 percent of GDP. During the same period, the Bank of Japan also rapidly increased its risky asset position from less than 1 percent of GDP in 2012 to 12.6 percent of GDP. These risky long positions (including domestic equities and foreign securities) are shown in the consolidated government balance sheet in Table 1, which grew from 18 percent of GDP to 51 percent of GDP from 1997 to 2012 and then rapidly increased to 104 percent of GDP in 2024. The government's equity position, equivalent to 42 percent of GDP, is equal to about two-thirds of the total equity holdings of all Japanese households combined (as will be discussed later in the paper and shown in Table 3). Since 2012, Japan's government has increased its exposure to equities by 100 percent, and it has increased its exposure to foreign risky assets by more than double.

The composition of liabilities of these funds has changed significantly over the past three decades. In 1997, around 46 percent of GDP in funding came from deposits with the postal savings bank through Japan's Fiscal Investment and Loan Fund. In 2001 Japan abolished the requirement that the postal bank provide funding to FILF, and as this source of FILF funding gradually dried up, the Japanese government issued more government bonds and other fixed income securities instead. From 1997 to 2012, the funding from postal savings deposits through FILF declines from 46 percent of GDP to only 1 percent of GDP, while the outstanding stock of Japanese government bonds rose from 42 percent to 162 percent of GDP.

The unconventional monetary policy conducted by the Bank of Japan plays an important role in the evolution of the public sector balance sheet. As a result of quantitative easing and yield curve control, the Bank of Japan has purchased a significant portion of the government's longer-maturity debt, primarily funded by the creation of reserves held at the Bank of Japan. After 2012, the government fully replaced the Fiscal Investment and Loan Fund deposits with bank reserves held at the Bank of Japan as a source of cheap domestic funding. In the second quarter of 2024, the Bank of Japan held long-term government bonds equivalent to 93 percent of GDP (recorded as −93.3 percent of GDP in Table 1 on the consolidated public sector liability side). This position offset government debt in the consolidated balance sheet, reducing the debt held by the public to just 103.8 percent of GDP. Without holdings by the Bank of Japan, outstanding government bonds and bills would approach 200 percent of GDP. Meanwhile, the central bank bond purchases were largely matched by an expansion of bank reserves, which stood at about 91 percent of GDP in 2024. In other words, through central bank bond purchases,

government securities amounting to 91 percent of GDP were converted into overnight bank reserves on the consolidated public sector balance sheet.

On the public sector balance sheet expressed in market values, total liabilities minus the financial assets equal the net liabilities, which have to be backed by the present discounted value of future primary budget surpluses (provided that a "no-bubble condition" holds). The net liabilities of the public sector are reported in the last row of Table 1. As a result of the growth of its financial assets, the net liabilities have only grown from 24.7 percent in 1997 to 52.9 percent in 2024. In fact, Japan's net liabilities decreased over the past ten years from 118.4 percent to 77.6 percent. However, these gains have come from taking risks—in particular, the risks of duration mismatch and currency mismatch.

Duration Mismatch

The current composition of the Japanese public sector holds risky long-duration assets, such as equities, equivalent to 102.5 percent of GDP, while a large fraction of its liabilities consisted of short-duration instruments. More than 40 percent of its liabilities have zero duration; that is, 91 percent of GDP was held in reserves at the Bank of Japan, and another 20 percent of GDP in cash. Government short-term bills and longer-term bonds accounted for only 104 percent of GDP. As a result, the Japanese government has engineered a sizable duration mismatch on its consolidated balance sheet. It borrows primarily at floating rates while investing in longer-duration, risky assets to harvest risk premia, resulting in a duration mismatch.

According to the workhorse expectations hypothesis of the term structure of interest rates, borrowing short-term to invest in long-term bonds when the yield curve is upward-sloping will not yield higher returns, because the slope of the yield curve signals higher short-term rates ahead, making the expected return on long-term bonds equal to the expected cost from rolling over short-term borrowing. But the expectations hypothesis fails in the data so that the "naive strategy" of chasing high yields along the yield curve works, but it is risky (for example, Campbell 1995). Investors who buy longer-dated bonds when the yield curve slope upwards harvest the term premium, which compensates them for the interest rate risk they are taking on.

To understand the duration risk exposure of financial assets and liabilities for Japan's government, we estimate the duration of each type of financial asset in the government's balance sheet.

Currency, deposits, and bank reserves have zero or near-zero duration. For all loans on the government's balance sheet, we assume a duration of three years. The duration of equities is perhaps a less familiar idea, but just as the duration of a bond is determined by the time it takes for the present value of the bond to be fully repaid, the duration of an equity is the time it takes for the present value of the equity to be repaid—which in turn will depend on the dividends paid by the stock. A shorter-duration equity will have higher near-term dividends; a longer-duration equity is expected to pay back its value later in the future. The Gordon (1959) growth model is the standard formula for inferring the duration of an equity from

its expected flow of dividends. For the price-to-dividend ratio of Japanese stocks, we use the Jordà-Schularick-Taylor Macrohistory Database (Jordà et al. 2019). Using the Gordon formula, the average duration for Japanese stocks over this period is 75.6 years, a high value reflecting the elevated valuation ratios of Japanese stocks during this time. For government bonds, we use the effective duration from one of the indexes calculated by the Intercontinental Exchange and Bank of America, the ICE-BofA Japan Government index, which averages 7.19 years over our sample period. The duration of FILF deposits is calibrated to one year.[1]

Foreign securities primarily consist of two components: (1) foreign reserves; and (2) securities held by social security funds. For social security foreign securities, we calculate a weighted average duration based on US stocks (59 years) and bonds (7.19 years). Given the 50/50 portfolio allocation of Japan's social security fund, the resulting duration is 33 years. The portfolio composition of Japan's foreign reserves, mostly held in the form of US Treasury debt, is not publicly available. We approximate its duration using the average maturity of foreign official holdings of US long-term debt, which is approximately 8.5 years. The Treasury International Capital (TIC) data indicate that most of Japan's holdings of US Treasury debt are long-term. For instance, in January 2023, Japan held approximately $1.1 trillion in US Treasury securities, but only $72 billion in T-bills (as shown at Board of Governors of the Federal Reserve System 2023)

Putting these estimates together, and using the portfolio reported in Table 1, we can then calculate the duration of the consolidated balance sheet. In the second quarter of 2024, the duration of its risky asset position is around 24.5 years. The high duration is mostly due to its equity position. The duration of its liabilities is quite short, only 3.2 years. The short duration of its liabilities is a direct result of the quantitative easing and yield curve control policies adopted by the Bank of Japan.

The "net duration" is the weighted average of asset and liability duration. The consolidated government has a net financial liability position of 77.6 percent of GDP in 2024, the net result of a debt position of 270 percent of GDP, and an asset position of 192.4 percent of GDP. As a result, in the perspective of net liability position, the smaller asset position contributes more negative duration than the positive duration contributed by its larger debt. This means that the duration of the government's net liability position is large and negative. In 2024, the public sector's asset duration is 49 years longer than its liability duration, leaving the liability duration equal to −49 years. As a result, an increase in Japan's interest rates by 100 basis points reduces Japan's asset value more than its liability value, increasing its net debt by 49 percent. Given the 2024 net liabilities position at 77.6 percent of GDP (Table 1), a 49 percent rise in net debt is about 38 percent of GDP. This is mainly due to the long position in high-duration risky assets.

---

[1]Our duration calculation excludes other assets and liabilities for two reasons. First, their durations are difficult to estimate without detailed portfolio information. Second, their shares are relatively small compared to the overall balance sheet.

Any primary budget surpluses for Japan are far in the distant future (for more detailed analysis, see Chien, Cole, and Lustig 2023). Hence, an increase in Japan's interest rates increases the market value of its net debt, because the value of its financial assets declines by more than the value of its debt, but greatly reduces the present discounted value of surpluses. As a result, Japan would come under heavy pressure to cut spending and/or to increase taxes. In short, interest rate increases destroy fiscal capacity in Japan. To avoid this duration mismatch, Japan would have had to align the duration of its net debt with the duration of its future surpluses by issuing lots of long-term debt, thus locking in low long yields.

Currency Mismatch

In addition to duration risk, Japan's public sector is heavily exposed to currency risk. The Japanese public sector is engaged in a global currency "carry trade," which refers to trades in which an investor borrows in low-interest-rate currencies and invests in high-interest-rate currencies without hedging the risk that exchange rates will shift. According to the theory of uncovered interest rate parity, this naive strategy of chasing high interest rates will not yield high returns, because these currencies will depreciate, but in the data, this naive strategy does work. The empirical evidence shows that low-interest-rate currencies do not appreciate sufficiently against high-interest rate currencies to offset their interest rate differentials, making carry trades consistently profitable on average (Lustig, Roussanov, and Verdelhan 2011; Hassan and Mano 2019). Currency carry trade investors harvest the currency risk premium, which compensates them for the currency risk they are taking on. In bad times for global investors, high interest rate currencies depreciate against low interest rate currencies.

Japan's government borrows at low short-term rates in Japanese yen, a typical funding currency for carry trades, and invests 62 percent of GDP in risky foreign assets without hedging currency risk.[2] This currency carry trade has been associated with large gains for Japan's public sector balance sheet over the past three decades, as the Japanese yen has depreciated by 47 percent in real terms since 1997 (Bank of International Settlements 1997–2025).

As shown in Table 1, the public sector's net exposure to foreign securities was equal to 62 percent of Japan's GDP in 2024. Of this, an amount equal to 27.7 percent of GDP was held in foreign securities by public pension funds, primarily through the Government Pension Investment Fund (GPIF). Notably, the GPIF maintains a near-zero hedge ratio on its dollar-denominated assets, as documented in Du and Huber (2024), leaving public sector finances highly vulnerable to exchange rate fluctuations.

Decades of ultra-low interest rates have cemented the Japanese yen as a canonical funding currency in global carry trade. The funding currency refers to the low-interest-rate currencies in which an investor borrows to implement a currency carry trade. This strategy is highly sensitive to shifts in global risk appetite and

---

[2]This currency exposure compounds the duration risk, as the Japanese yen tends to depreciate in real terms when Japanese long-term real interest rates decline below foreign rates (Lustig, Stathopoulos, and Verdelhan 2019).

monetary policy expectations. The sheer scale of Japan's public sector holdings makes it the world's largest carry trade investor, borrowing in yen-denominated liabilities and investing in higher-yielding foreign securities, predominantly denominated in US dollars. Indeed, the Japanese government's total foreign investments of $2.5 trillion dwarf existing estimates of the total yen-funded global carry trade by private investors. For example, Aquilina et al. (2024) provide a rough estimate (acknowledging that significant data gaps remain) of approximately $250 billion in yen carry trades by private investors.

While carry trades have been profitable on average, and the yen's substantial depreciation over the past decade has benefited Japan's public sector balance sheet, the currency position of the Japanese public sector entails considerable risk. Assuming the public sector's foreign currency exposure is entirely unhedged, a 10 percent appreciation of the yen would result in a valuation loss equivalent to 6.2 percent of GDP on its foreign currency portfolio.

The carry trade is not a free lunch. Rising interest rates in Japan would reduce the attractiveness of borrowing in yen, potentially triggering a major unwinding of global carry trades and a sharp yen appreciation. A reversal in carry trades can trigger a feedback loop of heightened global market volatility and reduced risk appetite, which further accelerates deleveraging and the unwinding of carry positions. The market turbulence in the summer of 2024 underscores this dynamic. Following the Bank of Japan's unexpected interest rate hike on July 31, 2024, the yen appreciated sharply, while the Japanese stock market plummeted by 12 percent in a single trading session. These developments had broader effects on the global financial markets. Global equity markets and digital assets suffered significant declines, and the Chicago Board Options Exchange Volatility Index (VIX), which measures market uncertainty by looking at option prices to infer the range of expectations about 30-day movements in the S&P stock market index, surged from 16 percent to 38 percent. Historical evidence shows a strong negative correlation between carry trade returns and global market volatility (Lustig, Roussanov, and Verdelhan 2011).

In summary, the consolidated balance sheet of the Japanese government assumes substantial interest rate risk and currency risk. In return, it collects a combination of the term premium, currency carry premium, and equity premium on these investments. The Japanese public sector effectively operates a large sovereign wealth fund financed with domestically borrowed money. Large increases in Japan's interest rates and/or yen appreciation would cause substantial mark-to-market losses for the public sector, including the Bank of Japan, and could potentially induce a run on government-issued yen-denominated debt.

Sustainability of Public Debt?

The reasons behind the sustainability of Japan's high public debt, at least up to the present, include a combination of borrowing at low interest rates and harvesting premia on risky assets. Here, we consider whether these dynamics are sustainable

in the longer run. Additionally, we explore some pricing implications for Japanese government bonds and the associated risks for bondholders.

Risk Premium and Debt Sustainability

A nation's debt-to-GDP ratio records the cumulative impact of past macroeconomic events and government decisions, much like a glacier records past weather. In the spirit of Hall and Sargent (2011), the key factors are the past record of primary deficits relative to GDP (where "primary" means that interest spending on past debt is not included), the interest payments on past debt, and the nominal growth of GDP.[3]

When the real rate of return that the government needs to pay on its debt is lower than the real growth rate of the economy, the government can sustain steady-state deficits while maintaining a constant debt-to-GDP ratio. The size of the possible steady-state deficit will be determined by the gap between the government's cost of debt and the growth of nominal GDP. However, in the sample period from 1997 to 2024, Japan's real GDP grew at an average rate of −0.03 percent, while the average real return on the Japanese government bond portfolio was 1.30 percent. Thus, Japan's central government was not operating in the magical region where it could just roll over steady-state deficits at an interest rate that is below the growth rate of the economy (Blanchard 2019; Mehrotra and Sergeyev 2021).

However, if we start from the consolidated budget constraint for the entire public sector, which takes into account the government's holdings of financial assets and the returns on those assets, the net debt-to-GDP ratio for the public sector has some additional determinants: the accumulation of financial assets over time and the gross return on the portfolio of assets. In this setting, even if the real growth rate of GDP is lower than the real rate of return that the government needs to pay on its debt, the consolidated government can run steady-state deficits with a constant debt-to-GDP ratio, provided that both the real returns on its asset position exceed the government's funding cost and the asset position itself is sufficiently large.

Over the past decade, Japan's public sector has realized an average annual return net of the funding costs equal to 4.6 percent (Chien, Cole, and Lustig 2023). That is equivalent to a gross return of 6 percent of GDP per annum between 2013 and 2023. Because the public sector is harvesting substantial excess return, Japan's net liabilities increased by only about 53 percent of GDP (as shown in Table 1) since 1998, even though the cumulative primary deficits surpassed 133 percent of GDP. To evaluate the impact of the risk-taking strategy, we conduct a simple counterfactual exercise in which Japan does not engage in risky investments. Specifically, instead of maintaining large equity and foreign asset investment positions, Japan is assumed to allocate these holdings exclusively to six-month Japanese Treasury bills. The results show that, under this counterfactual, the rise in net liabilities would have been similar to the data between 1998 and 2012, but would then have increased

---

[3]For an algebraic treatment of the conditions discussed in this section, see Supplemental Appendix A.

rapidly to more than 180 percent of GDP in 2024, compared with the actual level of 81 percent.[4] These results clearly reflect the shift toward a riskier portfolio during Abenomics.

Looking ahead, Chien, Cole, and Lustig (2023) project that the consolidated Japanese government could earn an expected return around 3.9 percent of GDP on its current risky portfolio. This nearly 4 percent of GDP, harvested in the form of risk premia, may help to explain why the expansion of Japan's government balance sheet has not led to a large inflation spike.

Where Did the Risk Go?

The Japanese approach to its public debt resembles a "gamble for resurrection," akin to the strategy employed by certain California municipalities facing underfunded pension liabilities (Myers 2022), or certain financial institutions that are insolvent but not yet illiquid. The strategy is to take substantial risks with the belief that if the risks pay off with higher returns, then the financial catastrophe can be averted; on the other side, if the risks do not pay off, the financial position was already negative, and it only becomes more so.

Can governments earn large excess returns by simply running a sovereign wealth fund with money borrowed at a low interest rate? No, not if the debt of that government is priced correctly. Governments cannot simply make risk disappear. The additional systematic risk on the asset side in Table 1 cannot be engineered away (Jiang et al. 2020), and will be priced into the cost of funding on the liability side. Ultimately, the risk on the asset either has to be borne by the bondholders of the Japanese government or by its current and future taxpayers and/or recipients of government spending.

The essential long-run condition here is that the present value of all public sector liabilities must equal to the future budget surpluses and the value of the financial assets. According to Table 1, in 2024 the present value of future Japan's total liabilities is 270 percent of GDP while the value of its financial assets 192 percent of GDP in 2024. This condition implies that the present value of future Japanese budget surpluses will be equal to 78 percent of GDP. Because this budget constraint is satisfied period by period, it implies that the gain or loss from its risky investment has to be reflected by the value of its debt or the value of future surplus claims. In other words, the risk associated with holding risky assets must be borne by either the taxpayer through the path of future surpluses or by the bondholders through changes in the valuation of debt.

Another way to see this point follows directly from the Modigliani and Miller (1958) theorem, which implies that the risk of the government's total assets, including its future surpluses, must equal the risk of its liabilities, irrespective of the

---

[4]For details of this calculation, see Supplemental Appendix B .

"capital structure" of the public sector balance sheet. As the consolidated government loads up on interest rate, currency, or equity risk on the asset side, the riskiness of its liabilities will increase proportionately. The only alternative would be if Japan's government offset the increased riskiness of its financial assets by reducing the riskiness of its nonfinancial asset, the claim to future surpluses—effectively shifting risk onto taxpayers.

In a "good state" for Japan's fiscal situation, the returns to risky financial assets remain high. But what happens in a bad state? On the one hand, if the risk is fully borne by Japanese taxpayers, then they will face a large future tax bill in bad states of the world. That seems implausible, given that surpluses in advanced economies are strongly pro-cyclical. There is no historical precedent that would rationalize expecting higher fiscal surpluses in the bad state of the world.[5] On the other hand, if the risk of a bad outcome is borne by bondholders and other creditors, a scenario that seems more likely, then they face large real losses in the future. If so, this should be reflected in a large risk premium on bonds, if the bonds are priced correctly. But Japan's government bonds are not in fact reflecting that risk premium, as we discuss in the next section.

Sustaining Low Interest Rates

The ability of Japan's public sector to sustain high levels of debt depends on stable and low funding costs from bondholders. However, governments can employ various strategies to enable borrowing at below-market interest rates. In this scenario, Japan's public sector liabilities are overpriced and do not reflect all of the risks. Reinhart, Kirkegaard, and Sbrancia (2011) refer to these measures as "financial repression." Possibilities include "macroprudential regulations" that require banks and financial institutions to hold government bonds, direct lending to the government by domestic pension funds and banks, and moral suasion to encourage domestic banks to increase their holdings of government bonds.[6]

The Japanese government has implemented a range of financial repression policies to maintain low funding costs. As discussed earlier, Japan's postal bank and public pension funds were mandated to invest heavily in Japan's Fiscal Investment and Loan Fund program at below-market interest rates up to 2011. As financial liberalization eroded this cheap funding source, the consolidated Japanese public sector turned to large-scale asset purchases, replacing bank reserves held at the Bank of Japan with low-cost deposits from the postal savings bank and pension

---

[5]In a neoclassical model with complete markets, the allocation of risk may be irrelevant for a country's aggregate welfare, as taxpayers and bondholders perfectly share risks. In the Ricardian class of models, a single representative agent effectively serves as both the bondholder and the taxpayer, with the government investing on their behalf. However, in reality where financial market participation is low (for example, many Japanese households do not hold bonds) and taxation is distortionary, the distribution of risk matters for the overall welfare.

[6]For a recent example of financial repression during the Eurozone sovereign debt crisis in 2011–2012, see Acharya and Steffen (2015), De Marco and Macchiavelli (2016), and Ongena, Popov, and Van Horen (2019).

reserves. Because the return on these reserves is set by monetary policymakers, it does not fully reflect the risk exposure on the asset side of the balance sheet.

When looking at the consolidated fiscal position of the central government with the Bank of Japan and pension funds, approximately 90 percent of GDP in consolidated debt takes the form of reserves, compared to only 80 percent of GDP in bonds (excluding short-term bills). The interest rate on bank reserves at the Bank of Japan has averaged just ten basis points between 1997 and 2024. Without these exceptionally accommodative policies, it is unclear whether the Japanese government could sustain borrowing above 200 percent of GDP at such low interest rates (Hansen and İmrohoroğlu 2023; Koeda and Kimura 2021). The public sector also benefits from a strong domestic investor base for Japanese government bonds. Panel A of Figure 1 shows that foreign ownership of Japanese government bonds remains extremely low, with nonresidents holding less than 10 percent of outstanding bonds. Historically, domestic banks were the largest holders of Japanese government bonds. However, over the past decade, the Bank of Japan crowded out private investors as a result of its quantitative easing and yield curve control program. By the end of 2023, the Bank of Japan held ¥581 trillion in Japanese government bonds (excluding short-term bills), accounting for more than half of all outstanding bonds. Since the inception of the Bank of Japan quantitative easing program, banks have shifted away from Japanese government bonds, increasing their reserve holdings at the Bank of Japan instead. The combined holdings of the Bank of Japan and domestic banks have steadily grown, surpassing 66 percent in 2024. Among nonbank domestic investors, insurance companies, private pensions, and investment trusts remain the largest holders, collectively accounting for approximately 27 percent of the market.

Japanese government bonds receive favorable regulatory treatment for regulated financial institutions such as banks, pensions, and insurance companies. Given that the majority of Japanese government bonds are held by the Bank of Japan and regulated domestic financial institutions, price discovery in the Japanese bond market is likely to be significantly impaired. This distortion helps explain why interest rate and currency risks might not be fully reflected in pricing of Japanese government bonds. A growing body of evidence suggests that liquidity and price discovery in Japanese government bond markets have deteriorated over the past decade, particularly as the Bank of Japan transitioned from large-scale asset purchases to full yield curve control (for example, Financial Markets Department, Bank of Japan 2024). Over the past ten years, daily transaction volumes in the cash market for on-the-run (recently issued) bonds have declined, bid-ask spreads for inter-dealer transactions have widened, and most liquidity metrics of Japanese government bonds have worsened.

Lastly, despite the absence of strict capital controls and persistently low nominal interest rates in Japan, domestic financial institutions, including banks, insurance companies, and pensions, face significant costs when scaling up their overseas investments. Because these institutions primarily have yen-denominated liabilities, investing abroad requires them to either maintain substantial currency risk exposure or to hedge this risk using foreign exchange derivatives. In the post-crisis

Figure 1
Public Debt Holder Comparison

Panel A. Japan
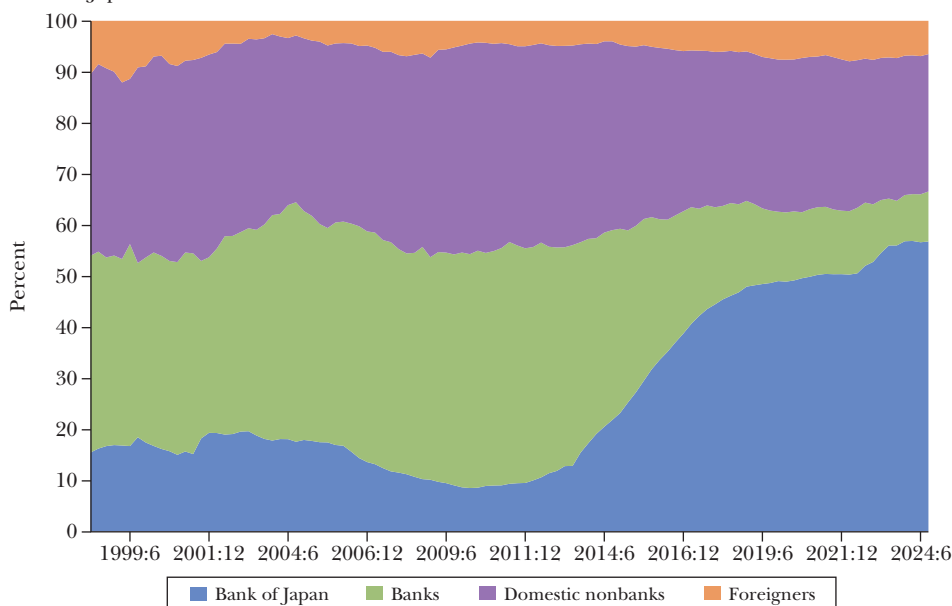
Panel B. United States
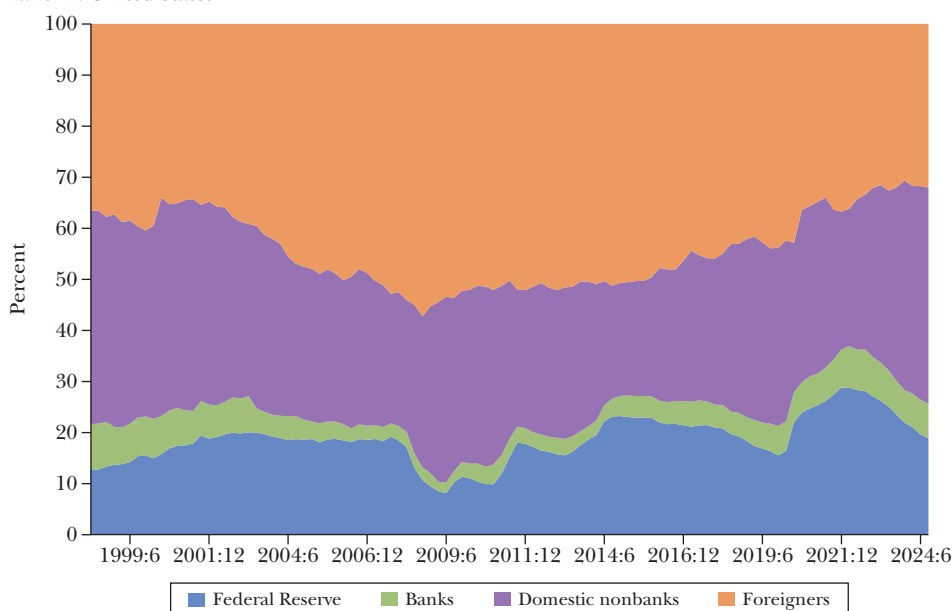
Source: Japan Flow of Funds, Federal Reserve Economic Data, and US Treasury Bulletin.
Note: The public debt considered here excludes intragovernmental holdings. These intragovernmental holders include state and local governments, public financial institutions, and public pension funds. Japan's public debt figure includes debt issued by public financial institutions, such as FILF, but excludes Treasury bills. Domestic nonbank holders include all other domestic investors, such as insurance companies, mutual funds, and households.

regulatory environment, constrained financial intermediaries have faced increasing pressure to sell US dollars and buy back yen in forward and swap markets for foreign exchange, leading to persistent deviations (Du, Tepper, and Verdelhan 2018; Du and Schreger 2022). This sustained demand for yen from hedging raises the cost of hedging for Japanese investors, effectively acting as a tax on private-sector institutions seeking to mitigate foreign currency risk. As a result, the rising cost of hedging discourages domestic investors from allocating capital abroad, reinforcing a strong domestic clientele for yen-denominated assets.

Implications for the United States

Over the past two decades, the US federal government debt-to-GDP ratio has doubled, rising from approximately 60 percent of GDP in 2004 to about 120 percent in 2024. According to projections from the Congressional Budget Office in 2024, the US government is expected to run persistently large primary deficits exceeding 2 percent of GDP per annum over the next three decades that can be traced to an aging population and rising healthcare costs, together with ever-rising interest expenses on past debt. Under this trajectory, the US government debt-to-GDP ratio is poised to continue rising.

The fiscal challenges now faced by the United States bear striking similarities to those Japan encountered several decades ago. In the early 2000s, Japan's national debt surpassed 100 percent of GDP, prompting widespread concerns that the mounting social security burden would lead to persistent fiscal deficits and, ultimately, a public debt crisis. While Japan did indeed run consistent fiscal deficits, such a crisis has yet to materialize. Instead, the Japanese public sector has strategically engineered a massive duration and currency mismatch on its balance sheet. By absorbing foreign exchange, aggregate, and interest rate risks, Japan's government has generated substantial excess returns on its risky investments. This additional income has played a crucial role in sustaining Japan's high debt levels and ongoing fiscal deficits. The Japanese experience, therefore, may potentially provide lessons as the United States navigates its own fiscal challenges.

Consolidated Public Sector Balance Sheet for the United States

We begin by conducting a consolidated balance sheet analysis for the United States, including the federal government, the Federal Reserve, and state and local governments, with the results reported in Table 2. Compared to Japan, the US government holds significantly fewer financial assets, amounting to just 45.7 percent of GDP in 2024. Most of these assets are concentrated in public employee pension funds or on the Federal Reserve's balance sheet in the form of equities, agency debt, and other debt securities. A substantial portion of the government's equity exposure comes through state and local public pension funds (in the table, S&L PPFs). In response to underfunded pension liabilities, US state and local governments have increasingly leveraged their balance sheets, employing an investment strategy

Table 2
Consolidated US Government Balance Sheet

Percentage of GDP, quarter end	1997Q4	2012Q4	2024Q2	Δ 1997–2024
Assets				
Deposits	5.0%	4.9%	4.3%	−0.7%
Agency debts	2.3%	10.5%	8.8%	6.5%
Of which held by Federal Reserves	0.0%	6.6%	6.8%	6.8%
Other debt securities	4.3%	4.9%	3.7%	−0.6%
Loans	5.4%	8.1%	8.8%	3.4%
Equities	14.7%	14.0%	12.7%	−2.0%
Of which held by S&L PPFs	12.0%	12.0%	11.4%	−0.6%
Other assets	3.8%	3.1%	7.3%	3.5%
Sum (assets)	35.6%	45.5%	45.7%	10.1%
Liabilities				
Currency	5.6%	7.4%	8.6%	2.9%
Bank reserve & repo	0.4%	9.6%	14.2%	13.8%
Marketable treasury securities	29.7%	50.9%	72.2%	42.5%
Municipal securities	12.1%	19.2%	11.2%	−1.0%
Non-marketable treasury securities	4.7%	3.4%	2.1%	−2.6%
Other liabilities	3.1%	6.5%	6.0%	2.9%
Unfunded pension	19.9%	29.0%	15.2%	−4.6%
Sum (liabilities)	75.6%	126.1%	129.4%	53.8%
Net liabilities	40.0%	80.6%	83.7%	43.7%

Source: US Flow of Funds and US National Income and Product Accounts. The net liabilities—defined as the sum of liabilities minus the sum of assets—should equal the present discounted value (PDV) of primary surpluses, which represents the implied market value of claims on future primary surpluses.
Note: Balance Sheet in Market Values as percentage of GDP. The consolidated US government includes the federal government, state and local government, social security funds, public employees defined-benefit pension funds, and Federal Reserves. S&L PPFs are state and local government defined benefit pension funds.

that parallels, in a less extreme way, Japan's approach to risk-taking (Myers 2022; Giesecke and Rauh 2023). Additionally, a large share of agency debt is held by the Federal Reserve as a result of its quantitative easing policies. The next most significant asset category is domestic loans, primarily student loans, which could potentially be subject to forgiveness.

Meanwhile, the public sector has borrowed extensively through debt securities, with national debt reaching 72.2 percent of GDP and municipal securities accounting for an additional 11.2 percent of GDP in 2024. Furthermore, unfunded pension liabilities across federal, state, and local governments amount to 15.2 percent of GDP. These pension liabilities are likely underestimated (Giesecke and Rauh 2023), further complicating the fiscal outlook. With total US public sector liabilities reaching 129.4 percent of GDP and a significantly smaller asset base, the US government's net liabilities amount to 83.7 percent of GDP—slightly higher than Japan's 77.6 percent of GDP. Moreover, while Japan's population has already reached an advanced stage of aging, the United States is only beginning to enter this

demographic shift. These factors challenge the common assumption that Japan's fiscal challenges are necessarily more severe than those facing the United States.

The Japanese government's strategy involves borrowing at extremely low interest rates and investing in risky, high-return assets. In contrast, this is not currently the case for the United States, at least not at the federal level. The US public sector balance sheet—particularly that of the federal government—shows a smaller scale of leverage relative to Japan. In 2024, risky equity accounts for only 12.7 percent of GDP, mostly held by state and local public pension funds. Most of its agency debt, around 6.8 percent of GDP, is held by the Federal Reserve and is declining due to the ongoing quantitative tightening policy. Furthermore, the risky investment strategy employed by Japanese public pension funds is not currently available to the US federal government. Unlike Japan, the US Social Security fund is restricted by the Social Security Act to holding only US Treasury securities. Because the US public sector does not currently engage in a risky investment strategy similar to Japan, its current asset holdings are highly unlikely to generate a significant excess rate of return above its funding cost.

A Sovereign Wealth Fund Based on Borrowed Money for the United States?

Could the United States duplicate Japan's strategy and sustain its high debt while running a consistent deficit? This question echoes recent proposals to establish a US sovereign wealth fund (for example, White House 2025; Cassidy 2025). Given the large net liabilities position of the US public sector (as show in Table 2), a US sovereign wealth fund would be funded by issuing additional debt, putting the United States on a path similar to that of Japan over the past two decades. As a result, the Japanese experience also provides insights into this issue.

As discussed earlier, the extra risks associated with the investments undertaken by sovereign wealth funds must be borne by either bondholders or taxpayers. However, it is unlikely that if bad outcome risks materialize, the government would then offload the additional costs onto taxpayers, because doing so would require raising more tax revenue during economic downturns. A more plausible scenario is that the extra risks are shifted to bondholders, who would, in turn, demand higher returns to compensate for the added risk. As a result, the returns earned by the sovereign wealth fund would be offset by higher funding costs, ultimately eliminating the positive spread. Without deliberate financial repression policies to suppress borrowing costs or significant financial market frictions, the government cannot cause bondholders to lend at below-market rates. The United States would probably find it more difficult than Japan to sustain low interest rates for government bondholders, for several reasons.

First, the investor base for US Treasury bonds differs significantly from that of Japanese government bonds. As shown in panel B of Figure 1, the Federal Reserve has reduced its holdings of Treasury securities to less than 20 percent of the total outstanding by stepping back from its quantitative tightening policy. Domestic banks now hold less than 10 percent of outstanding Treasury securities, while domestic nonbank financial institutions account for approximately 40 percent. Foreign

Table 3
Japan and United States Household Balance Sheet

	Japan		United States	
Percentage of GDP, year end	*1997*	*2023*	*1997*	*2023*
Assets				
Currency and deposits	127.6%	188.9%	43.4%	61.8%
Debt securities	11.7%	4.8%	20.0%	20.3%
Equities	20.4%	63.4%	128.2%	196.9%
Insurance & pension	63.1%	90.2%	112.0%	117.3%
Other assets	12.8%	11.9%	10.6%	9.7%
Sum (assets)	253.7%	352.9%	314.1%	406.0%
Liabilities				
Loans	64.8%	61.9%	63.2%	69.0%
Other liabilities	10.5%	2.4%	0.2%	0.1%
Sum (liabilities)	75.3%	64.4%	63.5%	69.1%
Net wealth	160.4%	294.8%	250.6%	336.9%

Source: Japan's Flow of Funds. Federal Reserve Board of Governors, Z.1 Financial Accounts of the United States, Table B.101.h.
Note: Unit: percentage of GDP.

investors hold about 30 percent of the market. Overall, the central bank and regulated financial institutions play a much smaller role in holding US Treasury securities, leaving a larger share to more price-elastic mutual funds and foreign investors. The overall liquidity of the Treasury market is significantly better than that of Japanese government bonds. These factors make a persistent mispricing of Treasury securities less likely.

    Second, US households exhibit greater financial sophistication than their Japanese counterparts. Table 3 contrasts the portfolio held by the Japanese household with the US household. Japanese households hold half of their wealth in low-return demand deposits, providing the government with an abundant and inexpensive funding source. In contrast, the US household allocates a much larger share of its wealth to stocks, bonds, and other long-duration assets rather than deposits. For Japanese households, a much higher share in deposits is mirrored in a much lower financial market participation rate (around 23 percent own stocks, bonds, or mutual funds), compared to US households (well over 60 percent). The portfolio of Japanese households holds the key for why the inflation rate of Japan remains low despite the massive liquidity injection by the quantitative easing and yield curve control policy by the Bank of Japan. Even if the US government is willing to implement low-rate policies at the expense of depositors, the effect on lowering its funding cost may be smaller than in Japan, because American households seem more willing to substitute away from deposits. Consequently, US policymakers face greater challenges in employing financial repression as a tool to support its fiscal deficits.

Third, a sustained low-interest-rate policy in the United States poses a risk to what is called the "convenience yield" on US Treasury securities—that is, the fact that many investors want to hold the securities themselves and to have immediate access to them. This convenience yield reflects the status of the US dollar as the global reserve currency. After Japan's decades of current account surpluses, it has one of the largest positive net external asset positions. In contrast, the US economy has run persistent current account deficits for three decades, accumulating a large negative net foreign asset position. To finance this position, the US economy relies on attracting foreign capital inflows. Given the central role of the US dollar in the international financial system, the United States has historically been able to borrow from the rest of the world at relatively low rates, in US dollars, to fund its current account and government deficits. In the past, US Treasury bonds have benefited from large convenience yields, particularly during periods of global financial stress. However, there is evidence that these convenience yields have largely disappeared (Du, Im, and Schreger 2018). As one sign, the spread between Treasury yields and AAA corporate bond yields has narrowed, pointing to a deterioration of the Treasury convenience relative to other relatively safe dollar debt (for example, Gómez-Cram, Kung, and Lustig 2024). If the United States was to maintain a prolonged low-interest-rate policy that led to sustained depreciation of the dollar against other currencies, its status as a provider of safe assets could be further undermined.

Fourth, a US sovereign wealth fund is unlikely to earn substantial currency risk premia from its investments. By investing in US dollar–denominated assets on net, without hedging this exposure, the Japanese public sector captures significant currency risk premia. However, for a US dollar–based sovereign wealth fund, the pool of foreign currencies offering higher yields, such as those of Australia, New Zealand, Canada, and emerging markets, is much more limited. Additionally, if the US public sector were to invest on net in considerable foreign assets, the US private sector would need to offset these outflows by attracting even greater foreign capital inflows, given the overall large negative net foreign asset position of the US economy.

Conclusion

To keep its promises to bondholders and to Japanese transfer recipients, the Japanese government has combined low-interest-rate policies with risk-taking on the asset side. The Japanese public sector is running a highly leveraged and increasingly risky sovereign wealth fund, exposing its taxpayers, bondholders, and depositors to significant interest rate, exchange rate, and equity risks. The risky investments are funded by borrowing from banks and ultimately depositors at below-market rates. This strategy has worked particularly well over the past two decades, amid a generally low-yield global macroeconomic environment. Returns on risky investments have sustained the debt-to-GDP ratio in Japan despite the lack of primary surpluses. Going forward, the sizable duration and currency mismatch on the public sector's

balance sheet exposes Japan to significant mark-to-market losses that could trigger a fiscal crisis and even a run on the Japanese currency.

The United States faces similar fiscal pressures but is unlikely to replicate Japan's approach. Unlike Japan, the US government has relied more on external funding to support its deficits. A prolonged policy that sought to help the federal government finance borrowing with below-market interest rates could erode the dollar's reserve currency status and further diminish the convenience yield on US Treasury bonds.

■ *The views in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Federal Reserve Bank of St. Louis or any other person associated with the Federal Reserve System.*

References

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

Aquilina, Matteo, Marco Lombardi, Andreas Schrimpf, and Vladyslav Sushko. 2024. "The Market Turbulence and Carry Trade Unwind of August 2024." BIS Bulletin 90.

Bank for International Settlements. 1997–2025. *Effective Exchange Rates.* BIS Data Portal. https://data.bis.org/topics/EER/BIS,WS_EER,1.0/M.R.B.JP (accessed on September 13, 2025).

Bank of Japan. 2025. *Flow of funds accounts: Japan.* https://www.stat-search.boj.or.jp/ssi/cgi-bin/famecgi2?cgi=\protect\T1\textdollarnme_a000_en&lstSelection=FF (accessed on September 9, 2025).

Blanchard, Olivier. 2019. "Public Debt and Low Interest Rates." *American Economic Review* 109 (4): 1197–229.

Bloomberg L. P. 2025. *Asset return data from Bloomberg terminal.* Bloomberg Terminal. (Accessed on September 25, 2025.) Access modality: file-based (CSV).

Board of Governors of the Federal Reserve System. 2023. *Foreign Portfolio Holdings of US Securities.* US Department of the Treasury.

Board of Governors of the Federal Reserve System. 2025. *Financial Accounts of the United States (z.1 release).* https://www.federalreserve.gov/releases/z1/ (accessed on September 25, 2025).

Cabinet Office, Government of Japan. 2025. *Quarterly estimates of Japan GDP: 1994–2024.* https://www.esri.cao.go.jp/jp/sna/data/data_list/sokuhou/files/2024/qe243_2/tables/gaku-mk2432.csv (accessed on September 25, 2025).

Campbell, John Y. 1995. "Some Lessons from the Yield Curve." *Journal of Economic Perspectives* 9 (3): 129–52.

Cassidy, Bill. 2025. "ICYMI: Cassidy Outlines Plan to Save Social Security in Op-Ed." Press Release, July 8. https://www.cassidy.senate.gov/newsroom/press-releases/icymi-cassidy-outlines-plan-to-save-social-security-in-op-ed/.

Chien, Yi-Li, Harold L. Cole, and Hanno Lustig. 2023. "What about Japan?" NBER Working Paper 31850.

De Marco, Filippo, and Marco Macchiavelli. 2016. "The Political Origin of Home Bias: The Case of Europe. FEDS Working Paper 2016-60.

Doi, Takero, and Takeo Hoshi. 2002. "Paying for the FILP." NBER Working Papers 9385.

Du, Wenxin, and Amy W. Huber. 2024. "Dollar Asset Holdings and Hedging around the Globe." NBER Working Paper 32453.

Du, Wenxin, Joanne Im, and Jesse Schreger. 2018. "The US Treasury Premium." *Journal of International Economics* 112: 167–81.

Du, Wenxin, and Jesse Schreger. 2022. "CIP Deviations, the Dollar, and Frictions in International Capital Markets." In *Handbook of International Economics*, Vol. 6, edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, 147–97. North-Holland.

Du, Wenxin, Alexander Tepper, and Adrien Verdelhan. 2018. "Deviations from Covered Interest Rate Parity." *Journal of Finance* 73 (3): 915–57.

Federal Reserve Bank of St. Louis. 2025. *Holders of U.S. Treasury Debt: Selected series from FRED.* https://fred.stlouisfed.org/. (accessed on September 25, 2025). Access modality: file-based (CSV).

Fernández-Villaverde, Jesús, Gustavo Ventura, and Wen Yao. 2025. "The Wealth of Working Nations." *European Economic Review* 173: 104962.

Financial Markets Department, Bank of Japan. 2024. *Liquidity Indicators in the JGB Markets.* Bank of Japan.

Giesecke, Oliver, and Joshua Rauh. 2023. "Trends in State and Local Pension Funds." *Annual Review of Financial Economics* 15: 221–38.

Gómez-Cram, Roberto, Howard Kung, and Hanno Lustig. 2024. "Government Debt in Mature Economies. Safe or Risky?" Preprint, SSRN. http://dx.doi.org/10.2139/ssrn.4935930.

Gordon, Myron J. 1959. "Dividends, Earnings, and Stock Prices." *Review of Economics and Statistics* 41 (2): 99–105.

Hall, George J., and Thomas J. Sargent. 2011. "Interest Rate Risk and Other Determinants of Post-WWII US Government Debt/GDP Dynamics." *American Economic Journal: Macroeconomics* 3 (3): 192–214.

Hansen, Gary D., and Selahattin İmrohoroğlu. 2023. "Demographic Change, Government Debt and Fiscal Sustainability in Japan: The Impact of Bond Purchases by the Bank of Japan." *Review of Economic Dynamics* 50: 88–105.

Hassan, Tarek A., and Rui C. Mano. 2019. "Forward and Spot Exchange Rates in a Multi-currency World." *Quarterly Journal of Economics* 134 (1): 397–450.

Hoshi, Takeo, and Ayako Yasuda. 2015. "Capital Market Regulation in Japan after the Global Financial Crisis." In *The New International Financial System: Analyzing the Cumulative Impact of Regulatory Reform*, edited by Douglas D. Evanoff, Andrew G. Haldane, and George G. Kaufman, 165–95. World Scientific.

Jiang, Zhengyang, Hanno Lustig, Stijn Van Nieuwerburgh, and Mindy Z. Xiaolan. 2020. "Manufacturing Risk-Free Government Debt." NBER Working Paper 27786.

Jordà, Òscar, Katharina Knoll, Dmitry Kuvshinov, Moritz Schularick, and Alan M. Taylor. 2019. "The Rate of Return on Everything, 1870–2015." *Quarterly Journal of Economics* 134 (3): 1225–98.

Koeda, Junko, and Yosuke Kimura. 2021. "Government Debt Maturity in Japan: 1965 to the Present." Tokyo Center for Economic Research Working Paper E-163.

Lustig, Hanno, Nikolai Roussanov, and Adrien Verdelhan. 2011. "Common Risk Factors in Currency Markets." *Review of Financial Studies* 24 (11): 3731–77.

Lustig, Hanno, Andreas Stathopoulos, and Adrien Verdelhan. 2019. "The Term Structure of Currency Carry Trade Risk Premia." *American Economic Review* 109 (12): 4142–77.

Mehrotra, Neil R., and Dmitriy Sergeyev. 2021. "Debt Sustainability in a Low Interest Rate World." *Journal of Monetary Economics* 124 (S): S1–18.

Modigliani, Franco, and Merton H. Miller. 1958. "The Cost of Capital, Corporation Finance and the Theory of Investment." *American Economic Review* 48 (3): 261–97.

Myers, Sean. 2022. "Public Employee Pensions and Municipal Insolvency." Preprint, SSRN. http://dx.doi.org/10.2139/ssrn.4533980.

OECD. 2022. *General Government Debt.* OECD. https://www.oecd.org/en/data/indicators/general-government-debt.html (accessed on August 10, 2025).

Ongena, Steven, Alexander Popov, and Neeltje Van Horen. 2019. "The Invisible Hand of the Government: Moral Suasion during the European Sovereign Debt Crisis." *American Economic Journal: Macroeconomics* 11 (4): 346–79.

Reinhart, Carmen M., Jacob F. Kirkegaard, and M. Belen Sbrancia. 2011. "Financial Repression Redux." *Finance and Development* 48 (1): 22–27.

United Nations. 1960–2024. *Population Ages 65 and Above (% of Total Population).* UN Population Division, distributed by World Bank Group. https://data.worldbank.org/indicator/SP.POP.65UP.TO.ZS (accessed on August 10, 2025).

US Bureau of Economic Analysis. 2025. *Table 1.1.5. Gross Domestic Product (GDP): Quarterly, Nominal.* https://apps.bea.gov/iTable/?reqid=19&step=2&isuri=1&categories=survey (accessed on September 25, 2025).

US Department of the Treasury. 2025. *Treasury Bulletin.* https://fiscal.treasury.gov/reports-statements/treasurybulletin/current.html (accessed on September 25, 2025). Access modality: file-based (PDF). Includes data from Table OFS-2: Estimated Ownership of US Treasury Securities.

White House. 2025. "A Plan For Establishing a United States Sovereign Wealth Fund." Presidential Action, February 3. https://www.whitehouse.gov/presidential-actions/2025/02/a-plan-for-establishing-a-united-states-sovereign-wealth-fund/.

World Bank Group. 2025. *Population Estimates and Projections.* World Bank Group. https://databank.worldbank.org/source/population-estimates-and-projections (accessed on August 10, 2025).

Putting US Fiscal Policy on a Sustainable Path

Karen Dynan and Douglas Elmendorf

F or decades, analysts have warned that the aging of the population and rising health care costs, together with repeated tax cuts that held revenue steady relative to output, were putting US fiscal policy on an unsustainable path. Sharp runups in federal debt spurred by the global financial crisis and the COVID pandemic have significantly accentuated the problem. Absent very favorable economic changes that are highly unlikely, the United States will ultimately need to scale back large and popular benefit programs, increase taxes relative to historical experience, or both.

Yet over the past 25 years, policymakers have taken no significant action to address rising debt. Instead, they have enacted legislation that, on balance, increases debt substantially. Two bipartisan efforts to forge agreement on policy changes that would lower the debt trajectory—the 2010 National Commission on Fiscal Responsibility and Reform (commonly known as the "Bowles-Simpson commission" after its cochairs) and the 2011 Joint Select Committee on Deficit Reduction (commonly known as the "Supercommittee")—failed to gain broad political support. This lack of action represents a distinct change from policymakers' behavior in the 1980s and 1990s. During those decades, policymakers enacted some

■ *Karen Dynan is Professor of the Practice of Economics, Harvard University, Cambridge, Massachusetts, and a Nonresident Senior Fellow at the Peterson Institute for International Economics, Washington, DC. Douglas Elmendorf is Professor of Public Policy, Harvard Kennedy School, Cambridge, Massachusetts. Dynan and Elmendorf are both Research Associates, National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are kdynan@fas.harvard.edu and doug.elmendorf@hks.harvard.edu.*

Figure 1
Federal Debt
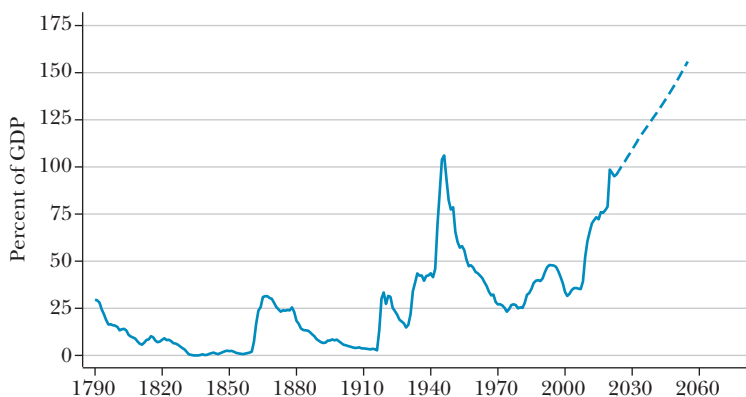
Source: Congressional Budget Office, *Budget and Economic Data*, available at https://www.cbo.gov/data/budget-economic-data.
Note: Solid line is history; dashed line is projection under current law as of March 2025.

legislation that increased debt, but they also enacted legislation that reduced debt. Recent estimates of the reaction function of fiscal policymakers confirm the impression that policymakers have stopped responding to rising debt in substantive ways (Auerbach and Yagan 2024).

As a result, the United States now faces a fiscal challenge that is unprecedented for the country. The publicly held debt of the US federal government is roughly equal to annual gross domestic product (GDP), a level that has been exceeded previously for only a few years around the end of World War II and that is nearly three times the level in 2007 on the eve of the global financial crisis. Moreover, the Congressional Budget Office (2025a) projected in early 2025 that, under then-current law, annual deficits over the coming decade would be close to 6 percent of GDP, which the country has never experienced apart from wars and economic downturns. The continuing large deficits would cause debt to rise rapidly further relative to GDP, as shown in Figure 1.

The projected path of debt is notably higher today than was projected under current law in 2007, as shown in Figure 2. That upward shift occurred not because aging and health care costs have had larger effects than expected; indeed, health care costs have risen less than projected in 2007 and account for some of the downward pivot in projected debt growth between 2020 and 2025. Instead, the jump owes to legislation making significant cuts in taxes and to the effects of two significant recessions and the policy responses to them. For further discussion of revisions to the budget outlook, see Dynan (2023).

We show in the next section that, even allowing for uncertainty about the future economy, US fiscal policies are almost certainly unsustainable. Therefore, the key

Figure 2
Vintages of Federal Debt Projections
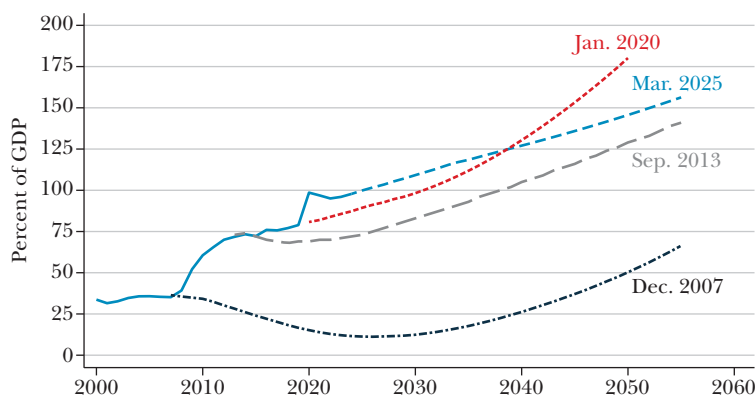
Source: Congressional Budget Office, *Budget and Economic Data*, available at https://www.cbo.gov/data/budget-economic-data.
Note: Solid line is history; dotted and dashed lines are projections under then-current law.

questions for policymakers are when and in what ways policies might be changed. We address those questions in turn and then conclude with comments on the politics of fiscal policy.

In this essay, we focus exclusively on the federal budget, although the budgets of many state and local governments also face major challenges. We also note that federal debt held by the public includes debt held by the Federal Reserve. A more comprehensive measure of the federal government's financial position would consolidate the Federal Reserve's balance sheet and also include other federal assets and liabilities. However, making those adjustments would not alter our main points. More generally, debt is just one of many commitments made by the federal government, but it has a special role legally, in the financial system, and in people's perceptions. Explicit government commitments beyond the current debt, such as future payments for Medicare and Social Security, are reflected in the projections we discuss.

Might Current US Fiscal Policies Turn Out to Be Sustainable Even Without Changes?

Future economic conditions are highly uncertain. To the extent that those conditions differ from what is expected today—in demographics, productivity, interest rates, health care costs, or other elements—future budget outcomes will differ from current projections as well. Might US debt stabilize relative to GDP without changing current tax and spending policies?

Assessing Sustainability

A useful starting point is to think of this year's change in the debt-GDP ratio as composed of two pieces. One piece is this year's primary deficit, which equals spending apart from interest payments less revenue. The other piece is the evolution of accumulated debt from previous years, which depends on the difference between interest rates and GDP growth: higher interest rates lead to higher interest payments, which raise the debt-GDP ratio, and faster GDP growth lowers that ratio.[1] Thus, the difference between interest rates and growth rates is key to debt dynamics, and the difference is especially consequential when debt is large relative to GDP. These various factors are connected in multiple ways; for example, the taxes and spending that make up the primary deficit are affected by economic growth, and economic growth and interest rates are affected by the amount of debt.

Between 1962 and 2007, the average primary deficit for the federal government was 0.1 percent of GDP, and the interest rate on federal debt averaged 0.4 percentage points less than the growth rate of GDP. The debt-GDP ratio changed little, on balance, and debt stood at 35 percent of GDP in 2007. Then the situation changed sharply, in two countervailing ways. From 2008 through 2024, primary deficits averaged 4.6 percent of GDP, and the interest rate on debt averaged 1.8 percentage points less than the GDP growth rate. The former shift had a much larger effect on the debt-GDP ratio, which jumped upward to 98 percent in 2024.

Looking ahead over the next few decades, the Congressional Budget Office (2025c) projected that, under then-current law, the primary deficit would be close to 2 percent of GDP. CBO also projected that interest rates would be close to economic growth rates, as shown in Figure 3 (which adjusts both interest and growth rates for realized inflation). Under those conditions, debt would be 156 percent of GDP by 2055 and rising rapidly (as shown in Figure 1).

Since that projection was made, Congress and the president approved legislation in summer 2025 that cut taxes appreciably and cut spending to a lesser degree, widening estimated budget deficits by more than $4 trillion during the decade between 2025 and 2034 (CBO 2025e). However, the president also imposed tariffs on imports that would narrow budget deficits by nearly the same amount (CBO 2025f). Those tariffs are currently under significant legal challenge, but the administration has stated that, if it loses in court, it will pursue alternative legal approaches to the same end. In light of the administration's determination, some significant revenue from tariffs seems more likely than not to become part of CBO's current-law baseline. Therefore, because the net deficit effects of these recent developments will probably be much smaller than the $21 trillion in deficits for the decade previously projected by CBO (2025c), this paper uses those previous projections.[2]

---

[1] In algebraic terms, $Change\ in\ Debt/GDP \cong Primary\ Deficit/GDP + \big((r-g)/(1+g)\big) \times Debt/GDP_{-1}$, where r is the interest rate on federal debt averaged across all outstanding securities and g is the growth rate of GDP. The equation is only approximately correct because it omits some government financial transactions.

[2] Projections from the Congressional Budget Office are based on the law at the time projections are made (including the executive branch's implementation of the law, for example regarding tariffs). If

Figure 3
Inflation-Adjusted Interest Rate and GDP Growth
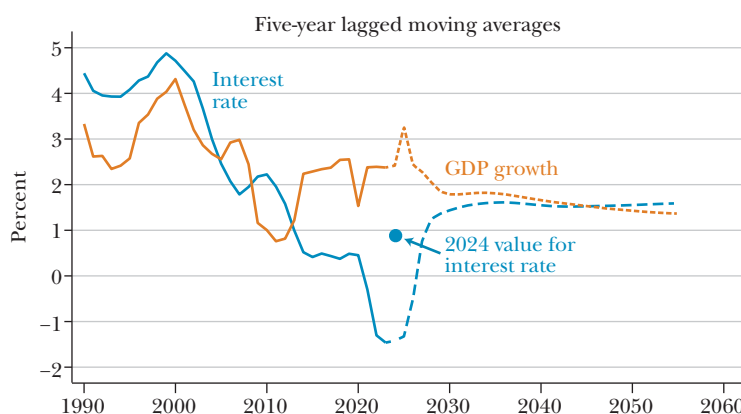
Source: Authors' calculations based on data from Bureau of Economic Analysis and Congressional Budget Office, *Budget and Economic Data*, available at https://www.cbo.gov/data/budget-economic-data.
Note: Solid lines are history; dotted and dashed lines are 2025 projections. Beginning in 1995, the inflation-adjusted interest rate is calculated as the nominal average interest rate published by CBO minus GDP price inflation. Prior to 1995, the inflation-adjusted interest rate is calculated as $100 \times$ *net interest/ federal debt* (CBO estimates) minus GDP price inflation.

Researchers have assessed the sustainability of fiscal policy (for the United States and other countries) in various formal and informal ways; for examples, see Debrun et al. (2019) and Mian, Straub, and Sufi (forthcoming). Our goal in this section is to offer an intuitive understanding of what developments would be needed to make current fiscal policies sustainable. We explore the possibilities that, relative to CBO's projections, primary deficits would turn out to be smaller, economic growth faster, or interest rates lower. To be sure, the opposite possibilities are present as well, and we draw out the implications of that uncertainty later in the paper. Because the outlook becomes increasingly uncertain at longer horizons, we follow CBO's practice of focusing on the next 30 years.

Primary Deficits
Might the difference between noninterest spending and revenue end up substantially less than 2 percent of GDP even without policy changes?

On the spending side, such narrowing seems unlikely. Spending for Social Security and the major health care programs—Medicare (net of premiums paid),

---

those laws include the expiration of some existing policies, as is generally the case, CBO's projections incorporate the effects of the expirations. Accordingly, budget analysts sometimes distinguish between projections based on current laws and projections based on continuing current policies. Because the deficit effects of the policies currently scheduled to expire are small relative to the deficits projected under current law, this paper employs the terms "current law" and "current policies" interchangeably.

Figure 4
Key Components of the Federal Budget
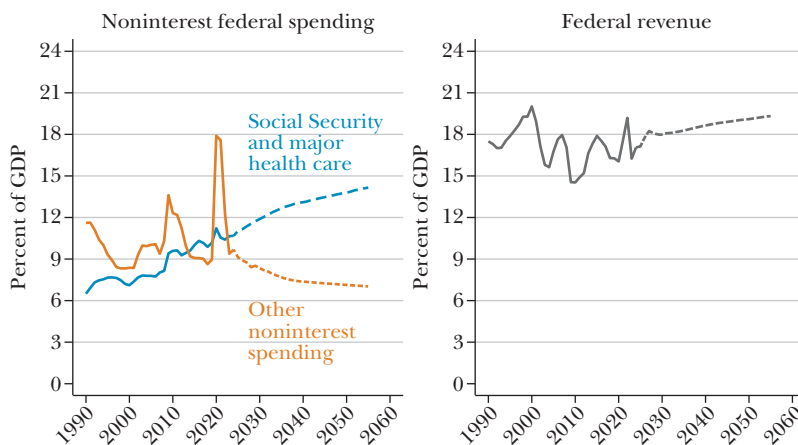
Note: Solid lines are history; dotted and dashed lines are 2025 projections.

Medicaid, the Children's Health Insurance Program, and premium tax credits established by the Affordable Care Act—was roughly 8 percent of GDP 30 years ago, is about 11 percent of GDP today, and is projected to be more than 14 percent of GDP 30 years ahead, as shown in the left panel of Figure 4. One key factor underlying those past and projected uptrends is the aging of the population; the other key factor is rising health care spending per person relative to output per person (adjusted for demographic change).

Population aging will certainly continue, but the future path of health spending is especially difficult to predict: On one hand, the growth rate has trended down significantly, because of both policy choices and other pressures to control costs; on the other hand, emerging new treatments may raise spending, because even when advances in medicine reduce health care spending at a point in time, the increases in lifespans can raise total spending for Social Security and health care. The Congressional Budget Office projections incorporate rising health care costs per person that allow for growth in spending between 2025 and 2055 equal to 1.4 percent of GDP; of course, if cost growth falls more rapidly, primary deficits would be smaller.

All other spending apart from interest payments includes outlays set in annual appropriations for defense and nondefense programs (with the latter including homeland security, highways, housing assistance, research, veterans' health care, and most of the government's operations) and outlays for benefit programs other than Social Security and health care (such as food stamps, veterans' compensation, and federal employee retirement benefits). CBO's methodology causes projected spending for this collection of activities to drop quickly below the lowest

mark relative to GDP seen in many decades, as shown in the left panel of Figure 4. Declines below CBO's projection seem very unlikely.

Developments on the tax side, in the absence of policy changes, also seem unlikely to materially shrink the projected gap between noninterest spending and revenue. Federal revenue has shown no trend relative to the size of the economy, equaling about 17 percent of GDP both 60 years ago and today, as shown in the right panel of Figure 4. The effect of the progressive tax code in increasing revenue relative to GDP as inflation-adjusted incomes rise has been offset over time by legislated tax cuts. The Congressional Budget Office projects that, under current law, revenue will rise to a little over 19 percent of GDP in 2055.

Tax revenue has varied considerably from year to year (even apart from the effects of policy changes), but sustaining higher revenue relative to GDP under current law would require more income in higher tax brackets, through either rapid broad-based growth of total income or a persistent increase in income inequality. For example, CBO (2024a) estimated that, in 2019, households in the top 1 percent of the pre-tax, pre-transfer income distribution paid an average federal tax rate of 30 percent, and households in the middle three quintiles paid an average federal tax rate of 14 percent.[3] CBO estimated also that households in the middle three quintiles received 43 percent of total income that year and the top percentile received 16 percent. If 6 percent of total income shifted from the middle quintiles to the top percentile, federal tax revenue would rise by about 1 percent of GDP ($0.06 \times [0.30 - 0.14]$). However, such a rise in inequality also would increase eligibility for federal benefits, which would offset the higher tax collections to some degree; for further analysis, see Pierce (2021).

Economic Growth

Overall economic growth can be divided into two parts: growth in total factor productivity and growth in the quantity and quality of the capital stock and the labor force. Faster growth of capital or labor would increase economic growth, but sustaining such changes for decades would require ever-rising investment or higher fertility or immigration rates, and significant changes of those sorts do not seem plausible. In any case, with annual potential labor force growth now running a full percentage point below its average pace in the 1980s and 1990s (leaving aside the effects of the immigration surge of the past few years), returning GDP growth to the 3.3 percent average pace of those decades would be especially challenging.

What about productivity? The Congressional Budget Office projects that total factor productivity growth in the nonfarm business sector will average 1.0 percent per year over the next 30 years. If productivity growth was 0.5 percentage points per year higher, the debt-to-GDP ratio in 2055 would be more than 40 percentage points below what it would be otherwise (CBO 2025d). In this analysis, faster productivity growth

---

[3] The CBO estimates mentioned here are through 2021, but we use the figures for 2019 to avoid distortions related to the pandemic and associated policy responses. CBO used broad measures of income and taxes, and it defined pre-tax, pre-transfer income to include social insurance benefits such as Social Security and Medicare.

causes the average interest rate on debt to increase by a comparable amount, so the difference between the interest rate and the growth rate changes little. Thus, the flattening of the debt path stems from the effect of faster growth on primary deficits.

Productivity growth might well exceed CBO's expectation for some period of time through the development and expanding use of artificial intelligence or other transformative innovations. But sustaining faster productivity growth for decades is more difficult, as evidenced by the computer and online revolutions, which generated fast growth for only limited periods. Using data since 1955 and measuring with the productivity series underlying estimated potential output, the 30-year moving average of total factor productivity growth has ranged from 1.0 percent to 1.3 percent. Therefore, sustaining growth that is 0.5 percentage points above CBO's projection of 1.0 percent seems very unlikely. In addition, faster productivity growth through artificial intelligence might displace workers or reduce work hours so that output would not rise in parallel with productivity and utilization of safety-net programs would increase.

Interest Rates

The inflation-adjusted average interest rate on federal debt holds fairly steady around 1.5 percent over the coming decades in CBO's projections, as shown in Figure 3. If the rate instead declined by 0.05 percentage points per year (for a cumulative decline of 1.5 percentage points over the 30-year projection period, putting the terminal value close to zero), the debt-to-GDP ratio in 2055 would be 35 percentage points below what it would be otherwise (CBO 2025d). Is such a decline in rates plausible?

US Treasury yields fell significantly during the 2000s and 2010s, with the nominal yield on ten-year Treasury notes dropping about 3.5 percentage points between 1999 and 2019 even as inflation changed little. The magnitude and persistence of the decline in inflation-adjusted Treasury yields—which was accompanied by declines in yields on both government and corporate debt in many other advanced economies—has generated a great deal of research regarding the global supply and demand for loanable funds. The drop in yields appears to have been caused by, among other factors, a reduced demand for investment stemming from slower trend growth of productivity and the labor force, along with an increased supply of saving as the effects of an older population and greater wealth inequality more than offset the impact of a surge in US federal debt (for example, see Gamber 2020; Auclert et al. 2025).

Many of the factors that held down yields during the 2000s and 2010s are still present, so yields may stay fairly low. However, yields are being pushed up by other forces, including continued increases in federal debt, greater investment in computing power and the energy transition, and a larger risk premium in light of the recent inflation experience. The CBO projections presume that interest rates will stay above their level during the 2010s but below their level in earlier years. Using data since 1962, the 30-year moving average of the inflation-adjusted interest rate on federal debt has ranged from 1.3 percent to 3.6 percent. Therefore, sustaining interest rates as low as in the alternative path explored by CBO seems very unlikely.

One significant source of uncertainty is the sensitivity of interest rates to increases in federal debt—a sensitivity that is difficult to estimate because of endogeneity (weak economic activity causes debt to rise and interest rates to fall, but the rise in debt does not cause the fall in rates) and also because the sensitivity might change over time. Gust and Skaperdas (2024) documented estimates by previous researchers that an increase of 1 percentage point in the debt-to-GDP ratio increases long-term interest rates by 1 to 6 basis points, with more-recent estimates that use a greater span of data on the lower side (see also Neveu and Shafer 2024; Plante, Richter, and Zubairy 2025).

A further source of uncertainty is that government policies can reduce interest rates on federal debt. In particular, regulations to lessen the liquidity and solvency risk of financial institutions can push those institutions to hold more federal debt than they would otherwise—if implemented to hold down government borrowing costs, this practice is known as "financial repression." Additionally, expansionary monetary policy can generate higher inflation that reduces realized inflation-adjusted yields and the value of outstanding debt relative to nominal GDP—if implemented to hold down government borrowing costs, a practice known as "fiscal dominance."

Both practices have been analyzed in substantial research literatures; as a starting point on financial repression, see Reinhart, Kirkegaard, and Sbrancia (2011), and as a starting point on fiscal dominance, see Leeper (1991, 2023) and Cochrane (2023). These practices are effectively taxation through alternative means; if banks are forced to hold debt at low interest rates, then bank owners or depositors are giving up wealth to the government, and if inflation rises, then holders of government debt are giving up wealth to the government. Moreover, both financial regulation and inflation have had quantitatively important effects on debt in the past: by one estimate, the drop in the debt-GDP ratio following World War II stemmed largely from unexpected inflation and the pegging of interest rates prior to the Treasury-Federal Reserve Accord of 1951 (Acalin and Ball forthcoming), and high inflation in 2021 through 2023 (stemming from multiple sources) temporarily stabilized the debt-GDP ratio despite large primary deficits.

Looking ahead, the incentives of political leaders to adopt these strategies will increase as debt rises relative to GDP—but still, the strategies seem unlikely to address a large share of the projected fiscal imbalance. Regarding financial repression, current regulations create strong incentives for financial institutions to hold government debt. However, attempts by regulators to strengthen those incentives have long faced substantial resistance based on concerns about constricting the supply of credit to private enterprises and driving risk-taking into less-regulated financial entities. Thus, significant tightening of restrictions on an ongoing basis would be surprising. Regarding fiscal dominance, the Federal Reserve's independence in monetary policymaking is under significant pressure as we write. The president is strongly criticizing the Federal Reserve's current policy stance and calling for a large reduction in the federal funds interest rate—explicitly, in part, on the grounds that lowering interest rates would improve the federal budget situation.

Yet, the higher inflation that would result from excessive monetary stimulus would be viewed very negatively by voters (Stantcheva 2024), which provides a strong incentive for policymakers to avoid it. In this regard, the Federal Reserve's reluctance to tighten monetary policy as inflation climbed in 2021 stemmed from a misinterpretation of incoming data rather than a deliberate accommodation of expansionary fiscal policy (Dynan and Elmendorf 2024). Moreover, higher inflation might not even improve the fiscal situation: It would reduce the burden of outstanding debt, but it would increase interest rates on new debt through higher expected inflation and a higher risk premium. Raising interest rates on new debt would be very costly for the government because the borrowing needed to both roll over outstanding debt and finance ongoing deficits is so large. For example, if inflation and interest rates were both 0.5 percentage points higher than projected, the estimates in CBO (2025b) imply that deficits over the coming decade would be about $1.6 trillion larger than currently projected. All told, monetization of substantial amounts of debt on an ongoing basis would be surprising.

Summing Up Prospects for Debt Sustainability

The authors of this paper have spent considerable time engaged in economic and budget forecasting, and we believe that forecasts should be offered with humility and received with skepticism. Nonetheless, we conclude that current US fiscal policies are almost certainly unsustainable, because stabilizing the debt-GDP ratio under current policies would require large and favorable shifts in interest rates, productivity growth, primary deficits given existing policies, or all three.

Moreover, unfavorable shifts in those factors are entirely possible as well. CBO (2025d) illustrates both the upside and downside possibilities for debt under alternative assumptions about future economic conditions. CBO's central projection is that, without changes in policies, the debt-GDP ratio would be 156 percent in 2055. If total factor productivity growth was 0.5 percentage points faster than CBO expects, then the debt-GDP ratio in 2055 would be 113 percent, but if productivity growth were 0.5 percentage points slower than CBO expects, that ratio would reach 203 percent. Similarly, if the average interest rate on federal debt declined by 0.05 percentage points per year, the debt-GDP ratio in 2055 would be 121 percent, but if the rate rose by that amount, the ratio would be 204 percent.

Because current fiscal policies are almost certainly unsustainable, the key questions for policymakers become *when* to change policies and *in what ways* to change policies. One might expect a third question of *how much* policies might be changed. However, that question is a variant of *when* policies might be changed, because the longer changes are delayed, the larger they need to be. That implication is straightforward for reaching a preselected debt-GDP ratio, because the additional years of rising debt would need to be offset by less debt accumulation later, but the implication also holds for stabilizing the debt-GDP ratio at any level, because higher debt leads to larger interest payments that need to be offset by a larger primary surplus. CBO (2025d) projected that keeping the debt-GDP ratio at its current mark would require a reduction in primary deficits relative to CBO's central projection

averaging 1.6 percent of GDP over the next three decades. Delaying policy changes until debt increases further relative to GDP would require larger changes in taxes and noninterest spending in order to stabilize the debt-GDP ratio.

Benefits of Restraining the Growth of US Government Debt

Acting sooner rather than later to restrain the growth of US government debt would reduce the costs of growing debt and broaden the policy options for limiting debt growth. We take up these effects in this section.

Raising National Savings

Restraining debt growth would increase national savings. Greater US savings would provide funds for greater capital investment in this country, which would raise the compensation of US workers through higher wages and benefits than would otherwise occur. Greater US savings also would provide funds for greater capital investment elsewhere in the world, which would raise the capital income received by US residents. Moreover, greater US savings (and reduced uncertainty about how an unsustainable fiscal situation would be resolved in the future) would boost the probability that US financial markets and the US dollar remain at the center of the global economy. These differences would enhance future living standards for Americans.

Because total factor productivity growth raises living standards over time, it is not obvious how much current generations should give up to increase the well-being of future generations. However, making sacrifices today to help today's children and others who will be living in the future has been part of our national ethos, and in the past, that approach has been carried out in part by limiting government debt. Moreover, current generations will be leaving to future generations many substantial challenges, including threats to national security, climate change and other environmental problems, and the possibility of economic and social disruptions from artificial intelligence and other technologies.

The quantitative impact of higher savings on future living standards is not as great as one might expect offhand. Ball and Mankiw (1995) estimated that eliminating the national debt—then equal to about 50 percent of GDP—would have raised national income by somewhat less than 6 percent. That effect on the level of income is small relative to the growth of national income over time but large relative to the effects of many other potential policy changes.

The economic value of increasing national savings depends on the scarcity of capital. As the decline in interest rates during the 2000s and 2010s became more apparent, Summers (2014, 2015) argued that the economy was suffering from "secular stagnation" and that the economic impact of government borrowing was very different under those conditions. Research has shown that most, but not all, of the possible reasons why Treasury yields declined signal that capital is less scarce than it was before and therefore the value of increasing national savings is lower

than it was before (for discussion, see Elmendorf and Sheiner 2017 in this journal; Eggertsson, Mehrotra, and Robbins 2019; Blanchard 2019, 2023; Ball and Mankiw 2023). Because Treasury yields have reversed only part of that decline, we infer that the value of increasing national savings probably remains somewhat lower than it was in the 1980s and 1990s.

Optimal national savings can also be affected by the aging of the population. Aging stems from a combination of lower fertility and greater longevity. Both of those changes increase optimal savings to provide for a larger retired population relative to the working population, but lower fertility also decreases optimal savings because a smaller labor force reduces the marginal product of capital (Cutler et al. 1990). In a calibrated model, Elmendorf and Sheiner (2017) found that the net effect at the time was to raise optimal saving by roughly 1.5 percent of GDP. In addition, population aging presents potential economic and budgetary challenges that are not included in the Elmendorf-Sheiner analysis (such as a possible damping effect on economic dynamism) or CBO's budget projections (such as a rising demand for federal support for eldercare). In these ways, the aging of the population provides a further justification for increasing national savings.

Providing Fiscal Space for Future Developments

Restraining debt growth also increases the government's capacity to borrow more when budgetary needs arise without causing interest rates to spike upward. Such fiscal space enables the government to respond more vigorously when some of today's uncertainties turn into tomorrow's problems.

One important use of fiscal space is countercyclical policy. The net decline in US interest rates over the past few decades has limited the Federal Reserve's room for cutting the federal funds interest rate when economic activity is weak—making tax reductions and spending increases especially useful. During and after the global financial crisis of 2007 to 2009, the government undertook fiscal stimulus equal to 11 percent of pre-crisis annual GDP, and after the COVID pandemic hit in 2020, the government undertook fiscal stimulus equal to 23 percent of pre-crisis GDP (Dynan and Elmendorf 2024). In addition, lower output and employment during those periods had the well-known "automatic stabilizer" effect of generating fiscal stimulus through lower taxes and higher spending without legislation. All told, debt held by the public jumped from 35 percent of GDP in 2007 to 70 percent in 2012 and from 79 percent of GDP in 2019 to 97 percent in 2021, for an increase of more than 60 percent during just seven extraordinary years. To use fiscal space by deliberately increasing debt (or just allowing automatic increases in debt) during and after recessions, fiscal space needs to be created at other points of business cycles.

In addition, fiscal space is useful if economic output falls short of expectations for prolonged periods rather than just during business-cycle downturns. Ball, Elmendorf, and Mankiw (1998) described deficits as a "gamble" that may have a desirable outcome if an economy prospers but an undesirable outcome if it does not, because increasing taxes and cutting spending to reduce debt has a larger welfare cost for people when an economy is not prospering. Indeed, people's desire

to protect themselves from the effects of slow economic growth can be seen in their acceptance of returns on federal debt that are lower than average returns on riskier private capital. Restraining debt growth during strong economic periods reduces the likelihood of needing to restrain it during weak ones.

Fiscal space also increases the government's capacity for responding to other challenges that can arise. Identifying all of the risks the country faces, much less quantifying those risks, would be difficult. But maintaining fiscal space gives future policymakers more scope to take action and protect people when problems develop.[4]

Reducing the Risk of a Fiscal Crisis

Yet another effect of restraining debt growth is reducing the risk of a fiscal crisis. A fiscal crisis occurs when investors become unwilling to hold a government's debt at typical interest rates, generally because government borrowing surges to respond to an emerging need or because investors lose confidence in how policymakers are handling government finances. The resulting upward spike in interest rates sharply increases the government's interest payments—an effect that is larger when debt is larger. The increase in payments worsens the outlook for debt, which can drive interest rates still higher, causing a path for debt that may have appeared sustainable to become quickly unsustainable. In addition, the runup in interest rates usually reverberates through a country's financial system, disrupting the flow of credit and impeding economic activity.

Without the ability to borrow as before, a government experiencing a fiscal crisis is forced to raise taxes and cut spending quickly and significantly to reduce issuance of new debt, default on outstanding debt that has reached maturity and cannot be refinanced at manageable interest rates, or both. Sudden and sizable changes in tax and spending policies are often painful for those who are affected directly, and they can reduce output, income, and employment broadly. Moreover, developing and implementing severe policy changes puts great strains on a country's political system.

Given the central role of US Treasury securities in global finance, a fiscal crisis in the United States would damage not only the US economy but also economies around the world. The damage would reduce Americans' employment and income directly, and it would diminish over time the global position of US financial markets and the US dollar, which would harm Americans further. In response to a fiscal crisis, the US government could choose to default on some of its outstanding debt—an outcome that may seem inconceivable, but effectively occurred in the 1930s when the US left the gold standard (Edwards 2018). The government also could choose to monetize some of its outstanding debt, with the effect of much higher inflation,

---

[4]Certain policy changes would help to insulate the government's finances from specific developments, such as indexing the eligibility age for Social Security to average life expectancy. However, that sort of change would not provide insurance against a broad array of possible developments, as a lower debt would.

as discussed above. More likely, in our view, is that the government would raise taxes and cut spending sharply.

Assessing the probability of a US fiscal crisis at different levels of debt is difficult. One approach is to examine periods of high government debt in other countries, as in this journal in Reinhart, Reinhart, and Rogoff (2012), among others. However, making strong inferences is difficult because sample sizes are small and countries differ in many respects. Reinhart et al. identified only 26 episodes where gross government debt exceeded 90 percent of GDP in advanced economies during the past two centuries.[5] Moreover, countries' situations differ not only in the amount of debt but also in the size of deficits, the share of debt in externally controlled currencies, the structure of financial markets and institutions, the expected path of debt, and factors that affect the malleability of that path, including the capacity of the political system to address budget issues constructively. As a result, debt levels that may precipitate crises in some places and times can be readily accommodated in others.

Consider the example of Greece, which experienced a very serious fiscal crisis beginning in late 2009 and lasting for many years. At the start, gross government debt equaled about 140 percent of GDP—a high figure, but not as high as in Japan, for instance, where gross government debt now exceeds 200 percent of GDP and no fiscal crisis has emerged (as discussed in this symposium by Chien, Du, and Lustig). Contributing factors to the crisis in Greece included a budget deficit close to 13 percent of GDP, a current account deficit that was almost as large, government debt denominated in euros, widespread tax evasion, and significant revisions to budget statistics that reinforced concerns about the quality of publicly available data. The crisis and the sharp fiscal austerity adopted in response caused Greek GDP to fall dramatically and recover very slowly. The result was prolonged hardship—and a marked rise in the debt-GDP ratio over the following decade, as the fiscal consolidation was offset by the weakness of GDP.[6]

Consider also the example of the United Kingdom, which experienced a far more limited fiscal crisis in 2022. Gross government debt equaled about 100 percent of GDP, and the government of Prime Minister Liz Truss announced new tax and spending proposals that investors feared would substantially increase debt. Concern about the government's fiscal management was reinforced by its decision to skip the usual review by the independent UK Office for Budget Responsibility. Yields on 30-year British bonds ("gilts") jumped 140 basis points over three days, with an initial increase in yields spurring bond sales by leveraged financial intermediaries that further pushed up yields, and the value of the British pound relative to other currencies fell sharply. To stabilize financial markets and limit spillovers to the

---

[5] Gross government debt is a broader measure than debt held by the public, which is the usual focus of analysis for the United States and is the focus in this paper. At the end of fiscal year 2024, US gross debt was 122 percent of GDP, and US debt held by the public was 98 percent of GDP (CBO 2025a).

[6] For more on the Greek crisis, see Lane (2012, in this journal) and Thomsen (2019). Debt data in this paragraph and the following one are from the Organisation for Economic Co-operation and Development (OECD 2024).

nonfinancial economy, the Bank of England implemented an explicitly temporary purchase of gilts, but only after the government retracted most of the proposals and Truss resigned did yields normalize.[7]

Four lessons of the Greek and British experiences seem relevant for the United States. First, investors' perspectives can change very quickly, as UK financial conditions seemed calm prior to the budget announcement but then deteriorated abruptly. Second, higher debt and deficits increase vulnerability to shifts in investors' perspectives; if those governments had been borrowing less, investor sentiment might have been less fragile, and the budgetary impact of higher interest rates would have been smaller. Third, the ability to adjust policy swiftly is crucial in limiting damage from a crisis; the Greek government faced such a large budgetary imbalance that changing course took time (and assistance from international organizations), whereas the UK government could withdraw proposed new policies nearly overnight. Fourth, faltering confidence in the judgment and competence of government leaders, as occurred in both the Greek and British cases, can precipitate and worsen a fiscal crisis.

Even with those lessons in hand, though, the probability of a US fiscal crisis in the near term is unclear. Current interest rates on US Treasury securities suggest that investors are not very worried about the US fiscal outlook. As we write, the yields on ten-year Treasury notes and ten-year inflation-indexed Treasury securities are in the higher ends of their ranges over the past two decades. However, that fact primarily reflects the exceptionally low interest rates of the 2010s and very early 2020s. Yields on ten-year nominal and inflation-indexed securities are no higher now than they were on the eve of the global financial crisis, when the debt-GDP ratio was only about one-third what it is now. Since 2007, financial markets have absorbed additional US government debt equal to nearly two-thirds of the country's current GDP with no net increase in the required return on that debt.

However, there is significant vulnerability to adverse developments. One source of vulnerability is the unprecedented nature of the current fiscal situation. The United States has almost no experience with federal debt larger than GDP, and it has no experience with sustained federal deficits close to 6 percent of GDP apart from wars and economic downturns—but those outcomes are now upon us. Another source of vulnerability is that policymakers have struggled in recent years to accomplish basic fiscal actions such as raising the statutory limitation on government borrowing and funding government operations on a timely basis. Perhaps with those figures and events in mind, participants in US financial markets reacted to the sharp changes in US trade policy in April 2025 by pushing up yields on Treasury securities and pushing down the value of the US dollar—in contrast with the typical movements in those markets during previous periods of global uncertainty. Some observers asserted that investors were losing trust in US governance and in the United States as a reliable partner to other countries.

---

[7]For more on the UK experience, see Chen and Kemp (2023) and Wilkins (2024).

A further source of vulnerability arises from the maturity structure of Treasury debt. Debt equal to nearly one-third of GDP is maturing within a year and needing to be refinanced; the corresponding figures before the global financial crisis and before the pandemic were roughly one-tenth and one-fifth, respectively (Slok, Shah, and Galwankar 2025). With so much short-term debt, shifts in investor sentiment can quickly change the government's interest payments by a substantial amount.[8]

An additional vulnerability stems from developments in the market for trading Treasury securities that are reducing the liquidity of that market. With less liquidity, a fiscal crisis could be more damaging (in this journal, Duffie 2025).

Indeed, increases in Treasury yields from the onset of the pandemic through at least mid-2024 were concentrated on days with unfavorable news about the fiscal outlook, and unfavorable fiscal news had a larger effect on yields during this period than it had earlier (Gómez-Cram, Kung, and Lustig 2024). Given the growing size of debt and deficits, greater salience of fiscal news and greater sensitivity of interest rates to such news is not surprising—and those changes increase the chance of a fiscal crisis.

All told, the risk of a near-term fiscal crisis is uncertain, but restraining debt growth would decrease that risk in multiple ways: It would leave more fiscal space to respond to emerging needs, help to reassure investors that policymakers are managing the government's finances appropriately, and reduce the impact of an interest-rate increase on the deficit (which would lower the chance of a self-reinforcing cycle of rate increases).

Broadening Policy Options by Acting Sooner

Deciding sooner rather than later how to restrain the growth of government debt also would broaden the policy options for doing so.

First, deciding sooner allows for smaller policy changes and more gradual implementation of those changes, which can enhance economic efficiency. For example, a smaller increase in tax rates applied sooner would distort behavior less than a larger increase in tax rates applied later.

Second, deciding sooner enables policymakers to adjust the impact of tax and spending changes across more birth cohorts than if they decide later. For example, discussions of changes to Social Security usually involve broad agreement that people who are already collecting benefits (or are close to doing so) should not face reductions, because they cannot adjust their work or saving behavior in response.[9]

---

[8]By contrast, foreign ownership of Treasury debt is not especially concerning. One potential trigger for a fiscal crisis would be a decision by foreign holders to sell large amounts of debt in a short time, motivated perhaps by geopolitical considerations. However, foreign holdings of Treasury debt have fallen to about 25 percent of the total today from nearly 35 percent a dozen years ago, and holdings of Treasury debt by foreign official institutions have declined in absolute terms over that period (Slok, Shah, and Galwankar 2025).

[9]Delay in addressing the unsustainable rise in debt tends to help older Americans. For example, delaying by ten years the increases in income tax rates that would be sufficient to stabilize the debt-GDP ratio has been estimated to *raise* the lifetime consumption of people in the lowest income tercile who were

Third, deciding sooner allows policymakers to implement tax increases or spending reductions when economic activity is relatively strong and fiscal stimulus is not needed to maintain full employment. By contrast, implementing such measures when economic activity is relatively weak would have larger short-term costs and might not even enhance fiscal sustainability (Summers 2014; Auerbach and Gorodnichenko 2017).

Putting the Pieces Together

Consistent with the preceding discussion, Reinhart and Rogoff (2009) showed that countries with gross public debt exceeding 90 percent of output have experienced notably slower growth than countries with less debt, Reinhart, Reinhart, and Rogoff (2012, in this journal) pointed to various researchers who have demonstrated a causal link from debt to growth, and Woo and Kumar (2015) showed that economies with higher debt-GDP ratios experience slower subsequent growth. However, this body of evidence does not point to a specific debt-GDP ratio that is optimal or a specific ratio that should not be exceeded.

We are left with a qualitative conclusion regarding the timing of US policy changes. Even if investors remain relatively unconcerned about the fiscal outlook, debt and deficits are at exceptionally high levels, and experience shows that confidence in a government's fiscal management can deteriorate quickly. Moreover, changing policies sooner rather than later would not only reduce the risk of a fiscal crisis but also would raise national savings, provide fiscal space for future developments, and broaden policy options. These effects have become more notable during the past several years because interest rates have risen and debt has increased sharply. No one can really know what might happen if US federal debt rose into uncharted territory well above 100 percent of GDP.

In What Ways Might US Policies Be Changed?

Suppose, for illustration, that US policymakers aimed to hold the federal debt-GDP ratio at its current value of roughly 100 percent during the next 30 years. Suppose as well that CBO's demographic and economic projections turned out to be correct. Then, to meet the goal, policymakers would need to adhere to the current-law discipline underlying CBO's projections *and* implement tax increases and spending cuts totaling 1.6 percent of GDP on average over those years (CBO 2025d)—or adopt other combinations of policies with the same budgetary effects. GDP now equals about $30 trillion and is projected by CBO to equal about $375 trillion over the coming decade, so 1.6 percent of GDP amounts to nearly $500 billion per year now and about $6 trillion over the coming decade.

---

born in the 1940s and 1950s by 0.3 percent and have little effect on their hours worked, and to *lower* the lifetime consumption of their counterparts born in the 2000s and 2010s by 1.1 percent while reducing their hours worked (CBO 2022).

Policy changes of that magnitude are quite possible, of course. Auerbach and Yagan (2024) estimated that, between 1986 and 1988, policymakers adopted deficit reduction equal to 2 percent of GDP. But examples of tax and spending changes illustrate how substantial the required policy changes would be. On the revenue side, CBO (2024b) estimated that raising all individual income tax rates (except the rates on long-term capital gains and dividends) by 1 percentage point would reduce deficits by roughly 0.3 percent of GDP and that eliminating the deductibility of state and local taxes for federal income taxes would reduce deficits by roughly 0.5 percent of GDP over the next decade.[10] On the spending side, CBO (2024b) estimated that raising premiums for Medicare Part B (which covers physicians' and other outpatient services) by 40 percent would reduce deficits by roughly 0.2 percent of GDP and that raising the full retirement age for Social Security gradually from 67 to 70 would reduce deficits by roughly 0.6 percent of GDP 30 years ahead.

With large policy changes looming at some point, the choices among potential changes will be especially important. Many considerations matter, including preferences about the role of the federal government, evolving national and global conditions, possibilities for increasing economic growth, and the desired distribution of government benefits and burdens within and across generations. Much has been written about those topics; for just two recent examples, see Grand Bargain Committee (2024) and Gale, Berlin, and Thorpe (2025). Here, we highlight just two key issues.

First, the legacy of current generations to future ones will include more than just the federal debt—it will also include physical capital, educational attainment, national security, social cohesion, the climate and environment, and more. Therefore, policy changes that reduced economic growth or adversely affected those other aspects of wellbeing would conflict with the core rationale for restraining debt. Yet, some significant elements of government taxes and spending affect economic growth. For example, federal outlays for infrastructure and for research and development are investments with substantial returns (Ramey 2021; Fieldhouse and Mertens 2023). In addition, some government programs providing resources to lower-income families improve earnings potential and other aspects of life trajectories; for examples of the expanding body of research in this area, see Aizer, Hoynes, and Lleras-Muney (2022) in this journal, as well as Bailey et al. (2024), Brown, Kowalski, and Lurie (2020), Hendren and Sprung-Keyser (2020), Hoynes, Schanzenbach, and Almond (2016), and Miller and Wherry (2019). Similarly, some elements of tax policy have notable effects on work, saving, and investment (Chodorow-Reich, Zidar, and Zwick 2024; Goodman et al. 2025).

Second, socioeconomic inequality has increased and income mobility has declined in the United States over the past several decades. For example, a report from the National Academies of Sciences, Engineering, and Medicine (2015) estimated that life expectancy was increasing across much of the income distribution,

---

[10]The enactment of significant tax legislation in 2025 has altered the deficit effect of those policy changes to some degree, but CBO has not provided updated estimates at this time.

but not in the lowest two quintiles (among much other research on these topics, see also Case and Deaton 2023; Chetty et al. 2017). Federal taxes and transfer payments make the distribution of after-tax, after-transfer income notably less unequal than the distribution of pre-tax, pre-transfer income (CBO 2024a, p. 31). Therefore, policy changes that reduced the progressivity of taxes and transfers could significantly reinforce the observed rise in inequality, and policy changes that increased the progressivity of taxes and transfers could provide a significant counterweight.

How Might the US Political Process Respond to Rising Debt?

A meaningful shift toward a sustainable debt path in the United States seems unlikely to be accomplished soon. In the current political climate, even routine fiscal actions, like raising the debt ceiling or enacting appropriations to keep the government running, have been difficult to achieve—so reaching bipartisan agreement on the much more challenging actions needed to lower the debt trajectory does not seem politically feasible. Yet, neither party has much incentive to restrain federal borrowing on its own, because most significant tax increases and spending cuts would generate popular resistance, and because the resulting fiscal space might be used by the other party to advance its own priorities when it next gains political control.

Moreover, although polls show that many people are concerned about rising debt, the dire consequences of debt that have been predicted by public figures for many years have not materialized, even though the debt-GDP ratio has nearly tripled since 2007. One partial exception is the inflation surge of 2021 and 2022 that has been linked to fiscal stimulus by some analysts (for example, Dynan and Elmendorf 2024). But the most important sources of economic disruption in the United States in the twenty-first century have been financial meltdowns, disease vulnerability, technological change, and globalization, so people's attention might reasonably be more focused on those issues than on government debt.

Looking ahead, the spur for political action to lower the path of US federal debt is not clear. Perhaps the depletions of the trust funds for Social Security and Medicare hospital insurance—now projected for the mid-2030s—will induce changes in spending or dedicated revenue for those programs. Perhaps rising federal interest payments—now headed to their highest level relative to GDP in at least 60 years—will heighten public concern about the fiscal situation. Perhaps interest rates will rise sharply at some point because of high and rising debt. If the increase in interest rates is so sudden and sizable as to constitute a fiscal crisis, policymakers would have little choice but to raise taxes and cut spending sharply. But prudent policymakers would prefer to put US fiscal policy on a sustainable path well before such a crisis occurs.

References

Acalin, Julien, and Laurence M. Ball. Forthcoming. "Did the US Really Grow Out of Its World War II Debt?" *American Economic Journal: Macroeconomics.*

Aizer, Anna, Hilary Hoynes, and Adriana Lleras-Muney. 2022. "Children and the US Social Safety Net: Balancing Disincentives for Adults and Benefits for Children." *Journal of Economic Perspectives* 36 (2): 149–74.

Auclert, Adrien, Hannes Malmberg, Matthew Rognlie, and Ludwig Straub. 2025. "The Race between Asset Supply and Asset Demand." Unpublished.

Auerbach, Alan J., and Yuriy Gorodnichenko. 2017. "Fiscal Stimulus and Fiscal Sustainability." In *Proceedings of the Jackson Hole Economic Policy Symposium*, 217–70. Federal Reserve Bank of Kansas City.

Auerbach, Alan J., and Danny Yagan. 2024. "Robust Fiscal Stabilization." *Brookings Papers on Economic Activity* 55 (2): 239–93.

Bailey, Martha J., Hilary Hoynes, Maya Rossin-Slater, and Reed Walker. 2024. "Is the Social Safety Net a Long-Term Investment? Large-Scale Evidence from the Food Stamps Program." *Review of Economic Studies* 91 (3): 1291–1330.

Ball, Laurence, Douglas W. Elmendorf, and N. Gregory Mankiw. 1998. "The Deficit Gamble." *Journal of Money, Credit and Banking* 30 (4): 699–720.

Ball, Laurence, and N. Gregory Mankiw. 1995. "What Do Budget Deficits Do?" In *Proceedings of the Jackson Hole Economic Policy Symposium*, 95–119. Federal Reserve Bank of Kansas City.

Ball, Laurence, and N. Gregory Mankiw. 2023. "Market Power in Neoclassical Growth Models." *Review of Economic Studies* 90 (2): 572–96.

Blanchard, Olivier. 2019. "Public Debt and Low Interest Rates." *American Economic Review* 109 (4): 1197–229.

Blanchard, Olivier. 2023. *Fiscal Policy under Low Interest Rates.* MIT Press.

Brown, David W., Amanda E. Kowalski, and Ithai Z. Lurie. 2020. "Long-Term Impacts of Childhood Medicaid Expansions on Outcomes in Adulthood." *Review of Economic Studies* 87 (2): 792–821.

Case, Anne and Angus Deaton. 2023. "Accounting for the Widening Mortality Gap between American Adults with and without a BA." *Brookings Papers on Economic Activity* 54 (2): 1–44.

Chen, Ruo, and Esti Kemp. 2023. "Putting Out the NBFire: Lessons from the UK's Liability-Driven Investment (LDI) Crisis." International Monetary Fund Working Paper 23/210.

Chetty, Raj, David Grusky, Maximilian Hell, Nathaniel Hendren, Robert Manduca, and Jimmy Narang. 2017. "The Fading American Dream: Trends in Absolute Income Mobility since 1940." *Science* 356 (6336): 398–406.

Chodorow-Reich, Gabriel, Owen Zidar, and Eric Zwick. 2024. "Lessons from the Biggest Business Tax Cut in US History." *Journal of Economic Perspectives* 38 (3): 61–88.

Cochrane, John H. 2023. *The Fiscal Theory of the Price Level.* Princeton University Press.

CBO (Congressional Budget Office). 2007–2025. *Budget and Economic Data (December 2007, September 2013, January 2020, January 2025, and March 2025).* Congressional Budget Office. https://www.cbo.gov/data/budget-economic-data (accessed April 17, 2025).

CBO (Congressional Budget Office). 2022. *The Economic Effects of Waiting to Stabilize Federal Debt.* Congressional Budget Office.

CBO (Congressional Budget Office). 2024a. *Trends in the Distribution of Household Income from 1979 to 2021.* Congressional Budget Office.

CBO (Congressional Budget Office). 2024b. *Options for Reducing the Deficit: 2025 to 2034.* Congressional Budget Office.

CBO (Congressional Budget Office). 2025a. *The Budget and Economic Outlook: 2025 to 2035.* Congressional Budget Office.

CBO (Congressional Budget Office). 2025b. *How Changes in Economic Conditions Might Affect the Federal Budget: 2025 to 2035.* Congressional Budget Office.

CBO (Congressional Budget Office). 2025c. *The Long-Term Budget Outlook: 2025 to 2055.* Congressional Budget Office.

CBO (Congressional Budget Office). 2025d. *The Long-Term Budget Outlook under Alternative Scenarios for the Economy and the Budget.* Congressional Budget Office.

CBO (Congressional Budget Office). 2025e. "Effects on Deficits and the Debt of Public Law 119-21 and of Making Certain Tax Policies in the Act Permanent." Letter to the Honorable Jeff Merkley.

Congressional Budget Office.

CBO (Congressional Budget Office). 2025f. "An Update about CBO's Projections of the Budgetary Effects of Tariffs." Congressional Budget Office (blog), August 22. https://www.cbo.gov/publication/61697.

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

Debrun, Xavier, Jonathan D. Ostry, Tim Willems, and Charles Wyplosz. 2019. "Debt Sustainability." In *Sovereign Debt: A Guide for Economists and Practitioners*, edited by S. Ali Abbas, Alex Pienkowski, and Kenneth Rogoff, 151–91. Oxford University Press.

Duffie, Darrell. 2025. "How US Treasuries Can Remain the World's Safe Haven." *Journal of Economic Perspectives* 39 (2): 195–214.

Dynan, Karen. 2023. "High and Rising US Federal Debt: Causes and Implications." In *Building a More Resilient US Economy*, edited by Melissa S. Kearney, Justin Schardin, and Luke Pardue, 17–43. Aspen Institute.

Dynan, Karen, and Douglas Elmendorf. 2024. "Fiscal Policy and the Pandemic-Era Surge in US Inflation: Lessons for the Future." Peterson Institute for International Economics Working Paper 24-22.

Edwards, Sebastian. 2018. *American Default: The Untold Story of FDR, the Surpreme Court, and the Battle over Gold*. Princeton University Press.

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.

Elmendorf, Douglas W., and Louise M. Sheiner. 2017. "Federal Budget Policy with an Aging Population and Persistently Low Interest Rates." *Journal of Economic Perspectives* 31 (3): 175–94.

Fieldhouse, Andrew J., and Karel Mertens. 2023. "The Returns to Government R&D: Evidence from U.S. Appropriations Shocks." Federal Reserve Bank of Dallas Working Paper 2305.

Gale, William, Ian Berlin, and Sam Thorpe. 2025. "Fiscal Consolidation: Lessons for the US." Preprint, SSRN. http://dx.doi.org/10.2139/ssrn.5467589.

Gamber, Edward N. 2020. "The Historical Decline in Real Interest Rates and Its Implications for CBO's Projections." Congressional Budget Office Working Paper 2020-09.

Gómez-Cram, Roberto, Howard Kung, and Hanno Lustig. 2024. "Government Debt in Mature Economies: Safe or Risky?" In *Proceedings of the Jackson Hole Economic Policy Symposium*, 113–85. Federal Reserve Bank of Kansas City.

Goodman, Lucas, Katherine Lim, Bruce Sacerdote, and Andrew Whitten. 2025. "How Do Business Owners Respond to a Tax Cut? Examining the 199A Deduction for Pass-Through Firms." *Journal of Public Economics* 242: 105293.

Grand Bargain Committee. 2024. *Toward a Potential Grand Bargain for the Nation*. Center for Collaborative Democracy and Bipartisan Policy Center.

Gust, Christopher, and Arsenios Skaperdas. 2024. "Government Debt, Limited Foresight and Longer-Term Interest Rates." FEDS Working Paper 2024-027.

Hendren, Nathaniel, and Ben Sprung-Keyser. 2020. "A Unified Welfare Analysis of Government Policies." *Quarterly Journal of Economics* 135 (3): 1209–318.

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

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

Leeper, Eric M. 1991. "Equilibria under 'Active' and 'Passive' Monetary and Fiscal Policies." *Journal of Monetary Economics* 27 (1): 129–47.

Leeper, Eric M. 2023. "Fiscal Dominance: How Worried Should We Be?" Preprint, SSRN. http://dx.doi.org/10.2139/ssrn.4665190.

Mian, Atif, Ludwig Straub, and Amir Sufi. Forthcoming. "A Goldilocks Theory of Fiscal Deficits." *American Economic Review*.

Miller, Sarah, and Laura R. Wherry. 2019. "The Long-Term Effects of Early Life Medicaid Coverage." *Journal of Human Resources* 54 (3): 785–824.

National Academies of Sciences, Engineering, and Medicine. 2015. *The Growing Gap in Life Expectancy by Income: Implications for Federal Programs and Policy Responses*. National Academies Press.

Neveu, Andre R., and Jeffrey Schafer. 2024. "Revisiting the Relationship between Debt and Long-Term Interest Rates." Congressional Budget Office Working Paper 2024-05.

OECD. 2024. *General Government Debt*. OECD. https://www.oecd.org/en/data/indicators/general-government-debt.html (accessed December 9, 2024).

Pierce, Brooks. 2021. "How Changes in the Distribution of Earnings Affect the Federal Deficit."

Congressional Budget Office Working Paper 2021-21.

Plante, Michael, Alexander W. Richter, and Sarah Zubairy. 2025. "Revisiting the Interest Rate Effects of Federal Debt." NBER Working Paper 34018.

Ramey, Valerie A. 2021. "The Macroeconomic Consequences of Infrastructure Investment." In *Economic Analysis and Infrastructure Investment*, edited by Edward L. Glaeser and James M. Poterba, 219–76. University of Chicago Press.

Reinhart, Carmen M., Jacob F. Kirkegaard, and M. Belen Sbrancia. 2011. "Financial Repression Redux." *Finance and Development* 48 (1): 22–26.

Reinhart, Carmen M., Vincent R. Reinhart, and Kenneth S. Rogoff. 2012. "Public Debt Overhangs: Advanced-Economy Episodes since 1800." *Journal of Economic Perspectives* 26 (3): 69–86.

Reinhart, Carmen M., and Kenneth S. Rogoff. 2009. *This Time Is Different: Eight Centuries of Financial Folly*. Princeton University Press.

Slok, Torsten, Rajvi Shah, and Shruti Galwankar. 2025. "The US Fiscal Outlook, Long-Term Interest Rates, Treasury Auctions, and the Dollar." Unpublished.

Stantcheva, Stefanie. 2024. "Why Do We Dislike Inflation?" *Brookings Papers on Economic Activity* 55 (1): 1–46.

Summers, Lawrence H. 2014. "US Economic Prospects: Secular Stagnation, Hysteresis, and the Zero Lower Bound." *Business Economics* 49 (2): 65–73.

Summers, Lawrence H. 2015. "Demand Side Secular Stagnation." *American Economic Review* 105 (5): 60–65.

Thomsen, Poul M. 2019. "The IMF and the Greek Crisis: Myths and Realities." Speech, London School of Economics, September 30, 2019. https://www.imf.org/en/News/Articles/2019/1%1/sp093019-The-IMF-and-the-Greek-Crisis-Myths-and-Realities.

Wilkins, Carolyn A. 2024. "Financial Stability and Monetary Policy: Lessons from the UK's LDI Crisis." Princeton University Griswold Center for Economic Policy Studies Working Paper 336.

Woo, Jaejoon, and Manmohan S. Kumar. 2015. "Public Debt and Growth." *Economica* 82 (328): 705–39.

Sovereign Debt and Fiscal Integration in the European Union

Zsolt Darvas, Lennard Welslau, and
Jeromin Zettelmeyer

Macroeconomic integration in the European Union occupies a halfway house. Most of the European Union shares a common currency and a single monetary policy, but fiscal policy remains national. The EU budget is small (just over 1 percent of GDP, excluding temporary funds) and must be balanced. Though EU debt has recently increased after a one-off attempt to use common borrowing as a stabilization instrument during the pandemic, debt markets are nationally fragmented. Financial systems also remain largely national, though banking supervision has become centralized since the euro crisis of 2010–2012.

In this setting, EU countries face several debt-related worries. First, each must worry about its own country-level debt sustainability. In addition, banks in the largest countries remain disproportionately exposed to debt issued by their home sovereigns. This pattern holds especially true in Belgium, France, Italy, Portugal, and Spain. Finally, each member of the euro area must in principle worry not just about debt risks at home but also about debt risks in other euro countries, which could spread by threatening the integrity of the single currency.

During the 2010–2012 euro crisis, the combination of these interlocking vulnerabilities almost extinguished the euro. Notably, however, the pandemic did

■ *Zsolt Darvas is Senior Fellow, Bruegel, Brussels, Belgium, and Senior Research Fellow, Corvinus University of Budapest, Budapest, Hungary. Lennard Welslau is a PhD Fellow, University of Copenhagen and Danmarks Nationalbank, Copenhagen, Denmark. Jeromin Zettelmeyer is Director of Bruegel, Brussels, Belgium, and Research Fellow, Centre for Economic Policy Research, London, United Kingdom. Their email addresses are zsolt.darvas@bruegel.org, lennard.welslau@gmail.com, and jeromin.zettelmeyer@bruegel.org.*

not have a similar effect, despite much higher debt levels. Also, monetary tightening during 2022–2023 did not lead to an EU debt crisis, as many had feared. This better-than-feared performance is often attributed to institutional improvements since the euro crisis, as well as European Central Bank policies geared to preventing a reemergence of sovereign debt risks.

The purpose of this paper is to take stock of EU sovereign debt risks in light of higher debt, higher interest rates, continuing fiscal fragmentation, but also better policies and institutions. We begin with some historical and institutional background. We then examine to what extent the euro area's institutional reforms have reduced the risks of what is commonly called the "sovereign-bank doom loop." Next, we present results from a debt-sustainability analysis for EU countries. For comparison purposes, we undertake the same analysis for the United States and the United Kingdom.[1] We finally turn to two policy debates: whether European fiscal rules require further reform and how common euro area public debt instruments might be expanded.

An Imperfect Macroeconomic Union

The European Economic and Monetary Union (EMU), established in the 1992 Treaty on the European Union in Maastricht, is one institution within the broader European Union. It involves a single currency, the euro (begun in 1999 and passing into widespread use in 2002). It is governed by an independent monetary authority, the European Central Bank (ECB). In contrast, fiscal and (initially) financial sector policies were largely left in the hands of national EU governments. The European Union does not have its own tax-raising powers, nor a macroeconomic stabilization instrument. The EU budget has focused on agricultural subsidies and "cohesion spending" aimed at supporting poorer regions, and has remained limited in size: currently, about 1.1 percent of EU GDP.[2]

With a single currency, the Economic and Monetary Union eliminated national-level monetary policy, as well as national-level devaluation and unanticipated inflation as a soft fiscal adjustment mechanism, and also eliminated national central banks as lenders of last resorts to their governments. But in contrast to federations such as the United States, EMU did not create fiscal mechanisms for macroeconomic stabilization or risk-sharing across member states. In these ways, EMU both required more from national fiscal policy and removed instruments to

---

[1] In this article, we use the terms "Europe" and "the European Union" roughly interchangeably. This said, many of the problems we ascribe to "Europe" also apply to European countries that are not members of the European Union. Furthermore, this paper considers sustainability in the United Kingdom in addition to that of EU members.

[2] The EU budget was supplemented by a temporary instrument during the COVID-19 pandemic—NextGenerationEU (NGEU), in which joint borrowing was used to finance expenditures amounting to 2.7 percent of the European Union's 2022 GDP over 2021–2026. However, the debt created by this instrument is intended to be fully repaid (rather than refinanced) over time, and no follow-up instrument is planned.

manage debt crises without an outright default or exit from the euro area. To create the fiscal space to conduct national countercyclical policy without endangering fiscal solvency, requiring the help of other EU members, or exerting pressure on the European Central Bank to monetize national public debt, the 1992 Maastricht treaty requires EU members to maintain or reduce their public debt and deficit levels to below 60 percent and 3 percent of GDP, respectively.

Figure 1 shows the distribution of sovereign debt as a share of GDP in advanced economies over the last 30 years. Debt/GDP ratios of EU members generally declined sharply in the second half of the 1990s, particularly in the top half of the debt distribution (Figure 1, panel A), unlike those of their non-EU peers (Figure 1, panel B). Following euro entry, debt levels in the lower half of the EU debt distribution continued to decline but edged up in the top quartile. The global financial crisis of 2008–2009 unsurprisingly triggered sharp increases in sovereign debt both in the European Union and elsewhere. But the rise in debt in the European Union continued for much longer, until 2014. This pattern was a consequence of the 2010–2012 euro crisis, when Greece restructured its government debt despite international financial assistance that initially sought to avoid this; debt issued by Ireland, Portugal, and Cyprus lost access to global financial markets and required international assistance; and debt issued by Spain and Italy came very close to losing market access.

The failure of the Maastricht rules to ensure fiscal solvency and prevent a euro-threatening crisis had two main reasons. In some countries, including Greece and Portugal, the fiscal rules failed to create the intended fiscal buffers. But more fundamentally, the EMU architecture failed to take into account the interdependence of fiscal and financial sector solvency and liquidity. In the absence of EU-wide banking supervision and resolution mechanisms, the fiscal liability for banking crises remained national, while banking systems continued to cater mostly to domestic credit needs. This gave rise to a "sovereign-bank doom loop" or "diabolic loop" (Brunnermeier et al. 2016). Bank exposure to government bonds constituted an important channel through which the fear of government default dragged down the economy, while bank losses addressed by government bailouts strained public finances.

While the interdependence between sovereign debt and bank balance sheets can lead to a vicious circle in any country, euro-area countries were particularly vulnerable because of their lack of instruments to stop a crisis in market confidence (De Grauwe 2012). Sovereign-bank interdependence contributed to the spread of the crisis from countries with unsustainably high debt in the private (Ireland) or public (Greece) sectors to countries such as Italy and Spain, and complicated the management of the crisis.

Since the euro crisis peaked in 2011–2012, the architecture of Economic and Monetary Union has undergone three sets of reforms that attempt to address these weaknesses.[3]

[3] For a deeper treatment of the policy debates and evolution of the EMU architecture until 2021 in this journal, see Bilbiie, Monacelli, and Perotti (2021) and Lane (2021). For the new fiscal rules, see Darvas, Welslau, and Zettelmeyer (2024).

Figure 1
Sovereign Debt as a Share of GDP in Advanced Countries, 1994–2030

Panel A. EU countries
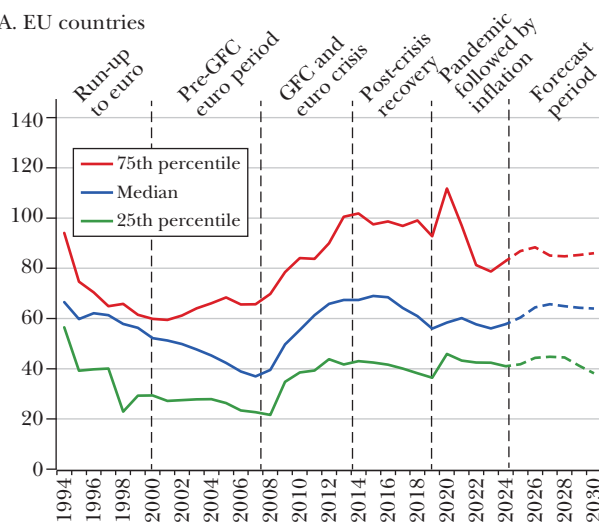
Panel B. Non-EU advanced countries
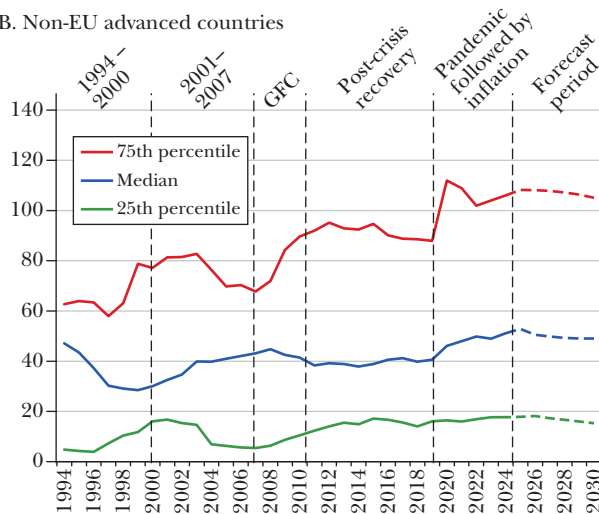
Source: IMF World Economic Outlook Database, April 2025.
Note: GFC is "global financial crisis." 2025–2030 values represent IMF forecasts. EU sample includes Austria, Belgium, Denmark, Finland, France, Germany, Greece, Italy, Netherlands, Portugal, Spain, and Sweden from 1994; Cyprus, Czech Republic, Estonia, Hungary, Ireland, Luxembourg, Malta, Poland, Slovak Republic, and Slovenia from 1995; Bulgaria, Croatia, Latvia, and Lithuania from 1998, and Romania from 2000. Non-EU sample includes Australia, Canada, Iceland, Japan, Korea, New Zealand, Norway, Singapore, Switzerland, and the United Kingdom from 1994; Taiwan Province of China from 1997; Israel from 2000; the United States from 2001; and Hong Kong SAR from 2004.

First, policy and institutional changes were introduced to prevent and mitigate sovereign illiquidity. After temporary lending programs supporting Greece, Ireland, and Portugal, a permanent institution, the European Stability Mechanism (ESM), was created in 2012 to provide conditional financial support to member states facing market pressure. The European Central Bank's Outright Monetary Transactions program announced in September 2012 offers unlimited secondary market purchases of sovereign bonds to countries in ESM programs. In 2020–2022, the Pandemic Emergency Purchase Programme (PEPP) showed that the European Central Bank was willing to deviate from its capital key in allocating bond purchases across its member states in an emergency. Finally, on July 21, 2022, the European Central Bank introduced a new instrument, the Transmission Protection Instrument, to "counter unwarranted, disorderly market dynamics" in bond yields of euro-area members.

The second set of reforms launched the European Banking Union, which centralized banking supervision under the Single Supervisory Mechanism in 2014. Banking resolution was also partially centralized through the Single Resolution Mechanism, which includes an industry-financed resolution fund. However, these reforms have fallen well short of creating a unified EU-level institution. Deposit insurance and most liquidation or resolution decisions remain national (Véron 2024).

Finally, European fiscal rules have undergone two sets of reforms: in 2011–2013, and again in 2024. The current system, passed in April 2024, requires all EU member states to regularly submit four-year "fiscal structural plans" that meet certain debt sustainability criteria. Countries with debt above 60 percent are required to implement a fiscal adjustment program over four years, but can obtain an extension to seven years if they submit public investment or structural reform plans that improve their growth or fiscal prospects and are endorsed by the EU council. This must ensure that, over the ten years following the adjustment period and assuming no changes in fiscal policy, the debt/GDP ratio is projected to decline with at least 70 percent probability, allowing for unexpected shocks. In addition, the budget deficit must remain below 3 percent of GDP during the ten-year post-adjustment period; backloading of fiscal adjustment is generally not permitted; and several "safeguards" apply that require a minimum speed of adjustment.[4]

The question addressed in the following sections is to what extent the three main sets of reforms have succeeded in reducing or containing sovereign debt risks in the European Union—both liquidity risks magnified by the "doom loop" between banks and sovereigns, and sovereign solvency risks.

---

[4]For countries with debt above 90 (60) percent of GDP, the debt ratio must decline by a minimum average amount of 1 (0.5) percent per year during the adjustment period. Countries with a structural budget deficit exceeding 1.5 percent of GDP must improve their fiscal position by 0.25–0.4 percent of GDP per year, depending on the length of the adjustment period. For countries with a headline deficit above 3 percent of GDP, a minimum annual fiscal adjustment of 0.5 percent of GDP is required.

Figure 2

Bank Exposure to the Respective Domestic Government in Selected Euro-Area Countries, 2001–2025 (in Percent of Equity Capital)
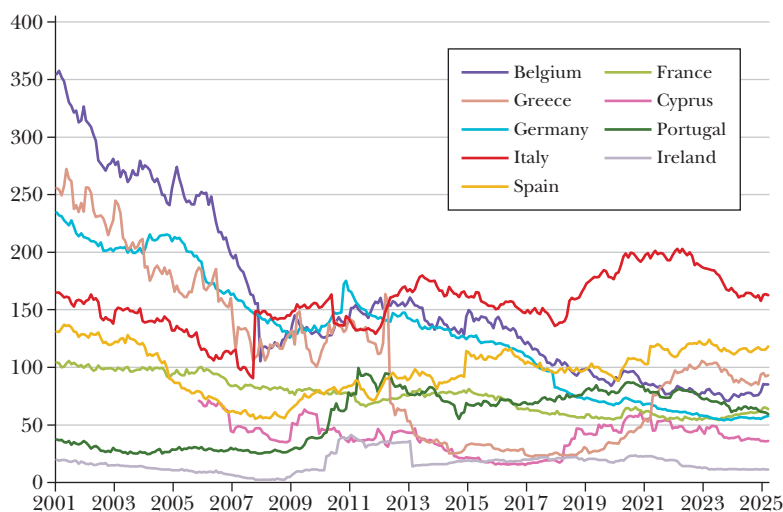
Source: ECB Balance Sheet Items (BSI) database.
Note: Both government securities holdings and loans to the government are considered. For data availability reasons, the chart uses equity capital rather than the more commonly used Tier 1 capital. Except for Germany, Tier 1 capital is lower than equity capital. Tier 1 capital refers to a bank's highest-quality capital, including common stock, retained earnings, and other disclosed reserves. It is defined in Article 25 of Regulation (EU) No. 575/2013, available at https://eur-lex.europa.eu/eli/reg/2013/575/oj/eng.

Fiscal Fragmentation and the "Doom Loop"

Following the euro crisis, bank exposures to their domestic sovereign came down in some euro area countries but remained high in France, Greece, Italy, and Spain (Figure 2). In these countries, as well as Belgium, aggregate exposures of banks to the domestic sovereign continues to exceed "Tier 1" core capital held by banks. Thus, the sovereign-bank nexus remained a potential channel for the transmission of shocks from governments to the banking sector at the time of the 2020 pandemic shock and remains so today. Yet, while high borrowing and the collapse in output during the pandemic led to a sharp rise in debt/GDP ratios in 2020 (Figure 1, panel A), this did not trigger a debt crisis.

A possible reason is that the reforms of 2012–2020 succeeded in weakening the doom loop by reducing the threat of sovereign illiquidity. However, the 2020 pandemic also led to an unprecedented, centralized fiscal crisis response called "Next Generation EU," in the form of EU-level common borrowing that was passed on to EU member states—particularly the fiscally weaker ones. Hence, the success of 2020–2021 may reflect a particular approach to crisis management that

unlike 2010–2012 crossed the line into fiscal mutualization. But this policy has been declared to be a one-off, motivated by the symmetric and exogenous nature of the pandemic shock, rather than a permanent improvement in the euro architecture.

A better test of the hypothesis that the architecture of the euro improved between 2012 and 2022 is to compare the market impact of early reforms to that of the most recent reform, the introduction of the European Central Bank's Transmission Protection Instrument (TPI) in 2022. If the pre-2022 reforms had a strong cumulative effect, then the incremental impact of the most recent reform should be far lower than that of the initial reforms.

To check this, we compare the impact of the 2022 Transmission Protection Instrument announcements to those of the 2012 Outright Monetary Transactions (OMT) announcements. As noted in the previous section, OMT offers unlimited secondary market purchases of sovereign bonds to countries undertaking an adjustment and reform program supervised by the European Stability Mechanism, while TPI commits to pushing back against "disorderly market dynamics" in bond yields without requiring such a program.

Figure 3, panels A and B, show the movements of sovereign and bank credit risk, measured by credit default swap spreads around the time of the 2012 Outright Monetary Transactions and 2022 Transmission Protection Instrument announcements, respectively.[5] In Figure 3, panel A, the first vertical line indicates July 26, 2012, when ECB President Draghi said that "the ECB is ready to do whatever it takes to preserve the euro"; the second, August 2, 2012, when he announced the ECB might undertake "outright open market operations"; and the third, September 6, 2012, when the European Central Bank announced the technical features of the OMT program. In Figure 3, panel B, the first vertical line indicates the June 15, 2022, ad-hoc Governing Council meeting of the European Central Bank that called for accelerated work on a new antifragmentation instrument, while the second vertical line indicates the July 21, 2022, meeting that approved the TPI.

Figure 3, panel A, shows that in Italy and Spain, spreads on both sovereign and bank credit default swap fell on or immediately after the days of the Outright Monetary Transactions announcements in 2012, settling at substantially lower levels in the weeks following the September announcement compared to those prevailing before the July announcement. Figure 3, panel B, indicates that the reactions to the Transmission Protection Instrument announcements in 2022 were different. For example, while sovereign credit default swap spreads fell on or soon after the two announcement dates in 2022, these falls merely reversed the increase in yields in the weeks ahead of the first announcement. In addition, the 2022 TPI announcements appear to have affected mainly sovereign spreads, with bank-related credit

---

[5] Credit default swaps are a financial derivative that works like an insurance contract. The party purchasing the swap makes regular payments, and if the financial security covered by the credit default swap stops making payments, then the party that purchased the credit default swap receives a payment. The price of credit default swaps—the "spread" as measured as a percentage of the notional amount being insured—reflects the probability of default.

Figure 3
Sovereign and Bank Credit Default Swap Spreads at the Time of the OMT and TPI Announcements

Panel A. July to September, 2012
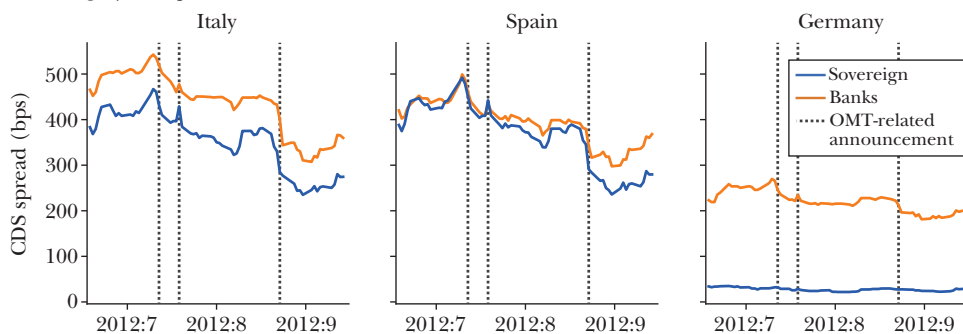
Panel B. June to August, 2022
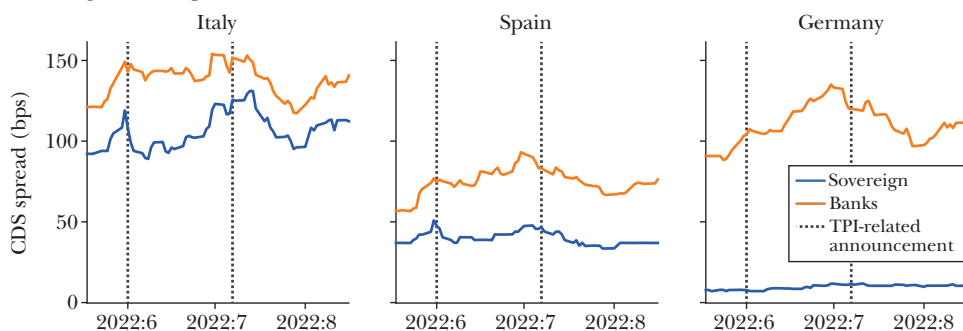
Source: Refinitiv.
Note: Vertical lines show the timing of OMT-related announcements (Figure 3, panel A) and TPI-related announcements (Figure 3, panel B), as explained in the text.

default swap spreads remaining largely unaffected after the June 15 announcement. The June event was more informative, as it followed an unexpected ad hoc ECB Governing Council meeting, whereas the July 2022 launch of the TPI was largely expected.[6]

---

[6] In a more formal test of the impact of the 2012 and 2022 ECB announcements on various credit default swap spreads, we estimate a regression model similar to that used by Altavilla, Giannone, and Lenza (2016), who analyzed the effects of the three OMT-related announcements in 2012 on sovereign bond yields, involving a regression of the five-year spreads on credit default swaps on a constant and a set of event dummies that take the value of one on the day of the announcement and the following day, thereby allowing for potential delayed market reactions. The results, shown in the Supplemental Appendix (Figure A1 and Table A1), confirm the visual impressions of Figure 3. That is, in Italy and Spain, both sovereign and bank CDS spreads experienced large declines during the 2012 OMT announcements. During the 2022 TPI announcements, sovereign spreads also fell in both countries—albeit by much less—but bank spreads moved minimally, or even increased.

In short, the impact of the Transmission Protection Instrument announce-ment in 2022 differed from that of the Outright Monetary Transactions in 2012 in two ways. First, the 2022 TPI announcement did not result in permanently lower sovereign spreads in Italy or Spain, unlike the 2012 OMT. Rather, its main effect appears to have been to offset, or "calm," spikes in the sovereign spreads that occurred immediately before the announcements. Second, it seems to have affected mainly sovereigns rather than both banks and sovereigns, suggesting a weaker sovereign-bank nexus in 2022 compared to 2012.

Debt Sustainability

Should high and rising sovereign debt levels in many European countries be a cause of concern?[7] Figure 1, panel A, does not look overly dramatic: the debt distribution is for the most part below its post-crisis peak, and the International Monetary Fund (IMF) expects it to remain roughly stable in the medium term. But this impression may be too sanguine for several reasons: it does not show the debt trajectories of individual countries that we may be worried about the most (in the upper-25th percentile of the debt distribution), it does not reflect uncertainty, and it considers only a relatively short horizon. This perspective may miss debt drivers that could push debt back up, including higher real interest rates since 2022, which take a while to make themselves felt as the debt stock is rolled over, and the impact of aging populations, both on the spending side and via lower potential growth.

This section addresses the question whether debt in European countries is sustainable, using the definition followed in most of the empirical literature and policy practice since the early 1990s (for references and a survey, see Willems and Zettelmeyer 2022). Debt is considered sustainable when the fiscal effort required to (eventually) stabilize the debt/GDP profile looks feasible.[8] If this is not the case, the only way to restore sustainability is a debt restructuring (or a surprise inflation, if debt is expressed in nominal currency units).

This definition implies a distinction between debt sustainability and the sustainability of current fiscal policy. *Fiscal policy* is unsustainable if the debt/GDP ratio explodes conditional on current and planned primary deficits. In contrast, *debt* is unsustainable if the debt/GDP ratio explodes under *any* feasible fiscal policy. Hence, the statement that debt is unsustainable is much stronger than the statement that fiscal policy is unsustainable. For example, as Figure 4

---

[7] This section, as well as the following sections on debt-stabilizing primary balances and fiscal reaction functions, draw on Darvas et al. (2025).

[8] Stabilization of the debt/GDP ratio is a stronger condition than ruling out government "Ponzi games" in which the value of the debt rises over time purely in the expectation that it can be resold in the future. Specifically, debt stabilization requires the asymptotic rate of growth of the debt/GDP ratio to be zero, whereas a no-Ponzi-game condition allows the debt/ratio to grow so long as it is lower than the growth-adjusted asymptotic interest rate. For a discussion, see Willems and Zettelmeyer (2022).

Figure 4

Sovereign Debt as Share of GDP in the European G-7 Countries and the United States
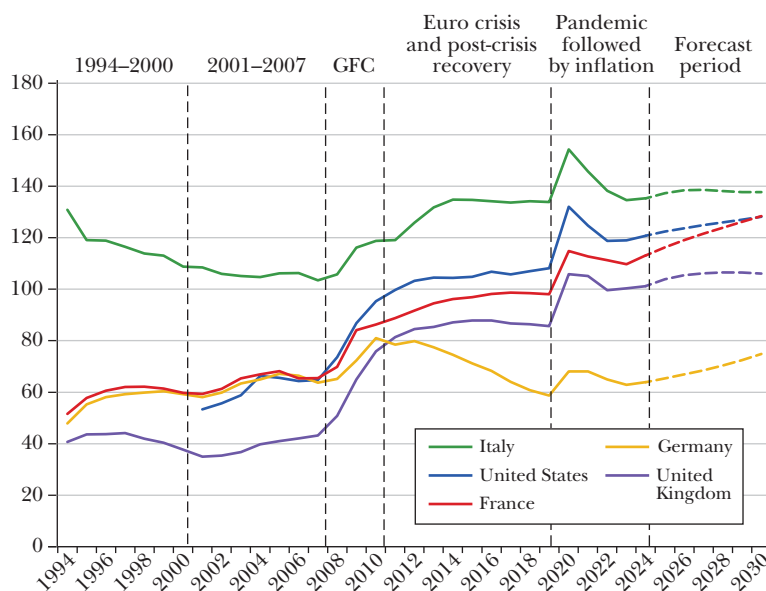
Source: IMF World Economic Outlook Database, April 2025.
Note: GFC stands for "global financial crisis." 2025–2030 values represent IMF forecasts.

indicates, the IMF projects a debt ratio increase for France, Germany, and the United States, suggesting that current fiscal policy is unsustainable (which is confirmed by longer-term projections, such as those prepared by the Congressional Budget Office for the United States). Whether debt is unsustainable, however, requires further analysis.

We undertake this analysis using two approaches. The first estimates the primary budget that would prevent the debt/GDP ratio from exploding, taking into account uncertainty, and assesses the plausibility of reaching and sustaining it. This approach is used by policy organizations such as the International Monetary Fund, the European Commission, or the European Central Bank. The second approach, due to Bohn (1998), estimates a fiscal reaction function—that is, how the country has adjusted its primary budget in the past in response to rising levels of debt—and assesses whether this feedback is strong enough to ensure debt stabilization over time.

Debt-Stabilizing Primary Balances

In the first approach, we project debt under current fiscal policies—the baseline—and use this to generate a stochastic debt projection in an adjustment

scenario in which debt stabilizes over the long term with a given probability (taken to be 70 percent).[9]

The first step is to generate a deterministic forecast of the debt/GDP ratio over 20 years, based on the current fiscal policy as represented by the 2024 structural primary fiscal balance, which is defined as the primary balance excluding its cyclical component and the impact of temporary fiscal measures, and forecasts of the remaining debt drivers—nominal growth, inflation, interest rates and exchange rates, as well as the so-called stock-flow adjustment.[10] Growth forecasts and stock-flow adjustments are taken from official sources, while interest rates and inflation are forecast from market sources.[11] For many EU countries as well as the United States this shows rising debt/GDP over the entire forecast horizon, reflecting high 2024 fiscal deficits.

To quantify uncertainty, we next build a "fanchart" around the deterministic debt/GDP ratio projection. This involves estimating a vector autoregression model of the debt drivers, which captures their interdependencies. For each year and debt driver, we take a random draw from the set of historical residuals of the estimated model and use these and the model's estimated coefficients to construct a new path for the debt drivers and calculate the associated debt path (following the approach of Bouabdallah et al. 2017). By repeating this process, we generate thousands of scenarios, each with a unique debt path, which allows us to approximate a probability distribution around the deterministic forecast.

The final step is to find the structural primary balance, denoted SPB*, that is just high enough for debt to stabilize over the long term. This requires defining an adjustment period during which the structural primary fiscal balance is assumed to rise linearly to an end-value, and a post-adjustment period in which it is held constant at that end-value, except for "aging costs": changes in expenditures and revenues due to demographic developments, such as costs of public education, pensions, and long-term care. For EU countries, projections for costs related to an aging population are taken from the most recent available estimates of the EU Working Group on Ageing Populations and Sustainability; for the United States, which we include for comparison, we use social and health care spending projections by the Congressional Budget Office.

For each structural primary balance path, one can construct a debt fanchart, following the same steps as in the baseline scenario, except that during the

---

[9] For details, see the Supplemental Appendix available with this paper at the JEP website and Darvas et al. (2025).

[10] In principle, the debt at the end of a given year should equal the debt at the end of the previous year plus the budget deficit during the year. In practice, however, certain transactions, valuation effects, and statistical discrepancies can create differences, referred to as "stock-flow" adjustments. Such adjustments have had a significant impact on the debt dynamics of EU countries (see the annex of Boivin and Darvas 2025). In our calculations, we incorporate the European Commission's stock-flow assumptions, which are available for 2025–2026 for 24 EU countries and for 2025–2044 for Finland, Greece, and Luxembourg.

[11] Market sourced interest rate and inflation rate forecasts refer to forward rates, which are derived from the difference in yields of publicly traded inflation-indexed or nominal bonds of varying maturities.

Figure 5
Illustration of Stochastic Debt Sustainability Methodology for France

Panel A. Fanchart based on continuing 2024   Panel B. Fanchart based on debt-stabilizing
structural primary balance   adjustment over seven years
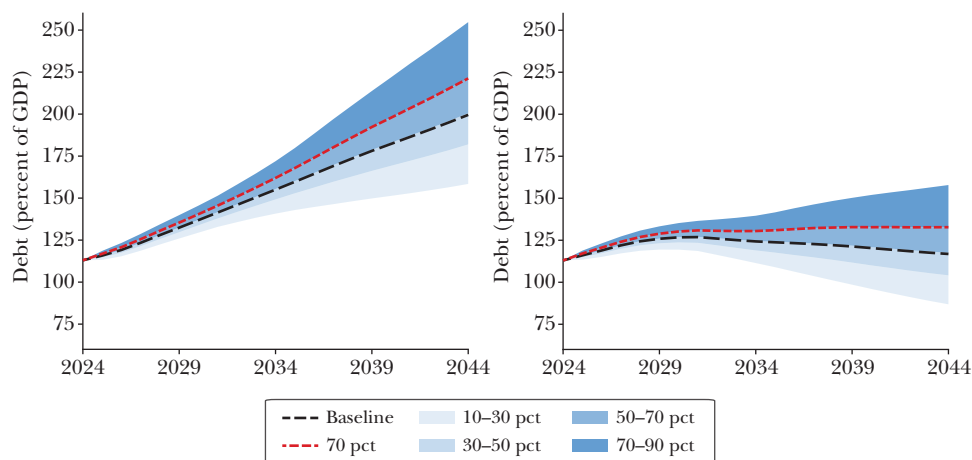
Source: Darvas, Welslau, and Zettelmeyer (2025).
Note: This figure illustrates the projected distribution of the French debt to GDP ratio up to 2044 under
two scenarios: (1) maintaining the 2024 structural primary balance unchanged, and (2) implementing a
fiscal adjustment from 2024 to 2031 that stabilizes the 70th percentile of the distribution between 2039
and 2044.

adjustment period the primary balance is treated as a policy variable (and hence as
deterministic). SPB* is defined as the level of the structural primary balance at the
end of the adjustment period such that the average slope of the 70th percentile of
the debt distribution in the last five years of the 20-year fanchart equals zero. In line
with the choice made by most high-debt EU countries under the new EU fiscal rules,
we assume that the length of the adjustment period is seven years.[12] Consequently,
we consider a post-adjustment period of 13 years. In general, SPB* is not very sensi-
tive to the choice of the length of the adjustment period.

Figure 5 illustrates this approach for France. Figure 5, panel A, shows the
fanchart associated with the baseline debt projection, on the assumption that the
structural primary balance remains at its 2024 level of −3.7 percent, except for
changes in aging costs. If current deficits continue, debt would explode from its
current level of about 110 percent of GDP to more than 180 percent of GDP by 2044
with 70 percent probability (the upper bound of the lightest shaded area). Figure 5,
panel B, shows the same projections on the assumption that the structural primary
balance is raised, in equal annual steps during the adjustment period, such that the

---

[12] European fiscal rules require adjustment within four to seven years. Most high debt EU countries have
requested the seven-year period.

debt stabilizes with 70 percent probability by the end of the adjustment period. This is shown by the flattening of the red dotted line, representing the 70th percentile of the debt fanchart distribution. The end-of-adjustment-period structural primary balance that achieves this stabilization is 1.2 percent of GDP. Hence, according to this estimate, debt stabilization in France requires a medium-term fiscal adjustment by 4.9 percent of GDP spread over seven years—that is, an average adjustment of 0.7 percent of GDP per year.

Table 1 shows the results of applying the same steps to EU countries with above-median debt levels, plus Slovakia, Poland, and Romania, which had large deficits in 2024, along with the United Kingdom and the United States for comparison. The first three columns show preliminary 2024 outcomes for the debt ratio, the fiscal balance, and the structural primary balance.[13] Column 4 contains the main result: SPB*, the structural primary balance at the end of a seven-year adjustment period required to stabilize the debt ratio with 70 percent probability by the end of the 20-year forecast horizon. This ranges from 0.3 percent of GDP for Germany to over 3 percent of GDP for Greece and Hungary. For Italy, SPB* is estimated at 2.5 percent of GDP, for the United States, at 1.8 percent of GDP.

Our results in Table 1 are generally close to the debt sustainability analysis of the European Commission (2024), and the differences do not impact the conclusions that follow, with one exception: Greece, where our methodology results in a much higher adjustment target than the Commission's.[14] This reflects the large output and interest volatility during the Greek sovereign debt crisis, which results in a rise of the higher percentiles of the debt fanchart at longer horizons. Hence, a methodology focused on long-run stabilization will result in a higher adjustment requirement.

Column 5 shows the adjustment requirement associated with these estimates, that is, the difference between SPB* and the structural primary balance in 2024. The country with the highest adjustment requirement is Romania (close to 9 percent of GDP), followed by the United States, Poland, and Slovakia (around 5.5 percent of GDP). In all cases, this reflects very large 2024 structural primary deficits rather than particularly high debt-stabilizing primary balances. Other countries with high

---

[13] The 2024 SPB estimates for EU countries are taken from the European Commission's May 2025 calculations, which may differ from other institutions' estimates due to methodological differences. For example, for Italy, the European Commission's May 2025 estimate was −0.3 percent of GDP, compared with the IMF's April 2025 estimate of 0.2 percent of GDP, which was later revised to 0.4 percent of GDP in the IMF July 2025 Staff Report for the Article IV consultation with Italy. The discrepancy was even larger for Spain, with the Commission estimating −0.7 percent of GDP in May 2025, compared to −2.0 percent by the IMF in April 2025.

[14] For a direct comparison of the results, see Table A1 in the Supplemental Appendix. The main differences with respect our methodology are that the European Commission (1) requires debt to stabilize with 70 percent probability in the first years after the end of the adjustment period rather than by the end of a 20-year period; (2) ignores uncertainty in the debt drivers during the adjustment period; and (3) assumes symmetric probability distributions for the debt drivers rather than allowing skew. The first and third of these differences tend to make the Commission's approach more conservative; the second, less conservative.

Table 1

Debt-Stabilizing Structural Primary Balances, Adjustment Requirement, and Comparison with Medium-Term National Fiscal Plans

	2024 outcomes			*Required adjustment*		*Comparison with medium-term fiscal plans (MTPs)*	
Debt	*Fiscal balance*	*SPB*	*SPB**	*SPB* − SPB (2024)*	*SPB*(MTP)*	*SPB*(MTP) − SPB**
(1)	(2)	(3)	(4)	(5) = (4) − (3)	(6)	(7) = (6) − (4)
Greece	154	1.3	4.0	3.2	−0.8	2.3	−0.9
Italy	135	−3.4	−0.3	2.5	2.7	3.2	0.7
United States	124	−7.5	−4.0	1.8	5.7	−4.0	−5.7
France	113	−5.8	−3.7	1.2	4.9	1.8	0.6
Spain	102	−3.2	−0.7	2.3	3.0	2.3	0.0
Belgium	105	−4.5	−2.0	1.7	3.7	1.6	−0.1
United Kingdom	101	−6.0	−2.5	1.7	4.3	1.0	−0.7
Portugal	95	0.7	2.4	2.9	0.5	2.9	0.0
Finland	82	−4.4	−1.1	0.6	1.7	2.6	2.0
Austria	82	−4.7	−2.5	0.6	3.1	1.2	0.6
Hungary	73	−4.9	0.5	3.7	3.2	2.8	−0.9
Cyprus	65	4.3	4.1	1.4	−2.8	2.4	1.1
Slovenia	67	−0.9	0.1	1.8	1.7	0.5	−1.3
Germany	62	−2.8	−1.0	0.3	1.3	1.1	0.8
Slovakia	59	−5.3	−3.8	1.7	5.5	1.0	−0.7
Poland	55	−6.6	−3.9	1.8	5.7	0.4	−1.4
Romania	55	−9.3	−6.4	2.3	8.7	1.7	−0.6

Source: Darvas, Welslau, and Zettelmeyer (2025).
Note: SPB* denotes the structural primary balance at the end of a seven-year adjustment period that stabilizes debt with 70 percent probability according to DSA methodology described in the text. For the EU countries, SPB*(MTP) refers to the structural primary balance promised after four to seven years in the EU countries' medium-term fiscal plans agreed with the European Commission in late 2024 for most countries, while the plans of Austria, Belgium, and Germany were approved in 2025; for the United Kingdom, to March 2025 projections of the Office for Budget Responsibility, based on the latest fiscal plans announced by the UK government; and for the United States, to May 2025 CBO projections. Debt and deficit numbers refer to the general government for all countries except the United States, for which only the federal government is considered. Pink shading highlights SPB* in excess of 2 percent of GDP; yellow shading highlights adjustment requirement in excess of 3 percent of GDP.

adjustment requirements include France, the United Kingdom, and Belgium. For Greece and Cyprus, the adjustment requirements are negative, reflecting large structural primary surpluses in 2024.

The question is whether these fiscal adjustments should be considered feasible. To answer this question, both the level of SPB* (column 4) and the distance between the 2024 SPB and SPB* (column 5) matter. However, the feasibility of achieving the latter—for example, reducing the deficit of France by almost 5 percentage points of GDP—is sensitive to the time horizon of adjustment. A fiscal adjustment of 5 percent of GDP will generally be unfeasible over a short period but could be feasible if stretched over time. The feasibility of extending the adjustment horizon,

however, depends on the patience of financial markets or (if a country loses market access) of official lenders. This is hard to assess.

In contrast, it is easier to assess the feasibility of sustaining a particular *level* of SPB* (column 4). While there are episodes of very high primary surpluses over short periods (for example, Belgium and Italy achieved primary surpluses of over 5 percent of GDP in the run-up to the euro, and Greece and Cyprus achieved primary surpluses of over 4 percent in the aftermath of their debt crises in the last decade), sustained high primary surpluses in excess of 3 percent are rare. The record holder in the European Union is Belgium, which achieved this level in 15 out of the last 45 years, followed by Italy and Cyprus, with 7 out of the last 45 years.[15] Hence, an SPB* above 3–3.5 percent of GDP would normally be viewed as a red flag.

In this light, the results in Table 1 are generally reassuring. Except for Hungary and Greece—where the high volatility of debt drivers during the Greek crisis likely distorts the results by widening the fan chart—our estimates of SPB* fall within a range that can be considered sustainable. Moreover, columns 6 and 7 show that most EU countries and the United Kingdom are currently targeting fiscal adjustments that would raise their structural primary balance above or close to SPB*. In this respect, the United States is an outlier: according to May 2025 projections of the Congressional Budget Office (2025), projected US structural primary balances are about the same in 2030–2035 as in 2024. At −4 percent, the latter falls well short—by over 5 percent of GDP—of the primary surplus that would be required to stabilize the debt according to our calculations.

At the same time, the medium-term adjustment plans of most EU countries and the United Kingdom look optimistic, for four reasons.

First, many EU countries and the United Kingdom are seeking to increase their defense spending rapidly, a factor not yet taken into account in most of the medium-term adjustment plans shown in Table 1.[16] As a result, in March 2025, the European Commission (2025) proposed, and the Council of Finance Ministers approved, that defense spending could be increased by up to 1.5 points of GDP over the next four years compared to the approved fiscal paths. But while helpful from the perspective of managing the requirements of the EU fiscal rules in the short term, this exception merely postpones the additional adjustment requirement associated with higher defense spending and does not help with debt sustainability.

Second, historical precedents for the magnitude of adjustment indicated in column 6 of Table 1 are exceedingly rare for several countries. According to Darvas et al. (2025), for example, the frequency of a fiscal adjustment of 4.9 percent of GDP,

---

[15] Data are from the IMF Public Finances in Modern History database, available at https://www.imf.org/external/datamapper/datasets/FPP. See Eichengreen and Panizza (2016) and Zettelmeyer, Kreplin, and Panizza (2017) for related historical analyses.

[16] At the June 25, 2025, NATO summit, NATO allies (which include all EU members except for Austria, Cyprus, Ireland, and Malta) agreed to raise military expenditures to at least 3.5 percent of GDP annually by 2035 from the previous NATO target of 2 percent. Some of the medium-term plans underlying column 6 of Table 1—including those of Poland, the Baltic countries, and Germany—already envisage increases in defense spending of this magnitude, but most did not.

as required for France, is about 11 percent of all observed seven-year changes in the structural primary balance since 1970, and zero percent in France during this period.

Third, many of the national medium-term plans of EU countries are based on nominal growth assumptions that are more optimistic than those of the European Commission, and may well be too optimistic (Boivin and Darvas 2025; Zettelmeyer, Darvas, and Welslau 2025). Because the operational fiscal target under EU fiscal rules is not the structural primary balance but rather nominal expenditure growth, use of optimistic assumptions implies the structural primary balance in seven years may turn out substantially lower than envisaged in the medium-term plans, even if countries comply with their expenditure growth ceilings.

Fourth, and likely related to the previous three points, the International Monetary Fund's April 2025 forecast for the primary balance is substantially less optimistic than the announced medium-term targets for almost all EU countries (see Supplemental Appendix, Table A2).

The bottom line is that debt appears sustainable in all or almost all of the countries shown in Table 1, but with an important caveat. In many cases—about a dozen, according to the table—stabilizing the debt ratio will require large adjustments, several of which will likely require more time than the maximum seven years envisaged by the European Union's fiscal rules. Such an adjustment may be workable, so long as countries receive the market financing for their debt that allows them to stretch out their adjustment. If they do not—that is, if they lose market access for their government debt—they will need to resort to public crisis assistance, likely led by the European Union and/or the International Monetary Fund.

Fiscal Reaction Functions

An alternative approach to thinking about the sustainability of debt focuses on the historical feedback from debt or expected deficits to the primary fiscal balance (Bohn 1998). The question is whether this feedback—if it continues in the future—is strong enough to ensure debt stabilization over time.

If the real growth rate of the economy exceeds the real interest rate paid on the debt, then debt/GDP ratio will decline over time—in effect, the denominator of the ratio will outgrow the numerator. If the real interest rate exceeds the growth rate of the economy, however, debt will only be sustainable if the government reacts to an increase in debt by increasing the primary surplus by an amount that exceeds the product of the debt ratio and the difference between the real interest rate and the real growth rate. If so, the debt ratio will stabilize.[17] Bohn (1998)

---

[17] More specifically, Bohn (1998) estimates:

$$s_t \,=\, \rho d_{t-1} + z_t + \epsilon_t,$$

where s_t is the primary balance as a share of GDP, d_t is debt as a share of GDP, z_t is a set of controls, ε_t is the error term, and ρ is a parameter. If $1 + \rho$ is larger than $(1 + r_t)/(1 + g_t)$, which is approximately the same as saying that $\rho > r_t - g_t$, the difference between the real interest rate and the real growth rate—then the debt ratio will stabilize.

estimated the respective parameter to be about 0.05—much more than the typical difference between interest rates and growth in advanced countries, which has historically been negative or in the order of a few basis points (less than 0.03). Thus, governments were in general reacting to existing debt in a way that made debt sustainable.

But has that pattern continued in the twenty-first century? In a study of the sustainability of the US debt/GDP ratio, Auerbach and Yagan (2025) looked at how the US fiscal reaction function has changed over time. They split their sample into two periods: the period before the global financial crisis and the period after it. They find that between 1984 and 2004, the US government responded to debt in a way that made debt sustainable—but has not been doing so in the last couple of decades.

We apply a similar analysis to a broader sample of countries including EU countries, as well as the United Kingdom and the United States between 1990 to 2024.[18] Our fiscal reaction regression has the primary budget surplus as a percent of GDP as the dependent variable. The key explanatory variable is the debt/GDP ratio. We also include a variable that measures the cyclical state of the economy, the ratio of actual to potential GDP, as well as controls for inflation, trade, political stability, crisis-related expenditures, and fixed effects for countries.

There are two obvious concerns in such a regression. First, the cyclical state of the economy is not exogenous. Second, the dependent variable, the primary balance, may be correlated with previous observations. We address these issues by following Plödt and Reicher (2015) and Checherita-Westphal and Žďárek (2017) in employing an instrumental variable approach to control for the endogeneity of the cyclical state of the economy and by including lagged values of the primary balance to control for autocorrelation.[19]

We first estimate the average fiscal response function for three country groups: Western EU countries, Eastern EU countries, and a sample including Western EU countries, the United Kingdom, and the United States. Between 1990 and 2007, we observe a robustly positive fiscal reaction in the first and third samples. Here, a one percentage point increase in debt was associated with an increase in primary surpluses in the following year between 0.04 and 0.05 percentage points—similar to Bohn's (1988) results. For Eastern EU countries, the estimated coefficient was not significantly different from zero. For the second sample period, point estimates for the Western European Union, the United Kingdom, and the United States remain significant but fall to circa 0.02, while the coefficient for Eastern EU countries rises to circa 0.03 and becomes statistically significant.

---

[18] For the new EU member states in our sample, Bulgaria, Cyprus, Czechia, Estonia, Croatia, Hungary, Lithuania, Malta, Poland, Romania, Slovakia, and Slovenia, data are only available from 1995 onwards.

[19] Instruments include the lagged output gap, the second lag of the debt ratio, and the first and second lag of output growth gap. Because the inclusion of lagged values of the dependent variable can result in biased estimates when using a fixed effects estimator, we instead use the Arellano-Bond Generalized Method of Moments estimator. Full details, results, and robustness checks for the regression are available in the Supplemental Appendix.

To identify between-country differences, we decompose the aggregate effect of the Western EU, UK, and US sample by interacting the effect of debt ratios with the country-level fixed effect variables. Country-specific estimates are positive and significant for all but one country (Spain) in the first sample half, ranging from 0.04 (France) to 0.13 (Sweden). In the second sample half, however, the fiscal response is only significantly different form zero for seven countries (Germany, Finland, France, Netherlands, Portugal, and Sweden).

Figure 6 illustrates the fiscal reaction function coefficients and 90 percent confidence intervals for a selection of countries based on a rolling 15-year estimation window for Western EU countries, the United Kingdom, and the United States. The estimates are positive for much of the sample in most countries, but decline over time in the years since the global financial crisis. For Belgium, France, Italy, and the United States, this results in negative point estimates for the last 15-years; for Denmark, Germany, Greece, Portugal, Spain, and the United Kingdom, estimates are close to zero.

A possible explanation for the smaller fiscal reactions since the global financial crisis could be the higher level of the debt. Ghosh et al. (2013) suggest that fiscal adjustment diminishes at high debt levels, a phenomenon they call "fiscal fatigue." To test their hypothesis, they estimated a fiscal reaction function with a third-order polynomial in the lagged debt/GDP ratio. A negative and significant coefficient of the cubic term implies fiscal fatigue. We follow their approach and add quadratic and cubic terms of the lagged debt/GDP ratio to our fiscal reaction function. Between 1990 and 2007, we only find evidence of fiscal fatigue for Eastern EU countries. For the second sample half, however, we find evidence of fiscal fatigue for the two samples including Western EU countries, the United Kingdom, and the United States.

The presence of fiscal fatigue implies a "debt limit," above which debt becomes unsustainable (Ghosh et al. 2013). This is the level of debt at which the primary surplus is exactly sufficient to cover the growth-adjusted interest service of an extra unit of debt. When debt exceeds this limit, the primary surplus is not enough to cover interest payments, debt grows continuously, and the government defaults. With an interest-growth differential around zero, the debt limit is equal to the largest real root of our estimated fiscal reaction function. For Western EU countries in the period since the global financial crisis, the debt limit is 236 percent of GDP. For the sample also including the United Kingdom and the United States, it is 217, about 30 percentage points above estimates by Ghosh et al. (2013).

What does this imply for debt sustainability? The positive correlation between primary surpluses and debt ratios appears to have reduced significantly amidst the fiscal turmoil of the past decade and a half. Although a fiscal response around 0.02 need not result in an explosive debt path (assuming equilibrium interest growth differentials remain below this level), this is not an encouraging development.

The results also show that fiscal reactions to debt ratios may change over time. One reason why policy reactions may change, for example, is the evolution of fiscal institutions—such as the new EU fiscal framework—and a recognition, shared by policymakers, that some countries could be too close to their debt limits. The fact that most EU countries have promised adjustment in line with their

Figure 6
Reaction of Primary Surpluses to Lagged Debt, Rolling 15-Year Sample Estimates
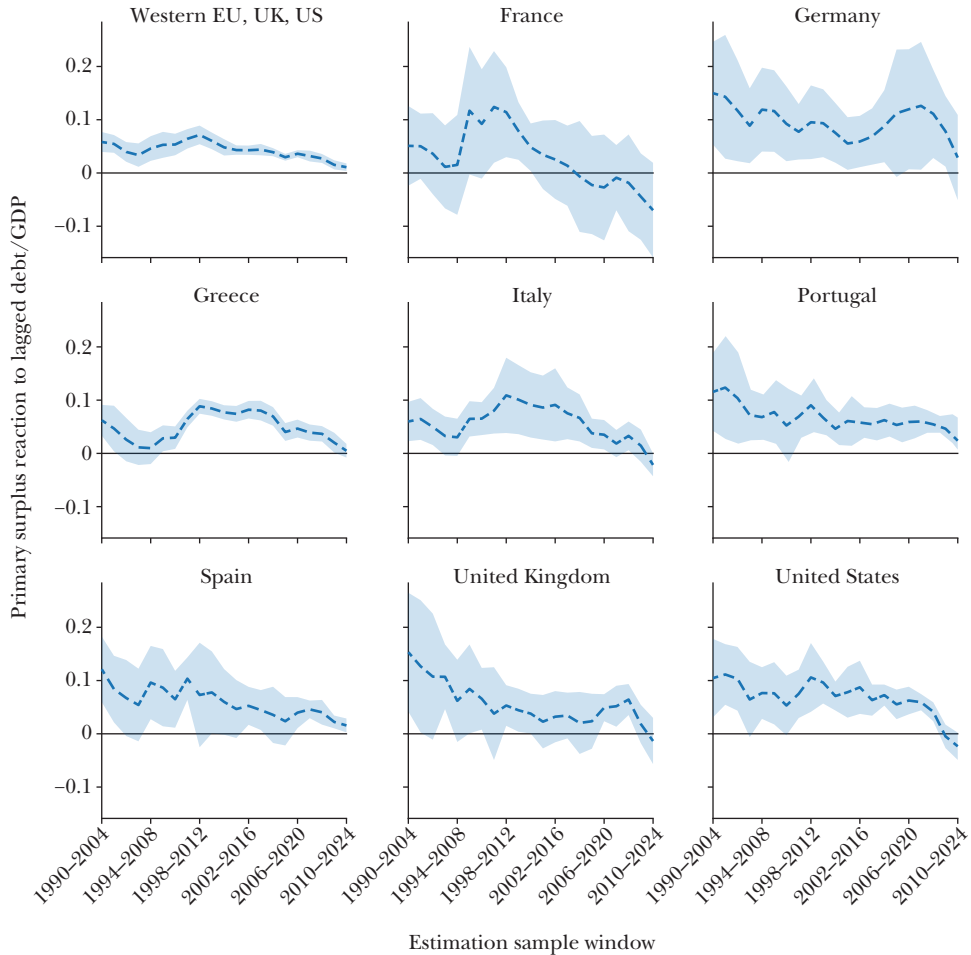
Source: Authors' calculations based on IMF World Economic Outlook, IMF Public Finances in Modern History, IMF Direction of Trade Statistics, International Country Risk Guide (ICRG), European Commission AMECO, and Bureau of Economic Analysis.
Note: Figure is based on 15-year rolling window fiscal reaction estimates from a sample including Western EU, United States, and United Kingdom. The model includes controls for country fixed effects, lagged primary balance, actual to potential output ratio, irregular expenditure, trade as share of GDP, inflation, and political stability using the Arellano-Bond GMM estimator. Dashed blue lines are point estimates of the effect of lagged debt ratios on primary surpluses. Shaded areas are 90 percent confidence intervals based on cluster robust standard errors.

medium-term adjustment requirements, as shown in the previous section, could indicate such a change. However, if the decrease in fiscal responsiveness continues, or interest-growth differentials increase substantially, debt in Europe may become unsustainable.

Do EU Fiscal Rules Require Another Reform?

The EU fiscal rules enacted in 2024 have generally gotten off to a good start. As of mid-July 2025, all 27 EU countries submitted "medium-term fiscal-structural plans" to the European Commission that by and large respected the spirit and letter of the new rules. The main caveat, as discussed above, is that these plans are in many cases based on optimistic macroeconomic assumptions. This implies that even though countries promise to meet their adjustment targets ex ante, they are likely to miss them ex post—in some cases by a wide margin.

Aside from this point, the limitations of the new rules have become apparent in two areas, which may require a "reform of the reform."

First, the medium-term fiscal plans proposed would raise EU public investment by just 0.25 percent of GDP by 2028, which would do little to fill the major investment gaps identified by Draghi (2024). Adjustment pressures under the new rules have likely contributed to this outcome. As Boivin and Darvas (2025) illustrate, fiscal adjustment was associated with substantial declines in public investment in EU countries in the 2008–2013 crisis period—from the global financial crisis through the euro crisis. This negative correlation persists in the 2024–2028 mediumterm fiscal plans. Hence, one of the stated objectives of the new rules—to protect public investment in times of fiscal adjustment and encourage higher investment in the aggregate in the European Union—does not appear to have been reached.

Second, EU countries whose debt is above 60 percent of GDP do not benefit from the flexibility of raising their debts to finance EU-endorsed investment, even if they could do so without raising their borrowing costs and without endangering their debt sustainability. In the past, this constraint had little practical effect, because the largest country in this category—Germany—had no inclination to spend more, in part because it was constrained by a constitutional fiscal rule, the "debt brake," that capped Germany's federal cyclically adjusted fiscal deficit at 0.35 percent of GDP. But in March 2025, Germany's Parliament mustered a two-thirds majority to replace the debt brake with a provision that allows unlimited borrowing for defense purposes, while maintaining a cap on nondefense spending, except for spending from extrabudgetary funds, which also require a two-thirds parliamentary majority. The same parliamentary act created such a fund, worth about 11 percent of GDP, to be spent within 12 years.

However, as argued by Steinbach and Zettelmeyer (2025), this extra spending cannot be reconciled with the EU fiscal rules unless nondefense spending is compressed far more than the reformed "debt brake" requires or Germany makes very optimistic assumptions about nominal growth, helping the government project a stronger improvement in the structural primary balance than the likely outturn. The medium-term plan presented by Germany in July 2025 combines both elements: very optimistic macroeconomic assumptions together with a backloaded adjustment plan, which implies a much higher compression of spending at the end of the adjustment period than Germany's new national debt brake would require (Zettelmeyer, Darvas, and Welslau 2025). Both are a bad idea: the former, because it

undermines the purpose of fiscal rules; and the latter, because it leads to an undesirable swing in fiscal policy.

A better solution would be a further reform of the fiscal rules. This should include two elements.

First, a "fiscally responsible public investment rule" could exempt public investment plans endorsed by the EU council from quantitative debt and deficit reduction requirements, provided that the additional investment is consistent with debt sustainability. Such a rule might require additional fiscal adjustment in the noninvestment budget to ensure the structural primary balance at the end of the adjustment period is sufficiently high to stabilize debt (Darvas, Welslau, and Zettelmeyer 2024, Appendix 1).

Second, the rules could offer more flexibility to countries that do not pose fiscal risks to maintain or raise their debt levels, even if their debt exceeds 60 percent of GDP. One way of doing so would be to increase the EU treaty's debt "reference value" from 60 to 90 percent of GDP—leaving most other elements of the new fiscal rules unchanged.[20] Making this change would require unanimity among EU governments. Another approach, proposed by Pench (2025), would be to give countries with debt above 60 percent, but low fiscal risks, leeway to increase their debt before they are required to reduce it. Low fiscal risks could be defined, for example, by the level of fiscal adjustment required to stabilize the debt, or as assessed by a sovereign risk framework like those operated by the European Commission (2024) or the International Monetary Fund (2022).

Can the EU Expand Its Pool of Supranational Safe Debt?

At the end of 2024, EU-level supranational, AAA-rated debt stood at about €1.1 trillion, roughly 7 percent of EU GDP. Over half of this is EU debt was issued during and since the pandemic under two programs: the Next Generation EU (NGEU) funds created to support public spending and prevent a fiscal crisis during the pandemic (with a with face value of about €410 billion when the program is completed in 2026), and the €98 billion SURE (Support to mitigate Unemployment Risks in an Emergency), a temporary loan facility to finance employment schemes and mitigate pandemic-related income losses. The European Union has also borrowed to finance financial assistance loans for non-EU countries, about €50 billion so far, which is set to double based on commitments made to Ukraine.

The remainder of the outstanding EU-level debt consists of debt issued by EU agencies: the European Investment Bank (€220 billion), the European Stability Mechanism (€67 billion), and the ESM's predecessor, the European Financial

---

[20] In addition, the so-called "deficit resilience safeguard" rule, which envisages minimum steps of fiscal adjustment until the deficit is below 1.5 percent of GDP, would have to be eliminated or waived for borrowing that finances EU-endorsed public investment.

Stability Facility (EFSF, €197 billion). While an additional €350 billion is expected to be issued by the end of 2026 to finance the remaining commitments under NGEU, EU supranational debt remains small both as a share of GDP and relative to the remaining pool of euro-denominated AAA-rated bonds, issued by Germany, the Netherlands, and Luxembourg, which amounted to about €2.6 trillion at the end of 2024.

Expanding the pool of euro-denominated EU-level AAA-rated debt could have two beneficial effects. First, if this debt was held by euro-area banks in lieu of sovereign bonds, they would eliminate the bank-sovereign doom loop without raising credit risk (Brunnermeier et al. 2017; Alogoskoufis and Langfield 2020). Second, it could contribute to a unified euro capital market, creating a homogenous pool of euro area bonds that attracts foreign investors, and potentially lower the cost of borrowing in the euro area (Monti 2010). This argument has gained prominence recently as some investors and central banks have expressed openness to diversifying away from US Treasury debt to escape policy uncertainty and fiscal risks in the US economy (Merler 2025).

How to Expand Safe AAA-Rated EU Debt

The question is how to do it. Expanding EU level-debt issuance would require a political decision to centralize additional fiscal functions, which is not likely. This leaves a possibility explored by a literature going back to Delpla and von Weizsäcker (2010) and Monti (2010), namely, to create safe EU-level debt synthetically through a combination of tranching and pooling. "Tranching" would create debt instruments that follow a seniority structure, for example, with levels or senior and junior debt. If the junior tranches that are the first to absorb any losses are sufficiently "thick," the senior tranches could even feature AAA ratings. "Pooling" would combine debt from different countries into a single pool. These proposals can be divided into two categories, depending on whether pooling or tranching comes first.

In the proposals that call for pooling followed by tranching (Brunnermeier et al. 2011, 2016, 2017), financial intermediaries would purchase euro area sovereign bonds roughly in proportion to the shares that EU national banks hold at the European Central Bank, up to some limit that ensures the liquidity of the market in the remaining bonds, while simultaneously issuing multi-tranche "sovereign bond backed securities" backed by this portfolio. Debt service received by the intermediary would be paid to the holders of these securities using a "waterfall" structure: in the event of a default on some of the underlying bonds, the holders of the senior tranche would be paid first. According to simulations by Brunnermeier et al. (2017), the size of the junior tranches required to ensure that the five-year expected loss rate of the senior tranche corresponds to that of the AAA-rated German bund is about 30 percent. That is, the face value of the senior tranche—referred to as "European Senior Bond," or "ESBie"—would be 70 percent of the face value of the cover pool.

In the proposals that call for national-level tranching followed by pooling (Leandro and Zettelmeyer 2018; Giudice et al. 2019), euro area sovereigns would

issue bonds in several tranches. The most senior tranche would be limited as a common share of GDP or share of total debt (or by the minimum of the two). Financial intermediaries could purchase the senior tranche while simultaneously issuing a single, AAA-rated security based on the pool of senior debt. Alternatively, investors wishing to reduce risk through diversification as well as seniority could buy a pool of senior tranches themselves.

Both approaches exist in two variants: with a single public intermediary, and with regulated private intermediaries. Based on the Brunnermeier et al. (2017) sovereign default model, Leandro and Zettelmeyer (2019) showed that either approach could generate a maximum volume of safe euro-nominated bonds in the order of 20 percent of euro area GDP, corresponding to about €3 trillion euros in 2024—a modest amount compared to the US Treasury market, but more than enough to replace all national debt in euro area bank balance sheets.

The main difference between the two approaches is that in the second national-level tranching, all new sovereign issuance would occur through the junior bond(s). While this shift would not affect the average cost of debt, it would increase its marginal cost—an advantage from the perspective of encouraging fiscal discipline, but a significant disadvantage from the perspective of lower-rated sovereign issuers concerned about market access. Perhaps for this reason, the national-level tranching approach was never pursued (though Blanchard and Ubide [2025] have recently tried to revive it).

In contrast, the pooling-then-tranching approach was explored by a High-Level Task Force of the European Systemic Risk Board (ESRB 2018). This report focused on developing a market for *privately* issued sovereign bond backed securities, and found that this approach raised operational challenges: Would the tranches within the pool of securities be truly homogenous and liquid? How would the pool and tranches be affected by eventual debt restructuring cases? To what extent would the role of a private intermediary add risks of its own?

These issues would be much easier to address by a lightly capitalized public intermediary than by private intermediaries. But in the aftermath of the euro crisis from 2010–2012, the public intermediation approach was rejected by the "frugal," higher-rated euro countries, who feared that it would give rise to implicit guarantees and moral hazard on the side of the weaker-rated sovereign bond issuers.

Steps for EU Policymakers

To seize the opportunity of the new interest in the euro area safe debt as a complement and possible alternative to the US Treasury market, EU policymakers should take two steps.

First, existing EU debt stemming from NextGenerationEU grants should be rolled over ad infinitum, rather than wound down. It makes no sense to add an additional burden to EU fiscal adjustment for the purpose of eliminating an asset that provides significant value at a small public cost.

Second, EU policy makers should reconsider the Brunnermeier et al. (2011, 2017) ESBie proposal—but in the version involving a capitalized *public* intermediary

rather than the version with private intermediaries. A public intermediary could avoid or overcome many of the regulatory and operational problems that the High-Level Task Force of the European Systemic Risk Board identified, and endowing it with some capital would provide investors with an assurance that unanticipated problems related to tranching and pooling (as opposed to the underlying debt assets) could be handled. Hence, European Senior Bonds issued by a public European Debt Agency will likely find much better reception by investors than European Senior Bonds issued by private entities, even if they are backed by exactly the same cover pool. It would also serve as a symbol of European resolve to make the euro a widely used and traded global currency in an uncertain and shifting international environment.

Conclusion

This paper has taken stock of sovereign debt risks across countries of the European Union following a series of shocks over the past five years, including a pandemic, an energy price shock, an inflation shock accompanied by a sharp tightening of monetary policy, and mounting pressures to increase defense spending. Euro area countries faced these shocks with the added complications of a macroeconomic framework combining centralized monetary policy with decentralized fiscal policy. Our main findings can be summarized in three points.

First, although fiscal and financial fragmentation persists, the fiscal risks related to this fragmentation seem to have receded thanks to institutional reforms to strengthen banking systems, receded macroeconomic imbalances, and the much greater willingness of the European Central Bank to intervene in sovereign debt markets. This was true even before the announcement of the ECB's latest antifragmentation tool, the Transmission Protection Instrument, as the muted reaction to that instrument shows.

Second, with few exceptions, stabilizing the future trajectory of debt/GDP with reasonably high probability does not require unrealistically high primary surpluses. In this sense, debt remains sustainable in most EU countries. Debt stabilization will require very large fiscal adjustments in some countries. But the countries with the largest adjustment requirements—Romania, Slovakia, and Poland—have relatively low debt. Furthermore, judging from the medium-term fiscal plans agreed with the European Commission, these and most other countries seem to be willing to make an effort.

Third, for some countries, adjustments may take much longer than promised, potentially straining the patience of markets. In France and Romania, the adjustments agreed upon with the European Commission have few, if any, historical precedents. Consistent with Auerbach and Yagan's (2025) findings for the US debt/GDP ratio, fiscal reaction function estimates indicate that the feedback from high debt to fiscal tightening has weakened and is close to zero in many European countries—including Belgium, France, Italy, Germany, Greece, Portugal, Spain, and the United Kingdom. Debt stabilization will hence require a break

with past behavior. It remains to be seen whether the recently reformed fiscal framework and/or fear of market reactions are sufficient to produce this break.

Europe's fiscal challenges go beyond debt stabilization. Europe will also need to reprioritize spending in the direction of defense and investment, and it should seize the opportunity created by investor nervousness about policies in the United States to create a larger and more homogenous safe debt market. Addressing these challenges is technically feasible but will require significant additional reform.

References

Alogoskoufis, Spyros, and Sam Langfield. 2020. "Regulating the Doom Loop." *International Journal of Central Banking* 16 (4): 251–91.

Altavilla, Carlo, Domenico Giannone, and Michele Lenza. 2016. "The Financial and Macroeconomic Effects of OMT Announcements." *International Journal of Central Banking* 12 (3): 29–57.

Auerbach, Alan J., and Danny Yagan. 2025. "Robust Fiscal Stabilization." NBER Working Paper 33374.

Bilbiie, Florin, Tommaso Monacelli, and Roberto Perotti. 2021. "Fiscal Policy in Europe: Controversies over Rules, Mutual Insurance, and Centralization." *Journal of Economic Perspectives* 35 (2): 77–100.

Blanchard, Olivier, and Ángel Ubide. 2025. "Now Is the Time for Eurobonds: A Specific Proposal. " Realtime Economics (blog), Peterson Institute for International Economics, May 30. https://www.piie.com/blogs/realtime-economics/2025/now-time-eurobonds-specific-proposal.

Bohn, Henning. 1998. "The Behavior of US Public Debt and Deficits." *Quarterly Journal Economics* 113 (3): 949–63.

Boivin, Nicolas E., and Zsolt Darvas. 2025. "The European Union's New Fiscal Framework: A Good Start, but Challenges Loom." Bruegel Policy Brief 06/2025.

Bouabdallah, Othman, Cristina D Checherita-Westphal, Thomas Warmedinger, Roberta De Stefani, Francesco Drudi, Ralph Setzer, and Andreas Westphal. 2017. "Debt sustainability analysis for euro area sovereigns: a methodological framework." ECB Occasional Paper, (185).

Brunnermeier, Markus K., Luis Garicano, Philip R. Lane, Marco Pagano, Ricardo Reis, Tano Santos, David Thesmar, Stijn Van Nieuwerburgh, and Dimitri Vayanos. 2011. "European Safe Bonds (ESBies)." Unpublished.

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

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

Checherita-Westphal, Cristina, and Václav Žd'árek. 2017. "Fiscal Reaction Function and Fiscal Fatigue: Evidence for the Euro Area." ECB Working Paper 2036.

Congressional Budget Office. 2025. *The Long-Term Budget Outlook Under Alternative Scenarios for the Economy and the Budget.* CBO.

Darvas, Zsolt, Gonzalo Huertas, Lennard Welslau and Jeromin Zettelmeyer. 2025. "What Will It Take to Stabilise Debt in Advanced Countries?" Unpublished.

Darvas, Zsolt, Lennard Welslau, and Jeromin Zettelmeyer. 2024. "The Implications of the European Union's New Fiscal Rules." Bruegel Policy Brief 10/2024.

Darvas, Zsolt, Lennard Welslau, and Jeromin Zettelmeyer. 2025. *Data and Code for: "Sovereign Debt and Fiscal Integration in the European Union."* Nashville, TN: American Economic Association; distributed by Inter-university Consortium for Political and Social Research, Ann Arbor, MI. https://doi.org/10.3886/E238402V1.

De Grauwe, Paul. 2012. "The Governance of a Fragile Eurozone." *Australian Economic Review* 45 (3): 255–68.

Delpla, Jacques, and Jakob von Weizsäcker. 2010. "The Blue Bond Proposal." Bruegel Policy Brief 2010/03.

Draghi, Mario. 2024. "The Future of European Competitiveness: A Competitiveness Strategy for Europe" Publications Office of the European Union.

Eichengreen, Barry, and Ugo Panizza. 2016. "A Surplus of Ambition: Can Europe Rely on Large Primary Surpluses to Solve Its Debt Problem?" *Economic Policy* 31 (85): 5–49.

European Central Bank. 2024. *Financial Integration and Structure in the Euro Area.* ECB.

European Commission. 2024. *Debt Sustainability Monitor 2023.* Institutional Paper 271, Directorate-General for Economic and Financial Affairs, European Commission.

European Commission. 2025. *COMMUNICATION FROM THE COMMISSION: Accommodating increased defence expenditure within the Stability and Growth Pact.* C(2025) 2000 final.

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

Ghosh, Atish R., Jun I. Kim, Enrique G. Mendoza, Jonathan D. Ostry, and Mahvash S. Qureshi. 2013. "Fiscal Fatigue, Fiscal Space and Debt Sustainability in Advanced Economies." *Economic Journal* 123 (566): F4–F30

Giudice, Gabriele, Mirzha de Manuel, Zenon Kontolemis, and Daniel P. Monteiro. 2019. "A European Safe Asset to Complement National Government Bonds." Preprint, SSRN. http://dx.doi.org/10.2139/ssrn.3447173.

International Monetary Fund. 2022. *Staff Guidance Note on the Sovereign Risk and Debt Sustainability Framework for Market Access Countries.* IMF Policy Paper 2022/039.

James, Harold. 2014. *Making the European Monetary Union.* Harvard University Press.

Lane, Philip R. 2021. "The Resilience of the Euro." *Journal of Economic Perspectives* 35 (2): 3–22.

Leandro, Álvaro, and Jeromin Zettelmeyer. 2018. "The Search for a Euro Area Safe Asset. " Peterson Institute for International Economics Working Paper 18-3.

Leandro, Álvaro, and Jeromin Zettelmeyer. 2019. "Creating a Euro Area Safe Asset without Mutualizing Risk (Much)." *Capital Markets Law Journal* 14 (4): 488–517.

Merler, Silvia. 2025. "More EU Debt Issuance Would Be the Best Response to Trump's Tariffs." First Glance, Bruegel, April 11. https://www.bruegel.org/first-glance/more-eu-debt-issuance-would-be-best-response-trumps-tariffs.

Monti, Mario. 2010. *A New Strategy for the Single Market.* Report to the President of the European Commission.

Office for Budget Responsibility. 2025. *Economic and fiscal outlook.* CP 2389, OBR.

Pench, Lucio. 2025. "European Union Fiscal Rules: It's Already Time to Reform the Reform. " Analysis, Bruegel, July 15. https://www.bruegel.org/analysis/european-union-fiscal-rules-its-already-time-reform-reform.

Plödt, Martin, and Claire A. Reicher. 2015. "Estimating Fiscal Policy Reaction Functions: The Role of Model Specification. " *Journal of Macroeconomics* 46: 113–28.

Steinbach, Armin, and Jeromin Zettelmeyer. 2025. "Germany's Fiscal Rules Dilemma." Analysis, Bruegel, April 24. https://www.bruegel.org/analysis/germanys-fiscal-rules-dilemma.

Véron, Nicolas. 2024. *Europe's Banking Union at Ten: Unfinished yet Transformative.* Bruegel.

Willems, Tim, and Jeromin Zettelmeyer. 2022. "Sovereign Debt Sustainability and Central Bank Credibility." *Annual Review of Financial Economics* 14: 75–93.

Zettelmeyer, Jeromin, Zsolt Darvas, and Lennard Welslau. 2025. "What Germany's Medium-Term Fiscal Plan Means for Europe." Unpublished.

Zettelmeyer, Jeromin, Eike Kreplin, and Ugo Panizza. 2017. "Does Greece Need More Official Debt Relief? If So, How Much?" Peterson Institute for International Economics Working Paper 17-6.

# China's Lending to Developing Countries: From Boom to Bust

Sebastian Horn, Carmen M. Reinhart, and Christoph Trebesch

I n the last two decades, China has become the world's largest state creditor and a key player in the international financial system. China's lending footprint in the developing world currently rivals that of the World Bank. As part of this transformation, the Chinese government and its state-owned banks provided more than $800 billion in loans, mostly denominated in US dollars, to developing countries to finance infrastructure and energy projects (Dreher et al. 2022). China's lending to developing countries has been spearheaded by a small number of state-owned banks, and this bank-based approach departed from the trend among emerging markets towards bond financing since the 1990s. However, circa 2014–2015, many of the countries that had borrowed from China were faced with mounting debt-servicing difficulties from a decline in commodity prices, which a few years later was followed by the COVID-19 pandemic and then the Russia-Ukraine war. Lending from China went into reverse. As shown in Figure 1, net transfers from China to governments in emerging market and developing economies have been negative since 2019.

In this paper, we place China's overseas lending cycle in the historical context of past capital flow booms and busts, highlighting similarities and distinctions.

■ *Sebastian Horn is Professor of Economics at the University of Hamburg, Hamburg, and a Research Director, Kiel Institute for the World Economy, Kiel, Germany. Carmen M. Reinhart is the Minos A. Zombanakis Professor of the International Financial System at Harvard Kennedy School, Cambridge, Massachusetts. Christoph Trebesch is Professor of Economics at Kiel University, and a Research Director, Kiel Institute for the World Economy, both in Kiel, Germany. Their email addresses are Sebastian.horn@uni-hamburg.de, carmen_reinhart@ hks.harvard.edu, and christoph.trebesch@kielinstitut.de.*

Figure 1

The Reversal of Chinese Lending: Net Transfers between China and Emerging Market and Developing Economies
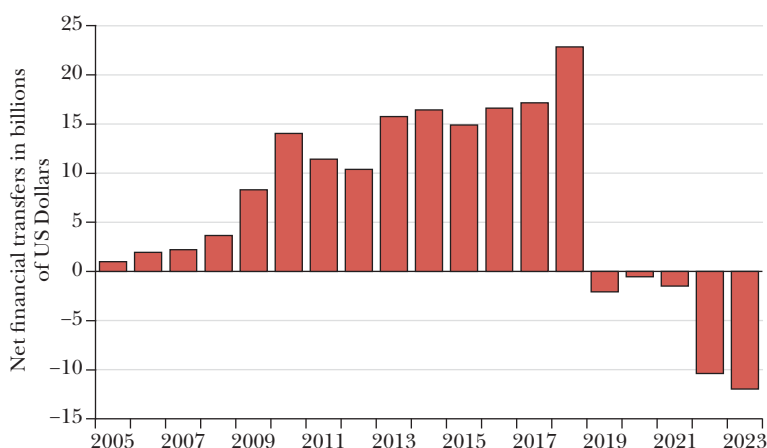
Source: World Bank (2024).
Note: This figure shows net transfers between Chinese state creditors and developing and emerging market country public borrowers. Net transfers are defined as the difference between new lending disbursements and repayments through principal and interest.

We find many parallels with the boom in US commercial bank-led lending to developing countries of the 1970s that came to an abrupt halt with the wave of defaults in the early 1980s. Besides the boom-bust pattern, the 1970s debt cycle was also bank-based, targeted to commodity-rich debtors, and nontransparent. We also discuss the broader implications of China's rise for the "international financial architecture," which has become increasingly fragmented, multipolar, and opaque in recent years, as new creditor powers challenge existing institutions and norms.

Despite China's current stature as an international creditor, the academic literature on the country's role in global finance remains comparatively small—though it is rapidly growing. Systematic empirical studies of China's international lending have been hindered by the opacity of its overseas lending practices. The Chinese government does not report on its official international lending and does not publish detailed data on outstanding overseas debt claims. Debt restructurings between China and debtor countries until very recently have been bilateral and confidential. For example, the major credit rating agencies—Fitch, Moody's, and Standard and Poor's, among others—monitor sovereign debt owed to private creditors (primarily Western banks and bondholders), but sovereign-to-sovereign lending (including by China's state-owned banks) is not their focus. The same is true for commercial data providers such as Bloomberg. China is not a member of the Paris Club, an informal group of 22 creditor governments that attempt to coordinate solutions to sovereign debt repayment problems. Nor is China a member

of the OECD, which tracks development assistance from official bilateral creditors from member countries. Documentation of China's international lending has fallen through the cracks—at least until recently.

To provide a systematic picture of China's lending cycle, we build on a series of research papers and data-collection efforts by us and others, in particular Horn, Reinhart, and Trebesch (2021, 2022), Horn et al. (2023a), Dreher et al. (2022), and Gelpern et al. (2023). The data allow us to document the scale, destination, and terms of China's overseas lending boom—as well as the lending bust and defaults that have followed.

The evidence shows that Chinese state banks are muscular and commercially savvy lenders. They charge higher interest rates than other official creditors and include numerous creditor-friendly provisions in their loan contracts (as one example, often requiring borrowers to post cash collateral in bank accounts held in China).

A benign interpretation is that creditor-friendly terms were a necessary precaution for China's "subprime" lending to high-risk developing countries, including some of the poorest and least stable nations where few other creditors were willing to engage. In this view, Chinese banks developed innovative contractual designs, including legal threats and sanctions, to address the notorious enforcement problems in cross-border sovereign lending (Bulow and Rogoff 1989). By deterring strategic defaults, China's lending model facilitated beneficial investment flows to countries that are chronically capital-scarce (Lucas 1990; Reinhart, Rogoff, and Savastano 2003; Alfaro, Kalemli-Ozcan, and Volosovych 2008).

A less benign interpretation is that China's lending practices impose excessive burdens on the debtor countries, made debt repayment problems and defaults more likely, rather than less, and aggravated coordination problems with other creditors. This interpretation is in line with theories of "debt dilution," which occurs when a borrowing country takes on new debt and thus reduces the likelihood that the old creditors will be fully repaid (Bolton and Jeanne 2009; Hatchondo, Martinez, and Sosa-Padilla 2016). If additionally, the new debt is senior to existing debt, as is potentially the case for China's loans, then debt dilution and inter-creditor equity problems are amplified. According to this view, China's lending style has created negative externalities for other creditors and makes orderly debt restructuring processes less likely—with potentially larger deadweight losses for debtors. Prolonged defaults are associated not only with significant economic losses but carry significant social costs (poverty, health, nutrition) as well (Farah-Yacoub, Graf von Luckner, and Reinhart 2024).

We begin with some historical context for the unfolding developing country debt cycle, starting with the boom phase. We discuss how China's emergence as a key creditor and its distinct style of lending affected emerging market and developing economies that borrowed. We provide insights on the drivers of China's overseas lending boom and the renewed accumulation of external debt by these countries. We then turn to the "bust" stage of the lending cycle, as a rising number of borrowers began to experience difficulties in servicing their external debts to China

and other creditors. We shed light on why the current boom ended abruptly and whether the terms of China's lending contribute to debt-overhang problems. The reversal in Chinese lending predates the COVID-19 crisis, as does the deterioration in economic fundamentals among many emerging market and developing economies. Our narrative shows that China has responded to the evolving debt-servicing problems of many of its debtors like private creditors of the past had in similar circumstances. Specifically, China has sharply curtailed new lending commitments and engaged in efforts to restructure existing debts. The effects and implications of China's preferred approach of direct lending, versus the prevalent bond financing from private Western creditors, is a theme that runs throughout the discussion.

There are important questions we do not address, some because they are beyond the scope of our earlier and ongoing studies, others because the bust phase of the cycle is still unfolding and a postmortem appears premature. In the former category, we do not provide answers to questions about how the borrowed funds were used by emerging market and developing economies. Infrastructure spending is a broad and imprecise term that encompasses everything from growth-enhancing, high-return projects to white elephants. Corruption has also been identified as a problem. A significant share of Chinese loans has involved kickback schemes that have benefitted leaders and elites in the borrowing country (Isaksson and Kotsadam 2018; Kern, Reinsberg, and Shea 2022).

We also do not discuss the broader geopolitical dimension of China's lending, which has received growing attention amid intensifying great power rivalry between China and the United States (Clayton, Maggiori, and Schreger 2024; Mohr and Trebesch 2025). China's loans are often directed to countries that already display a measure of anti-Americanism in their foreign policy (for example, Mações 2018; Moura 2025). Access to Chinese finance could cement and widen the wedge between the United States and these countries while enlarging the "China bloc." China may also derive strategic and military benefits from financing and operating critical infrastructure such as ports or telecommunication networks in developing countries, exemplified by China's establishment of a military base in Djibouti—a major recipient of Chinese loans (Lew et al. 2021). More rigorous research is needed on these questions.

In the second category—the "postmortem" discussion—we cannot assert with confidence, as the poet Lord Tennyson did, "Tis better to have loved and lost than never to have loved at all."[1] Ultimately, the overall effects of China's lending boom on recipient countries will depend crucially on whether the newly-built infrastructure, a legacy of the boom, will further the development goals of the debtors or not. It remains to be seen whether the potential long-term benefits will outweigh the high and more concrete costs of servicing China's debt or, alternatively, the multifaceted costs of a default in many countries.

---

[1] The famous quotation is from Alfred Tennyson's poem "In Memoriam A. H. H," a tribute to his friend Arthur Henry Hallam who died at age 22. The full text is readily available online, including at https://poets.org/poem/memoriam-h-h.

Overseas Lending and Debt Crises: Past and Present

China's overseas lending boom shares many features with earlier lending booms to developing countries (Reinhart and Rogoff 2009). The generic cycle begins with the rapid extension of cross-border credit and overoptimism about the sustainability of the boom. Lack of oversight of the caliber and magnitude of lending often fuels the boom. But in time, a rising share of borrowers face repayment challenges and some default and/or restructure their debts.

The parallels are particularly striking when comparing China's boom to the syndicated bank lending surge of the 1970s, which became the debt crisis of the 1980s (Cline 1995). In the 1970s, the caliber of lending was assessed to be sound. James (1996) provides a compelling narrative of the (over)confidence of bankers, debtor-country policymakers, and academics during the height of the 1970s surge in lending. He notes: "In an article published in 1981, for instance, Jeffrey Sachs of Harvard wrote: 'if my analysis is correct, much of the growth in LDC [less developed country] debt reflects increased investment and should not pose a problem of repayment.'" But looking back at the resulting debt crisis, James observes, "Nobody knew enough—in particular about the extent of bank loans. The debt crisis reflected a failure to share and make available information: in effect a failure of the surveillance principle."

The main national-level borrowers in the 1970s were resource-rich developing countries, as major oil shocks and high commodity prices made this group especially attractive to prospective lenders. The lenders were typically a syndicate of ten or more Western banks from the United States, Europe, and Japan. The borrowers were the central government or public companies and the loan's purpose was often to improve the country's infrastructure and/or its resource-extracting industries. Lending was primarily denominated in US dollars, maturities were typically between three and seven years, and interest rates implied risk premia, often more than 1 or 2 percent above the common benchmark at that time, the London Interbank Offered Rate (LIBOR). Much of the syndicated bank lending was not picked up by official statistics, so debtor countries and the International Monetary Fund and the World Bank had an incomplete picture on the resulting debt flows and debt sustainability risks (Díaz-Alejandro 1985; Horn et al. 2024). US and international interest rates were higher then, but not in real terms, as global inflation was also markedly higher.

International lending by China's government and state banks in the past two decades has also been bank-based, opaque, mostly denominated in US dollars, and close to market terms. While some of the borrowers are the same in both episodes, the 1970s debt boom also included some of the larger, higher-income emerging markets, such as Brazil, Mexico, or Poland (among those that defaulted) and Korea or Turkey (among the ones that avoided default). China's debtors markedly skew to low- and middle-low-income countries. This difference is consequential. Another key difference, as the next section discusses, is the pervasiveness of collateralized Chinese loans. In these cases, the loan contract specifies that the repayment is

secured by specific assets or revenue streams (often oil and other commodity revenues) pledged by the government as collateral. Thus, if the government defaults on the loan, the lender can seize or take control of the pledged collateral to recover the outstanding debt.

Despite the increase in debt defaults on Chinese lending in recent years, including Ghana, Zambia, and Sri Lanka, and near-defaults in Egypt and Pakistan, this debt crisis has been comparatively quiet. No systemic debt crisis has yet arisen across emerging market and developing countries. There is little evidence of cross-border contagion or a domino effect, like that triggered by the Mexican default of August 1982. Unlike the eurozone crisis or the financial crisis of 2007–2009, the coverage of recent debt defaults in the financial press has been sparse. Possibly the attention is limited because these crises pose little risk of destabilizing financial contagion (so far, the Russian default of August 1998 remains the last of the contagious crises in emerging market and developing countries). The bad news is that there may be less urgency in resolving the debt crises of nonsystemic, lower-income countries, relative to their wealthier and more consequential counterparts. In the 1980s crisis, it took a decade to solve the crisis in middle-income emerging markets through the "Brady deals" that granted deep debt relief.[2] However, in many low-income countries, it took more than twice as long—well into the 1990s and even the 2000s—to arrange the kind of debt relief that would enable countries to tackle their debt overhang.

There are other salient differences across the two boom-bust episodes. In the 1970s, external debt dominated the landscape, and external creditors were comparatively homogeneous. The modern reality is markedly different. Besides the vast sums owed to China, many developing countries also must deal with heterogeneous private creditors, including bondholders, banks, or commodity traders. Debts owed to multilateral institutions, which are usually exempt from debt restructuring, also loom large for many of these countries.[3] In addition, external debt is no longer the only game in town. Domestic debts have reached record levels over the past two decades in many countries; that is, debts issued under domestic law, denominated in the local currency, and held primarily by domestic banks. The specifics of the creditor base and debt instrument involved vary across debtor countries. These changes in both the creditor base and the greater varieties of debt make crisis resolution and debt restructurings more challenging than before.

---

[2]For a contemporaneous discussion of the strengths and weaknesses of the resolution of middle-income debt in the 1980s, a useful starting point is the five-paper "Symposium on New Institutions for Developing Country Debt" in the Winter 1990 issue of this journal. For a discussion of the later Highly Indebted Poor Countries debt relief initiative launched in the 1990s, see in this journal Arslanalp and Henry (2006).

[3]In effect, during the crafting of the Common Framework for Debt Treatments, China pressed for the inclusion of multilateral debt in debt restructuring negotiations; this initiative was widely rejected by other official bilateral creditors. The Common Framework seeks a coordinated approach to help low-income countries restructure their debt. The framework was launched in November 2020 by the G20 and the Paris Club.

Then and now, the lending boom and eventual bust transformed sovereign debt markets and necessitated deep changes to the international financial architecture. The securitization of bank loans during the US-led Brady Plan for debt restructuring in 1989 ushered in a new era for US-dollar-denominated emerging market bonds.[4] In the current boom-and-bust cycle, China is the largest bilateral creditor by far, but it remains to be seen whether China will follow that route. Resolution of international debt negotiations requires that creditors accept losses and recognize that ultimate repayment from many of its borrowers will be partial. To be clear, this acknowledgement took US banks a decade in the 1980s, as they waited for their balance sheets to improve. Given the scale of the stress on Chinese bank balance sheets from impaired domestic assets (housing in particular), the wait for China's banks to acknowledge losses is poised to again be long. These delays are unambiguously problematic for debtors in or near distress. Still, some signs of progress have emerged in Zambia's 2024 debt restructuring with China and its other creditors, which implied a present value "haircut" (creditor loss) of more than 50 percent—that is, the new debt contracts negotiated to replace the earlier ones are only worth half of the original loan value.

The Boom

Figure 2 summarizes China's rapid rise as a global creditor by comparing the aggregate claims of different official creditors. Throughout this paper, we will refer to "China" or to "Chinese official creditors" to denote the full set of state-owned Chinese entities that extend cross-border loans. Our definition follows the widely used OECD definition of official creditors and includes the Chinese government, China's central bank (the People's Bank of China or PBOC), and all state-owned banks and enterprises, even though several Chinese banks—most prominently China Development Bank—deny that they are official creditors and self-identify as private creditors.

In the past two decades, the Chinese government and its state-owned banks and enterprises extended more than $800 billion in loans to emerging market and developing country economies to finance infrastructure, energy, and mining projects (Malik et al. 2021; Dreher et al. 2022; Amendolagine, Presbitero, and Rabellotti 2024). As a result of this rapid credit expansion, China's outstanding loan portfolio of foreign governments surpassed those of the World Bank, the International Monetary Fund, and all 22 Paris Club creditor governments combined during the mid-2010s at the zenith of the lending cycle. When the World Bank scaled up its lending during the COVID-19 pandemic, China ceded its position as largest official

---

[4]Bonds were certainly not new to emerging market and developing countries, as foreign lending dominated by Britain in the nineteenth century and by the United States in the early twentieth century was exclusively bond led. For a discussion of the historical evolution of bond and loan finance in emerging markets, see Meyer, Reinhart, and Trebesch (2022).

Figure 2

Official Bilateral and Multilateral Creditors for Emerging Market and Developing Countries: 2000–2021 (billions, US dollars)
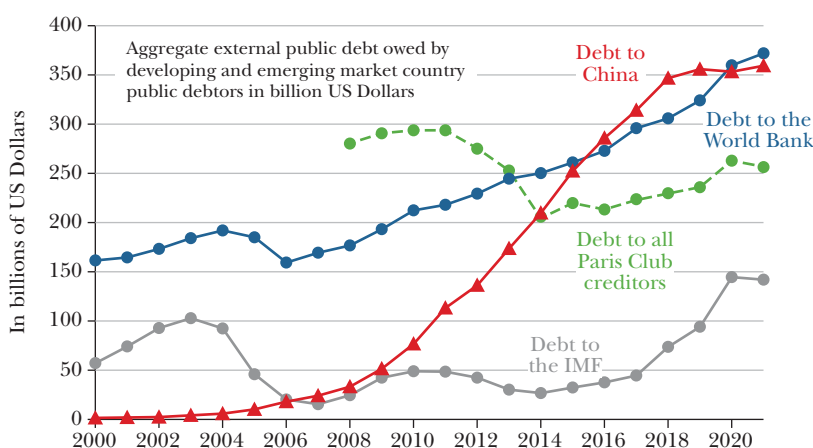
Source: Horn, Reinhart, and Trebesch (2021), Horn et al. (2023a), Franz et al. (2025), Custer et al. (2023), Dreher et al. (2022), IMF (2025), World Bank (2025), and Paris Club (multiple years).
Note: This figure plots aggregate external public and publicly guaranteed debt by emerging markets and developing countries to different official or state creditors. Debt owed to the World Bank (IDA plus IBRD), debt to the Paris Club, and debt to the IMF are taken from creditor sources. Debt owed to China is estimated by Franz et al. (2025) following the approach outlined in Horn, Reinhart, and Trebesch (2021). For a comparison of the scale of China's lending with private creditors, see Table 1.

creditor to the developing world. Still, the striking comparison shown in Figure 2 does not do justice to China's role as a global creditor. Although our focus here is on China's lending to developing economies, China's lending to advanced economies—notably the United States, through People's Bank of China purchases of US Treasury and other advanced economies' government securities—far surpassed the dollar value of loans to emerging market and developing countries shown in Figure 2.[5]

While Figure 2 places China's lending in perspective relative to other official creditors, it is silent on how China compares to private creditors to emerging market and developing countries. Table 1 fills this gap. As the table highlights, there are considerable differences across income groups. The share of public and publicly guaranteed debt owed to China to total public and publicly guaranteed debt has risen substantially in all three groups and is higher in low-income countries. The differences in the relative importance of private creditors across income groups is far more stark: for the low-income countries, debt owed to foreign private creditors

[5] Figure 2 also does not cover the claims of Chinese banks on private debtor entities in emerging market and developing economies, including rapidly growing lending to special purpose vehicles in the mining, infrastructure, and energy sector (Malik et al. 2021).

Table 1

Public External Debt: Share Owed to China versus Private Creditors (Public and Publicly Guaranteed Debt of Emerging Market and Developing Countries)

	Income group		
	Low	*Lower-middle*	*Upper-middle*
Number of countries in sample	27	47	44
Average percent of PPG debt owed to China			
2000–2009	2.4	1.9	1.5
2010–2021	15.5	14.6	11.1
Average percent of PPG debt owed to private creditors			
2000–2009	2.9	11.6	32.9
2010–2021	6.4	16.8	40.0
Share of countries which issue PPG bonds	17.2	59.6	86.4
Share of countries where debt to China > debt to private creditors	70.4	40.4	18.2

Source: Horn, Reinhart, and Trebesch (2021), Horn et al. (2023a), Franz et al. (2025), and World Bank (2024).
Note: The table focuses on external public debt stocks. "PPG" refers to public and publicly guaranteed debt, externally held. External private creditor totals combine debts to commercial bank creditors, bondholders (where applicable), suppliers' credits, and other private creditors. Numbers are for 2021, unless noted otherwise.

accounts for only about 6 percent of the total versus 40 percent for the upper-middle-income group; and for the lower-middle-income group at around 17 percent. The last row in Table 1 focuses on the incidence of China dominance versus private financing. In 70 percent of low-income countries, debts owed to China were larger than private creditor debt while the same pattern holds true for 40 percent of lower-middle-income group countries. Even in the upper-middle-income group, almost one-fifth of the countries have larger public debts with China. To sum up: for low-income countries as a group, China is a dominant creditor, outstripping private and other bilateral creditors. Even for the higher income groups, China remains the top creditor in a substantial number of countries.

This analysis also speaks to the large variation in loans versus bonds across emerging market and developing economies. The secular rise in national-level bond finance in the past three decades has largely bypassed low-income countries, only 5 countries out of the 29 in that group have external bonds.[6] While the number of lower-middle-income countries tapping international bond markets soared, as investors sought higher yields in the low international interest rate environment

---

[6] Table 1 lists 27 countries in the sample, as no debt data were reported in World Bank sources for South Sudan and North Korea. However, we are not aware that these countries have placed any internationally traded bonds.

of 2009–2022, about 40 percent of these countries continue to rely exclusively on loans. The dominance of bond finance is most evident in the higher-income emerging market and developing economies, such as Brazil, Mexico, South Africa, and Turkey, among others.

Several factors, within China and in the rest of the world, combined to fuel the lending bonanza and, years later, to bring it to an end. In China, record levels of investment and domestic savings had propelled real GDP growth, which averaged almost 10 percent per year during 2003–2015 (IMF 2024). China's appetite for primary commodities fueled a global commodity price boom—the second-longest since the late eighteenth century.[7] At the same time, China accumulated about $3 trillion in foreign exchange reserves during 2003–2015, supported by large current account surpluses and an undervalued exchange rate (according to IMF 2025). This reserves build-up mostly went into purchases of US Treasury securities, but also positioned China to accelerate its lending to emerging market and developing countries.

In 2013, Chinese President Xi Jinping announced the Belt and Road Initiative, with a broad goal of improving transportation links by road, rail, and sea—along with energy pipelines—ultimately improving China's connections to Asia, Africa, and Europe. At this point, China's lending boom had already been underway for about a decade. Its overseas lending was directed towards low- and lower-middle-income countries as well as primary commodity producers (like Angola, Ecuador, Russia, or Venezuela). When growth in China began to slow and overcapacities began to emerge, direct loans to developing countries—often tied to the procurement of Chinese goods and services—became an even more attractive outlet for China's US dollar reserves. These loans also helped to secure access to raw materials and large foreign markets for Chinese state-owned enterprises.

The recycling of China's US dollar reserves shares some similarities of the "recycling of petrodollars" in the 1970s. "Petrodollars" referred to the revenues earned by oil-exporting countries, which were commonly denominated in US dollars. Following the oil shocks of the 1970s, petrodollar deposits (from oil producers) in global financial center banks soared. This expansion fueled the surge in bank lending to emerging market and developing economies at that time. One difference is that China's US dollar recycling process was entirely handled by state-owned entities and spearheaded by the country's two major policy banks: China Export-Import Bank and China Development Bank.

Conditions in emerging market and developing economies in the first 10–15 years of the twenty-first century were also conducive to the borrowing bonanza. Many commodity-producing countries benefited from growing "hard currency" export earnings—again, often based on US-dollar-denominated export sales—that provided them with the means to service higher levels of external debt denominated in US dollars (at least while commodity prices remained high).

---

[7] The duration of the 1999–2011 boom in global commodity prices was only surpassed by the 1938–1951 boom (Reinhart, Reinhart, and Trebesch 2016).

In addition, public debt levels of many emerging market and developing economies were comparatively low early in the twenty-first century, partly due to the earlier debt relief operations such as the Highly Indebted Poor Country initiative, which in the late 1990s and early 2000s wrote off most of the external debt owed to traditional bilateral and multilateral creditors. Additionally, private creditors also wrote off debt owed by many of these countries (Cruces and Trebesch 2013). After two decades of serial crises and debt overhang, economic growth in emerging market and developing economies increased, as did the need to finance infrastructure and social programs. In this context, recipient countries in the developing world welcomed China's twenty-first century overseas lending with open arms. The lending further fueled economic growth, at least in the short run (Dreher et al. 2021; Müller forthcoming).

Low global interest rates were a key factor as well. Rates had already been declining during the 2000s and fell further after the global financial crisis of 2007–2009. The low expected returns on advanced country government bonds contributed to a global "search for yield," and China sought higher returns abroad by lending to countries that had little or no access to international capital markets. As noted, China was not alone. International investors were increasingly drawn to bonds issued by the governments of what became known as "frontier markets" (in the terminology of the IMF *Global Financial Stability Report*, October 2019). Many of these frontier markets issued dollar-denominated bonds in international markets for the first time in their histories (Graf von Luckner et al. 2025). External debt levels began to climb rapidly, bolstered by the expectation that *this time was different* and the higher commodity prices and economic growth would continue, echoing the overoptimism of the late 1970s (James 1996).

The flipside of China's lending boom was a rapid surge in debt burdens. As documented in Table 1, China's share in external debt also became more dominant. By our estimates, around 20 developing countries have debt totaling more than 10 percent of their GDP to Chinese state creditors. Among this group are commodity exporters in Africa and Latin America, including Angola (29 percent of GDP owed in debt to Chinese state creditors) and Zambia (28 percent). Among the countries with the largest exposures are also those in close geographic proximity to China and geographically important members of the Belt and Road Initiative such as the Kyrgyz Republic (21 percent), Laos (33 percent), and Pakistan (10 percent), as well as several smaller economies across the globe such as Djibouti (36 percent), the Maldives (27 percent), or Vanuatu (17 percent). The list also includes several countries from earlier Highly Indebted Poor Countries debt relief initiative, who benefitted from large-scale official debt relief in the 2000s and quickly releveraged with financing from Chinese and other commercial creditors (Horn, Reinhart, and Trebesch 2021; Kose et al. 2021). It is important to add that China's lending also included countries that had previously been shut out of private international capital markets. In more than 20 countries of the two lower-income groups (that is, both low-income and lower-middle-income countries) debt to private creditors is nil, comprising less than 1 percent of their total external debt.

Table 2
Financial and Nonfinancial Terms of China's Overseas Lending

Interest rates	Most of Chinese state lending comes on commercial terms. On average, Chinese creditors charge a margin of 250 basis points over a LIBOR reference rate. Some loans—particularly those that come with positive externalities for China's economy—contain significant subsidies. They typically charge fixed interest rates of 2 or 3 percent.
Currency	Most of the lending (75 percent) is denominated in US dollars. Only a small share is denominated in renminbi, the euro or local currencies.
Purpose	The largest share of Chinese lending supports development projects in the infrastructure, mining, or energy sector and is tied, that is, needs to be used for procurement of specific goods and/or services from Chinese suppliers and contractors.
Collateral	Around 50 percent of lending is collateralized with liquid, easily accessible assets. In the typical arrangement, the recipient government or state-owned enterprise commits to route foreign currency proceeds from commodity sales to an offshore bank accounts controlled by the lender. In the case of a default, the creditor bank can seize the cash accumulated in these accounts.
"No Paris Club"	Chinese loan contracts often include "no Paris Club" clauses which require the recipient to keep the debt out of collective restructuring efforts and not to seek restructurings on comparable terms. This clause effectively gives Chinese creditors discretion to decide bilaterally if, when, and how to restructure debt.
Confidentiality	Chinese loan contracts often include far-reaching confidentiality clauses that bar the borrower from revealing the terms or even the existence of the debt.

Source: Horn, Reinhart, and Trebesch (2021), Gelpern et al. (2023, 2025), Custer et al. (2023), and Dreher et al. (2022).
Note: This table summarizes key financial and nonfinancial features of China's overseas lending portfolio.

For low-income countries, interest payments on external debt to all creditors as a percent of exports of goods and services stood at 5.7 percent in 2023 versus 1.9 percent a decade earlier. For Egypt, Kenya, and Pakistan (middle-low-income), these amount to about 13 percent in 2023 (World Bank, *International Debt Statistics* 2024).

How China Lends

How do Chinese state creditors lend overseas? While there is considerable heterogeneity across lenders, what stands out is that much of China's state-driven overseas lending resembles commercial bank lending transactions. In terms of the debt instrument of choice (in this case, loans), this is quite distinct from emerging market and developing economies borrowing from the private sector through the international placement of bonds; however, it is in line with the loan-driven approach of other official creditors (bilateral and multilateral). Table 2 lists some of the key features. Most of China's loans are denominated in US dollars and interest rates tend to be nonconcessional, where a "concessional" loan has either a below-market interest rate, an extremely long maturity, or both. However, China's

loans have an average risk premium of more than 250 basis points over a London Interbank Offered Rate reference rate. For emerging markets and middle-income countries, most Chinese official loans are extended at market terms, meaning with interest rates that are close to those prevailing in private bond or loan markets. That said, there is also considerable heterogeneity in the financial terms of the loans, with substantial subsidies in loans provided for political prestige projects, for the benefit of local elites, and for trade infrastructure that yields positive externalities to Chinese exporters (Franz et al. 2025; Kern, Reinsberg, and Shea 2022).

Another striking feature of China's lending are contractual provisions to seek "seniority" over other creditors, which means that repayment of loans to China would take priority over other debt if the sovereign debtor cannot meet all its outstanding debt obligations (Schlegl, Trebesch, and Wright 2019). As noted earlier, China's loans make widespread use of collateral. This is another important departure from the lending practices of other official creditors. Around 50 percent of all cross-border lending is collateralized, most importantly through the proceeds of recipient country commodity exports, such as oil, that debtors need to deposit in offshore bank accounts and to which Chinese creditors have access (Gelpern et al. 2025). Other common contractual provisions bar debtors from including debts to China in collective restructuring efforts—say those organized through the Paris Club—and therefore give Chinese creditors scope to negotiate preferential treatment during debt restructuring negotiations (Gelpern et al. 2023).

These financial and legal features would be unusual for official international loans extended by OECD governments and Paris Club member countries during the past decades, which typically offer concessional financial terms, feature less stringent clauses, and are not backed by collateral. Indeed, one needs to go back to the nineteenth and first half of the twentieth centuries to find similar lending terms and the widespread use of collateral in sovereign debt contracts, at least on the scale that we observe in the Chinese lending in this century (Wynne 1951; Gelpern et al. 2023; Flandreau, Peitrosanti, and Schuster 2022).

Lack of Transparency and Hidden Debt

China's loans also often include unusual confidentiality provisions. While it is common for international loan contracts to include some nondisclosure provisions, some of the clauses used by Chinese creditors are unusual in their scope and in the fact that they bind the *debtor* country rather than the creditor, as is commonly the case in the lending relationship between a bank and its customer. The incidence of broad confidentiality clauses in Chinese loan contracts has increased markedly over time (Gelpern et al. 2023).

The use of confidentiality clauses is consistent with the overall treatment of China's overseas lending activities as a "state secret" (Bräutigam 2009, p. 2). Unlike most other major official lenders, the Chinese government does not release creditor statistics on the magnitude and composition of its overseas lending portfolio and is not

a member of the usual data-sharing arrangements such as the OECD Development Assistance Committee or the Paris Club. At the recipient end of the transactions, more than half of China's lending has been taken up by state-owned enterprises and special purpose vehicles. In high-risk countries with weak institutions and low debt management capacity, governments often have limited oversight on this type of public borrowing. The result is a lack of consolidated fiscal accounts and incomplete debt statistics.

Why are prospective borrowers willing to agree to confidentiality? As previously noted, numerous high-risk borrowers had no access or the prospect of access to private finance. Furthermore, official bilateral creditors scaled back their lending following the significant debt write-offs under the Highly Indebted Poor Countries program in the late 1990s and early 2000s.[8] For some of these countries, China was the only option. Beyond these extreme cases, theoretical models that consider the optimal allocation of sovereign bonds versus loans stress that the critical difference between bank lending and bond finance is that banks (in this case China's development banks) act as private monitors and keep their assessment of the borrower private, whereas rating agencies (having a rating is the usual prerequisite for issuing bonds) act as public monitors and disseminate this information not only to the existing bondholders but also to third parties—including potential future creditors (Tanaka 2006). Risky borrowers may welcome the confidentiality offered by Chinese lending. To illustrate this point, consider the case of Angola, which has restructured loans with different Chinese creditors four times since it exited from a default that also involved private creditors in 2003. Had those loan restructurings been instead on their privately held government bonds, one may well imagine that its credit rating (currently B−) would be lower still. Angola would have also faced the potential for bondholder litigation that Argentina and other sovereigns have faced in recent years.

A large share of China's overseas lending flows initially went unreported and created sizeable knowledge gaps (Horn, Reinhart, and Trebesch 2021). By the end of the lending boom, in 2019, around 50 percent of China's lending to emerging market and developing country governments had not been reported in the World Bank's main database on external public debt, the *International Debt Statistics*. The volumes of hidden debt were particularly large in around two dozen developing countries that had heavily borrowed from China (Horn, Reinhart, and Trebesch 2021). Hidden debt can create significant welfare losses for debtor countries: the emerging literature on the effects of hidden debt includes Alfaro and Kanczuk (2022), Kondo, Sosa-Padilla, and Swaziek (2025), Guler, Önder, and Taskin (2025), Gamboa-Arbalez (2023), and Horn et al. (2024). Uncertainty about the true level of external indebtedness undermines surveillance, leads to biased assessments of default risks (for example, in the International Monetary Fund and the World Bank debt sustainability assessments) and distorts asset pricing. Countries with

---

[8]The debt build-up in many advanced economies during the past decade may have also contributed to the declining appetite for overseas concessional lending.

underreported debt may, for example, issue sovereign bonds at a lower coupon rate than would have prevailed had the true extent of indebtedness of the borrower been known to the underwriter of the bond and the prospective buyers. As long as hidden debts remain undetected, cheaper rates will fuel overborrowing and thereby raise the risk of eventual default.

Hidden debts are neither new nor limited to developing countries or to state creditors, as shown in the systematic analysis by Horn et al. (2024). Hidden debt scandals have happened in Europe, such as in Greece 2009–2010, and they have involved many types of creditors. One example was Mozambique's "tuna bonds scandal" in which state-owned enterprises used government guarantees to borrow $1.2 billion from international banks Credit Suisse and VTB to finance a tuna-fishing project without disclosing the liabilities to bond investors, the International Monetary Fund, or the broader public. When the hidden debts came to light in 2016, the IMF and international donors suspended their support of the country, and Mozambique defaulted on its bondholders and triggered a protracted economic crisis. However, the widespread nature of the hidden debt phenomenon in China's overseas lending to its many borrowers is, to the best of our knowledge, historically unique in its scale.

In many countries, the full extent of external indebtedness to China only became evident when crises hit, or when the countries had to turn to the International Monetary Fund for assistance. Amid rising debt distress risks circa 2015, the official debt statistics of emerging market and developing economies became the subject of intense scrutiny by policymakers and academic researchers, resulting in more transparency, especially regarding China's lending. Since 2019 more than $600 billion in previously unrecorded loans—by Chinese and other creditors—have been added to the World Bank debt statistics. This large upward revision significantly contributed to the record debt levels that many countries faced at the height of the pandemic (Horn, Mihalyi, and Nickol 2022; Horn et al. 2024). Despite this progress in debt-data coverage, severe reporting gaps persist. The debt data of loans channeled through "special purpose vehicles" or tax havens, which are often explicitly designed to remain off the sovereign's balance sheet and that constitute an increasing share of China's overseas lending portfolio (Malik et al. 2021), are particularly incomplete.

The Bust: Rising Debt Distress and the Chinese Lending Reversal

In 2014–2015, non-oil commodity prices fell sharply to about 30 percent of their peak a few years earlier. Around this period, external debt had climbed, and growth slowed for many of China's borrowers. Sovereign credit rating downgrades became more commonplace, as shown in Figure 3, and the International Monetary Fund and World Bank began to classify a rising share of low-income countries as being at a high risk of debt-distress and default. China's growth had begun to slow as well, with an average yearly growth rate of 5.6 percent during 2016–2024 versus 9.8 percent for 2003–2015. China's foreign exchange reserves no longer grew, peaking in 2014 (IMF *World Economic Outlook*, October 2024, and *International*

Figure 3
Deteriorating Credit Quality: Sovereign Risk of China's Lending Portfolio
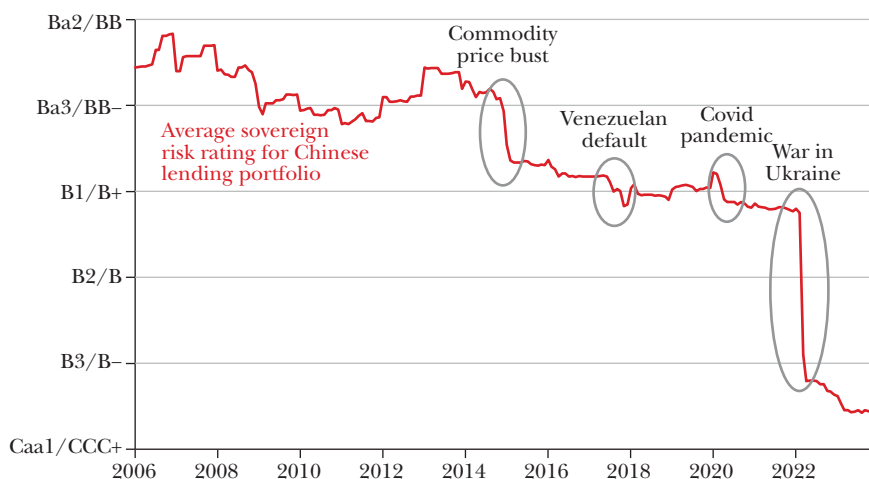
Source: Horn, Reinhart, and Trebesch (2021), Franz et al. (2025), and Custer et al. (2023).
Note: This figure shows the average sovereign credit risk ratings for a balanced panel of 112 developing and emerging market countries, for which data are available for the full sample period. The credit risk rating is an average across ratings from Moody's, Fitch, S&P, the Institutional Investor Rating (IIR), and the World Bank-IMF LIC DSF rating. To ensure comparability across rating scales and sources, all ratings are mapped to a numerical scale from −4 to 20 following the approach of Reinhart, Reinhart, and Trebesch (2017). The average rating of China's overseas lending portfolio was obtained by weighting ratings with estimated outstanding debt claims.

Financial Statistics). Hence, both "push" factors from China and "pull" factors from the borrowers signaled that the end of the boom phase of the cycle was near.

During the past ten years, China's overseas lending has been exposed to a series of global shocks and rising levels of default risk. The "commodity price shock" was followed a few years later by the COVID-19 pandemic, which brought about a deep and synchronous global contraction in economic activity and trade. The G20's Debt Service Suspension Initiative (DSSI) was an attempt to respond to the dislocations, allowing participating low-income countries to suspend payments to their official bilateral creditors between May 2020 and December 2021 (although private creditors refrained from participating in the DSSI). However, the DSSI was a temporary moratorium and not designed to resolve the problem of unsustainable debts. The pandemic also markedly worsened public finances in developing and advanced economies alike, as government revenues plummeted—even for the many low-income countries that could not afford fiscal stimulus packages. The result was a rapid build-up in domestic debt, adding further strains to debt servicing. On the heels of the pandemic came the Russia-Ukraine War, which drove imported food prices markedly higher—further straining the finances of many borrowers compromised China's overseas claims.

While these shocks affected all emerging market and developing economies, China's overseas lending has been particularly hard hit, as shown in Figure 3, reflecting China's subprime lending profile and its strong exposure to some of the world's riskiest sovereign borrowers during the boom years. The average quality of China's lending portfolio has always been significantly lower—by about two credit-rating notches—than a GDP-weighted benchmark portfolio of all rated emerging market and developing economies. After 2014, however, China's average portfolio rating declined much more than the benchmark. One explanation is that Chinese banks are particularly exposed to commodity exporters, and more exposed to geopolitical risk, with a comparatively large share of the portfolio supporting autocratic governments in geopolitically contested parts of the world. The default of Venezuela's oil-exporting, autocratic government in 2014 had a substantial impact on China's average portfolio quality, as did Russia's attack on Ukraine in 2022, due to China's large lending volumes to Russia, Belarus, and Ukraine. As of 2024, China's overseas lending portfolio is rated at B− on average, which is deep in "junk bond" territory and three notches below a GDP-weighted emerging market and developing economy benchmark.

The rating decline does not necessarily imply that a large share of China's portfolio is in default. In fact, there is a substantial knowledge gap on what happens to Chinese claims in situations of debt distress and default.

In the case of international bonds, failures to repay and reschedulings of debt are known "credit events" covered by credit rating agencies. However, credit events with Chinese (and other official) creditors are not collected and disseminated by the large international credit rating agencies or international organizations. We therefore lack details on the dates and scale of defaults vis-à-vis Chinese state creditors. We also do not know how large any "hidden" debt repayments are. As discussed earlier, many Chinese loans include innovative contractual designs with elaborate safeguards against financial and political risk. For example, debts to China may continue to be serviced through the proceeds of commodity exports that have been routed to (offshore) escrow accounts that Chinese creditor banks can seize in the case of default (Gelpern et al. 2025). In this sense, Chinese banks may be poised to deal with rising default risk better than other creditors and enjoy a higher degree of seniority. Venezuela, for example, notwithstanding its collapse of economic activity since 2017, has continued to partially service China's debts by shipping oil while in default on its external private creditors.

As default risk rose among emerging market and developing economies, Chinese banks briskly curtailed their cross-border lending, especially since 2018, as shown earlier in Figure 1, which illustrated that net transfers between Chinese lenders and emerging market and developing economies have turned increasingly negative for five years in a row. In other words, principal and interest payments to Chinese creditors now clearly exceed fresh lending disbursements. In writing about private capital flows and crises in emerging markets, Calvo (1998) has dubbed this phenomenon a "sudden stop," reflecting the deteriorating economic or political fundamentals among many borrowers. In addition, Chinese banks also face growing pressures at home, including a crisis in the housing market, the insolvency

or near-insolvency of some provinces (which also relied heavily on bank credit), and slower economic growth (Rogoff and Yang 2024).

In hindsight, China's overseas lending boom looks remarkably procyclical, as was the case with US commercial bank lending in the 1970s and more generally with private capital flows historically. Procyclical overseas lending fuels booms, but makes busts more costly, in this way adding to the amplitude and volatility of the economic cycle. It also contributes to fiscal procyclicality, as reduced access to international finance during bad times may require governments to reduce fiscal outlays and/or rely more on inflationary finance (Kaminsky, Reinhart, and Végh 2004; Bianchi, Ottonello, and Presno 2023).
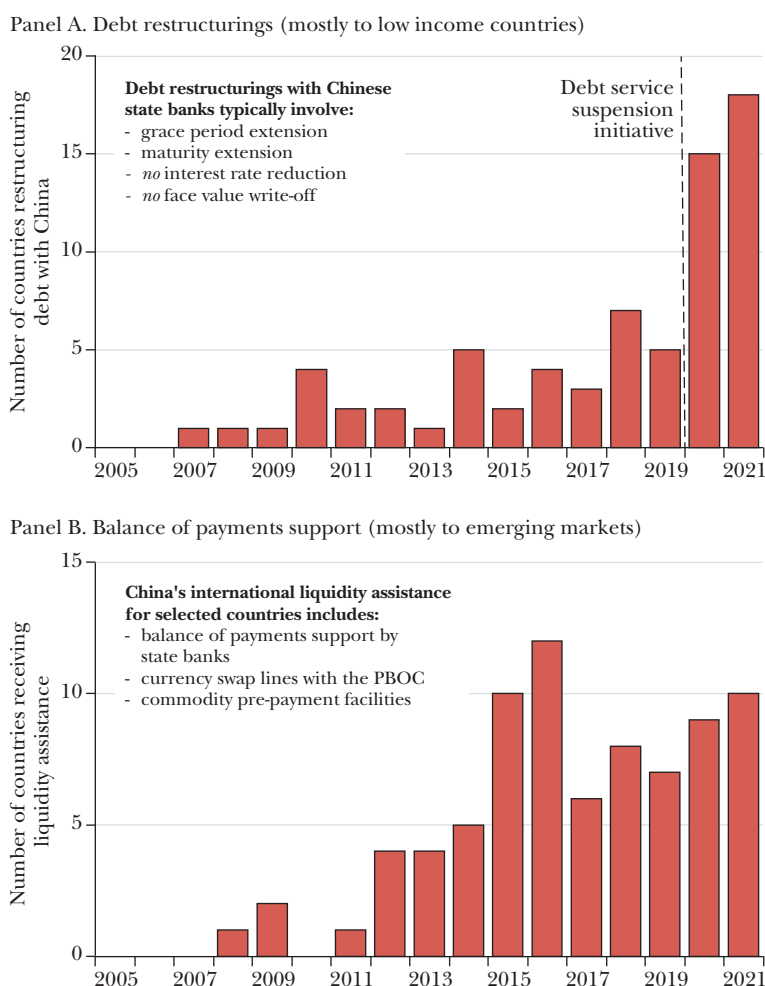
China lent record amounts when global commodity prices were high and international interest rates were low, but then slowed its lending and turned towards debt collection as global risks increased and growth slowed. China's "sudden stop" may amplify the downturn in recipient economies and may make it difficult for countries to refinance the large amount of Chinese lending that is now coming due. In its cyclical lending pattern, as well as in its terms of lending (shown earlier in Table 2), China more closely resembles private creditors rather than its official counterparts. As shown in Horn, Reinhart, and Trebesch (2020), the procyclicality of Chinese lending is at odds with the countercyclical lending practices of the dominant multilateral institutions, the International Monetary Fund, and the World Bank. For example, the international institutions ramped up lending during the COVID-19 years, while China retrenched (as shown earlier in Figure 2). Countries rarely, if ever, turn to IMF loans in good times.

Resolving Debt Distress and Default: China's Early Steps

How does China respond to the threat of default, or actual missed payments, on its loans? As noted, its response has closely resembled that of (Western) private creditors during previous debt crises, including in the 1980s. Then and now, the primary aim of the initial policy response has been to contain "collateral damage" for debtor countries, but especially for creditor banks. To do so, China's policy mix has included both debt restructurings and, more selectively, the provision of short-term financing. Low-income countries in distress are typically offered debt restructurings that involve maturity or grace-period extensions, but no new money. In contrast, selected distressed middle-income countries, where Chinese exposure is markedly greater, have also received new financing (Horn et al. 2023b). As the adage goes: "If you owe the bank a little money, that's your problem. If you owe the bank a lot of money, that's the bank's problem." Importantly, whether new money is offered or not, the debt restructuring agreements have seldom involved a write-off of principal and often do little to reduce the debt overhang.

Figure 4 describes China's policy response in greater detail. Panel A shows that debt restructurings with Chinese creditors have surged notably. Since 2008, Chinese creditors arranged at least 71 distressed debt restructurings—more than three times

Figure 4

Debt Restructuring and Liquidity Support, 2005–2021: China's Response to Debt Distress

Panel A. Debt restructurings (mostly to low income countries)

Debt restructurings with Chinese state banks typically involve:
- grace period extension
- maturity extension
- *no* interest rate reduction
- *no* face value write-off

Debt service suspension initiative

Panel B. Balance of payments support (mostly to emerging markets)

China's international liquidity assistance for selected countries includes:
- balance of payments support by state banks
- currency swap lines with the PBOC
- commodity pre-payment facilities

Source: Horn, Reinhart, and Trebesch (2022) and Horn et al. (2023a).
Note: This figure shows China's international response to debt distress by showing the number of emerging market and developing countries that have restructured their debts with Chinese state creditors (panel A) or that have received short-term liquidity assistance from China (panel B).

the number of sovereign restructurings with private external creditors and higher than the total number of Paris Club restructurings with distressed debtors during the same period (Horn, Reinhart, and Trebesch 2022). China has become the most important official player in international sovereign debt renegotiations. The modalities of the debt restructuring of Chinese lenders are also reminiscent of those of western bankers in the early 1980s. So far, Chinese lenders have almost never

provided deep debt relief by reducing the face value or the interest rates of the debt, except for symbolic debt cancelations of minor zero-interest loans. Instead, the deals have provided short-term cash flow relief by extending the maturity or grace periods. This has resulted in in serial debt restructurings—repeated restructurings of the same debt with the same sovereign creditor—that have also characterized the lengthy resolution processes of the 1930s and 1980s debt crises. Among numerous others, Angola, Ecuador, Seychelles, Sri Lanka, and Venezuela have already had two or more restructurings of debts owed to China.

Another parallel to Western creditors in the 1980s is that China is also increasingly engaging in international lending of last resort, by providing countries in its economic and political sphere of influence with short-term liquidity assistance. In Horn et al. (2023a), we document that China's international rescue lending is growing fast (panel B of Figure 4), even though aggregate net flows have been increasingly negative (repayment) in most countries since 2019. In the past 15 years, China has extended more than $250 billion in short-term financial assistance, including many roll-overs to the same group of countries. This repeated provision of financing is reminiscent of the International Monetary Fund's "serial lending" practice in recent decades (Reinhart and Trebesch 2016b) and the bridge credits by private creditors prevalent during the 1980s debt crisis (Cruces and Trebesch 2013).

Financial assistance has come through three distinct lending arrangements. First, state-owned Chinese banks have selectively offered balance-of-payments support and budget support loans that explicitly authorize the recipient countries to use the funds to repay existing debts. Second, Chinese state-owned enterprises have prepaid for large future deliveries of commodities, in particular oil, providing funds that recipient countries have used to finance budget deficits. This step is also reminiscent of historical initiatives by Western governments, like the US oil prepayments to Mexico during its debt crisis of 1982 (Boughton 2001). Third, China's central bank, the People's Bank of China, operates an international swap line network that is increasingly used by highly indebted Belt and Road Initiative countries. Recipient country central banks have used the swap lines to address short-term liquidity needs, to bolster gross foreign exchange reserves, and to repay external debts (Horn et al. 2023a). PBoC swap line drawings are typically scheduled for repayment in twelve months or less, but repeated rollovers have implied long de facto average maturities of 3.5 years, making them different from the very short-term US dollar swaps extended by the US Federal Reserve.

Taken together, China's policy response usually offers temporary cash-flow relief but has so far done little to address underlying solvency issues and to spur the economic recovery. Repeated restructurings and "rescue" lending at high interest rates have shifted repayment obligations to the future instead of providing substantial reduction in the present value of debt. As so often in the initial stages of a debt crisis, Chinese creditors reacted as if the borrowers faced a temporary liquidity shortage rather than a long-term solvency problem.

If, indeed, the problem for the borrowing government is its inability to make an imminent payment on principal or interest in accordance with a contract but

is otherwise thought to be capable of meeting future obligations (that is, they are solvent), then the creditor will usually allow for a minimalist restructuring that offers a grace period when no payment or partial payment, say of interest only, is due (for a discussion of the "typical debt resolution pattern", see Reinhart 2022). There is no reduction in the interest of the loan, and there may be no lengthening of maturity and no write-off of principal. If the borrower is important to the creditor and/or there is confidence in future repayment, the restructuring may include as part of the agreement some new money, often in the form of a bridge loan.

If the borrower's debt burden, however, is fundamentally unsustainable—that is, it cannot be serviced under realistic assumptions about future growth and fiscal policy—rollovers and payment deferrals are insufficient to restore economic growth. The legacy debt overhang and the prospect of large future repayments discourage new investments and result in protracted stagnation (Myers 1977; Krugman 1989). In this case, nominal debt write-offs and "deep" debt relief are typically needed to spur economic recovery (Reinhart and Trebesch 2016a).

China's approach so far may reflect the assessment (or hope) that the debt servicing difficulties are transitory. China cannot be singled out in this regard, as serial restructurings and serial "bridge loan-type" lending has been commonplace in past crises. In the 1980s, Western creditor banks and the Paris Club refused for a decade to accept write-downs on their loans to emerging market and developing economies, and instead arranged dozens of restructurings and refinancing facilities with little debt relief, which did not cure the debt overhang and instead resulted in a lost decade of growth (Reinhart and Trebesch 2016a).

China's approach also reflects the increasingly fragmented creditor landscape and the severe coordination issues that have paralyzed multilateral debt resolution efforts in recent years. While it is optimal for creditors to *jointly* provide deep debt relief in the face of debt overhang, the unilateral provision of debt relief is delayed by collective action problems (Sachs 1989). Indeed, a key concern of both Chinese and Western creditors has been that other creditors may free-ride on their concessions—and vice versa. Overcoming these coordination issues will further delay resolution—notwithstanding some recent progress in the context of the Common Framework.

It is too early to assess whether the unfolding situation will result in a similar delay as in the 1980s and whether much of the developing world will suffer from another lost decade. As the evidence on past serial restructurings shows, however, a crisis resolution approach that treats insolvency as a liquidity problem is bound to fail (Graf von Luckner et al. 2025). If history is any guide, the debt difficulties in affected developing countries will take many more years to resolve.

The Road Ahead

While economic factors suggest China will be in no hurry to agree to systematic haircuts in its overseas loan portfolio, geopolitical factors may weigh in the other

direction. Tensions between China and the United States make daily headlines. The United States is moving in the direction of economic isolationism, with plans for higher tariffs for many countries and further plans to scale back on overseas assistance. In this setting, China may choose to seize the opportunity to strengthen its ties with emerging market and developing countries around the world by opting for a speedier timeline to offer deep debt relief to its distressed debtors.

Will China's overseas lending resume or will its retrenchment continue? If history is any guide, the experiences of creditors in the aftermath of debt crises suggest that a resurgence in lending is unlikely over the near term. Paris Club creditors never came fully back after the deep haircuts agreed to through the Heavily Indebted Poor Countries debt relief agreements of the 1990s. Neither did US banks. Going further back in time, it took more than 50 years between the end of the US-led bond lending boom of the "Roaring 1920s" and the resurgence of US external lending to developing countries in the 1970s. It seems reasonable to expect China's lending to developing country governments to remain subdued in the coming years. The data presented here suggest that, to date, China's overseas lending has been markedly procyclical.

Nevertheless, China will remain an important player in the sovereign debt markets of developing countries, given the large outstanding amounts and the sheer scale of China's footprint in the global economy. However, its future lending may be more selective, more strategic, and more responsive to recipient countries' risks. There are signs that Chinese state creditors are increasingly engaging with higher-income countries with lower credit risk.

Looking ahead, we expect further deep changes in the international financial system. In the past two decades, the pendulum swung from the traditional North-South official creditor-debtor model, as exemplified by the Paris Club, to China's South-South model, which represents a hybrid of official finance with many commercial creditor features. Many of the issues that we have discussed here are not unique to China but extend to other non–Paris Club creditors that have deep pockets and/or use state finance to further their economic and political objectives abroad. Further research on the overseas lending of Saudi Arabia and the United Arab Emirates is an excellent place to start. Because of the growing importance of these and other new creditor powers, there has been a steady shift to a more opaque, fragmented, and multipolar international financial architecture.

References

Alfaro, Laura, Sebnem Kalemli-Ozcan, and Vadym Volosovych. 2008. "Why Doesn't Capital Flow from Rich to Poor Countries? An Empirical Investigation." *Review of Economic and Statistics* 90 (2): 347–68.

Alfaro, Laura, and Fabio Kanczuk. 2022. "Undisclosed Debt Sustainability." *AEA Papers and Proceedings* 112: 521–25.

Amendolagine, Vito, Andrea F. Presbitero, and Roberta Rabellotti. 2024. "Chinese Infrastructure Lending in Africa and Participation in Global Value Chains." *Review of World Economics* 161 (1): 7–48.

Arslanalp, Serkan, and Peter Blair Henry. 2006. "Policy Watch: Debt Relief." *Journal of Economic Perspectives* 20 (1): 207–20.

Bianchi, Javier, Pablo Ottonello, and Ignacio Presno. 2023. "Fiscal Stimulus under Sovereign Risk." *Journal of Political Economy* 131 (9): 2328–69.

Bolton, Patrick, and Olivier Jeanne. 2009. "Structuring and Restructuring Sovereign Debt: The Role of Seniority." *Review of Economic Studies* 76 (3): 879–902.

Bulow, Jeremy, and Kenneth Rogoff. 1989. "Sovereign Debt: Is to Forgive to Forget?" *American Economic Review* 79 (1): 43–50.

Bräutigam, Deborah. 2009. *The Dragon's Gift: The Real Story of China in Africa.* Oxford University Press.

Calvo, Guillermo A. 1998. "Capital Flows and Capital-Market Crises: The Simple Economics of Sudden Stops." *Journal of Applied Economics* 1 (1): 35–54.

Clayton, Christopher, Matteo Maggiori, and Jesse Schreger. 2024. "A Framework for Geoeconomics." NBER Working Paper 31852.

Cline, William R. 1995. *International Debt Reexamined.* Institute for International Economics.

Cruces, Juan J., and Christoph Trebesch. 2013. "Sovereign Defaults: The Price of Haircuts." *American Economic Journal: Macroeconomics* 5 (3): 85–117.

Custer, Samatha, Axel Dreher, Thai-Binh Elston, Brooke Escobar, Rory Fedorochko, Andreas Fuchs, Siddhartha Ghose, et al. 2023. *Tracking Chinese Development Finance: An Application of AidData's TUFF 3.0 Methodology.* AidData.

Díaz-Alejandro, Carlos. 1985. "Good-Bye Financial Repression, Hello Financial Crash." *Journal of Development Economics* 19 (1–2): 1–24.

Dreher, Axel, Andreas Fuchs, Bradley Parks, Austin Strange, and Michael J. Tierney. 2021. "Aid, China, and Growth: Evidence from a New Global Development Finance Dataset." *American Economic Journal: Economic Policy* 13 (2): 135–74.

Dreher, Axel, Andreas Fuchs, Bradley Parks, Austin Strange, and Michael J. Tierney. 2022. *Banking on Beijing: The Aims and Impacts of China's Overseas Development Program.* Cambridge University Press.

Farah-Yacoub, Juan P., Clemens M. Graf von Luckner, and Carmen M. Reinhart. 2024. "The Social Costs of Sovereign Default." NBER Working Paper 32600.

Flandreau, Marc, Stefano Peitrosanti, and Carlotta E. Schuster. 2022. "The Puzzle of Sovereign Debt Collateral: Big Data and the First Age of Financial Globalization." CEPR Discussion Paper 17826.

Franz, Lukas, Sebastian Horn, Bradley Parks, Carmen M. Reinhart, and Christoph Trebesch. 2025. "Creditor Returns on China's Belt and Road." Unpublished.

Gamboa-Arbalaez, Juliana. 2023. "Hidden Chinese Lending." Preprint, SSRN. http://dx.doi.org/10.2139/ssrn.4617798.

Gelpern, Anna, Omar Haddad, Sebastian Horn, Paulina Kintzinger, Brad Parks, and Christoph Trebesch. 2025. "How China Collateralizes." CEPR Discussion Paper 20379.

Gelpern, Anna, Sebastian Horn, Scott Morris, Brad Parks, and Christoph Trebesch. 2023. "How China Lends: A Rare Look into 100 Debt Contracts with Foreign Governments." *Economic Policy* 38 (114): 345–416.

Graf von Luckner, Clemens M., Josefin Meyer, Carmen M. Reinhart, and Christoph Trebesch. 2025. "Sovereign Haircuts: 200 Years of Creditor Losses." *IMF Economic Review* 73 (1): 150–95.

Guler, Bulent, Yasin Kürşat Önder, and Temel Taskin. 2025. "Sovereign Debt Disclosure." *Journal of International Economics* 157: 104118.

Hatchondo, Juan Carlos, Leonardo Martinez, and César Sosa-Padilla. 2016. "Debt Dilution and Sovereign Default Risk." *Journal of Political Economy* 124 (5): 1383–422.

Horn, Sebastian, David Mihalyi, and Philipp Nickol. 2022. "Systematic Underreporting of Public Debt Statistics: 50 Years of Evidence and Recent Progress." World Bank blog, March 3. https://blogs.worldbank.org/en/developmenttalk/systematic-underreporting-public-debt-statistics-50-years-evidence-and-recent.

Horn, Sebastian, David Mihalyi, Philipp Nickol, and César Sosa-Padilla. 2024. "Hidden Debt Revelations." NBER Working Paper 32947.

Horn, Sebastian, Bradley C. Parks, Carmen M. Reinhart, and Christoph Trebesch. 2023a. "China as an International Lender of Last Resort." NBER Working Paper 31105.

Horn, Sebastian, Bradley C. Parks, Carmen M. Reinhart, and Christoph Trebesch. 2023b. "Debt Distress on China's Belt and Road." *AEA Papers and Proceedings* 113: 131–34.

Horn, Sebastian, Carmen M. Reinhart, and Christoph Trebesch. 2020. "Coping with Disasters: Two Centuries of International Official Lending. NBER Working Paper 27343.

Horn, Sebastian, Carmen M. Reinhart, and Christoph Trebesch. 2021. "China's Overseas Lending." *Journal of International Economics* 133: 103539.

Horn, Sebastian, Carmen M. Reinhart, and Christoph Trebesch. 2022. "Hidden Defaults." *AEA Papers and Proceedings* 112: 531–35.

International Monetary Fund. 2019. *Global Financial Stability Report.* International Monetary Fund.

International Monetary Fund. 2024. *World Economic Outlook—October 2024.* International Monetary Fund.

International Monetary Fund. 2025. *International Financial Statistics.* International Monetary Fund.

Isaksson, Ann-Sofie, and Andreas Kotsadam. 2018. "Chinese Aid and Local Corruption." *Journal of Public Economics* 159: 146–59.

James, Harold. 1996. *International Monetary Cooperation since Bretton Woods.* Oxford University Press.

Kaminsky Graciela, Carmen M. Reinhart, and Carlos A. Végh. 2004. "When it Rains, It Pours: Procyclical Capital Flows and Policies" In *NBER Macroeconomics Annual,* Vol. 19, edited by Mark Gertler and Kenneth Rogoff, 11–53. University of Chicago Press.

Kern, Andreas, Bernhard Reinsberg, Patrick E. Shea. 2022. "IMF Programs, Chinese Lending, and the Political Economy of Leader Survival." AidData at William and Mary Working Paper 118.

Kondo, Illenin O., César Sosa-Padilla, and Zachary Swaziek. 2025. "Borrowing in the Shadow of China." Unpublished.

Kose, M. Ayhan, Peter Nagle, Fraziska Ohnsorge, and Naotaka Sugawara. 2021. *Global Waves of Debt: Causes and Consequences.* World Bank.

Krugman, Paul. 1988. "Financing versus Forgiving a Debt Overhang." *Journal of Development Economics* 29 (3): 253–68.

Lew, Jacob J., Jennifer Hillman, Gary Roughead, and David Sacks. 2021. *China's Belt and Road: Implications for the United States.* Council on Foreign Relations Task Force Report 79.

Lucas, Robert E., Jr. 1990. "Why Doesn't Capital Flow from Rich to Poor Countries?" *American Economic Review* 80 (2):: 92–96.

Maçães, Bruno. 2018. *Belt and Road: A Chinese World Order.* C. Hurst US and Co.

Malik, Ammar A., Bradley Parks, Brooke Russell, Joyce Jiahui Lin, Katherine Walsh, Kyra Solomon, Sheng Zhang, Thai-Binh Elston, and Seth Goodman. 2021. *Banking on the Belt and Road: Insights from a New Global Dataset of 13,427 Chinese Development Projects.* AidData at William and Mary.

Meyer, Josefin, Carmen M. Reinhart, and Christoph Trebesch. 2022. "Sovereign Bonds Since Waterloo." *Quarterly Journal of Economics* 137 (3): 1615–80.

Mohr, Cathrin, and Christoph Trebesch. 2025. "Geoeconomics." *Annual Review of Economics* 17: 563–87.

Moura, Rodrigo. 2025. "Profit or Power: Determinants of Chinese Financing to the Developing World." *Journal of Contemporary China* 134 (125): 198–213.

Müller, Joris. Forthcoming. "The Domestic Political Economy of China's Foreign Aid." *Review of Economics and Statistics.*

Myers, Stewart C. 1977. "Determinants of Corporate Borrowing." *Journal of Financial Economics* 5 (2): 147–75.

Paris Club. Multiple years. *Amounts due to Paris Club creditor countries by foreign sovereign and other public debtors.* Paris Club.

Reinhart, Carmen M. 2022. "From Health Crisis to Financial Distress." *IMF Economic Review* 70 (1): 4–31.

Reinhart, Carmen M., and Vincent R. Reinhart. 2009. "Capital Flow Bonanzas: An Encompassing View of the Past and Present." *NBER International Seminar on Macroeconomics* 5 (1): 9–62.

Reinhart, Carmen M., Vincent Reinhart, and Christoph Trebesch. 2016. "Global Cycles: Capital Flows, Commodities, and Sovereign Defaults, 1815–2015." *American Economic Review* 106 (5): 574–80.

Reinhart, Carmen M., Vincent R. Reinhart, and Christoph Trebesch. 2017. "Capital Flow Cycles: A Long, Global View." Unpublished.

Reinhart, Carmen M., and Kenneth S. Rogoff. 2009. *This Time is Different: Eight Centuries of Financial Folly.* Princeton University Press.

Reinhart, Carmen M., and Kenneth S. Rogoff. 2011. "From Financial Crash to Debt Crisis." *American Economic Review* 101 (5): 1676–706.

Reinhart, Carmen M., Kenneth S. Rogoff, and Miguel A. Savastano. 2003. "Debt Intolerance." *Brookings Papers on Economic Activity* 34 (1): 1–74.

Reinhart, Carmen M., and Christoph Trebesch. 2016a. "The International Monetary Fund: 70 Years of Reinvention." *Journal of Economic Perspectives* 30 (1): 3–28.

Reinhart, Carmen M., and Christoph Trebesch. 2016b. "Sovereign Debt Relief and Its Aftermath." *Journal of the European Economic Association* 14 (1): 215–51.

Rogoff, Kenneth, and Yuanchen Yang. 2024. "Rethinking China's Growth." *Economic Policy* 39 (119): 517–48.

Sachs, Jeffrey D. 1989. "The Debt Overhang of Developing Countries." In *Debt, Stabilization, and Development: Essays in Memory of Carlos Diaz-Alejandro*, edited by Guillermo Clavo, Ronald Findlay, Pentti Kouri, and Jorge Braga de Macedo. Blackwell.

Schlegl, Matthias, Christoph Trebesch, and Mark L. J. Wright. 2019. "The Seniority Structure of Sovereign Debt." NBER Working Paper 25793.

Tanaka, Misa. 2006. "Bank Loans versus Bond Finance: Implications for Sovereign Debtors." *Economic Journal* 116 (510): C149—71.

World Bank. 2024. *International Debt Report 2024.* World Bank.

World Bank. 2025. *World Bank Finances One.* World Bank.

Wynne, William H. 1951. *State Insolvency and Foreign Bondholders: Selected Case Histories of Governmental Foreign Bond Defaults and Debt Readjustments.* Yale University Press.

Industrial Policy, Asian Miracle Style

Reda Cherif and Fuad Hasanov

In the second half of the twentieth century, a small group of East Asian economies—Korea, Taiwan, Singapore, and Hong Kong, following in the path of Japan before them—achieved a remarkable feat, starting from low levels of per capita income and joining the group of affluent *industrialized* economies within a couple of generations. Their experience is one of the most consequential pieces of evidence of what is possible in terms of economic development. Lucas (1993) described this process with an article titled "Making a Miracle." Similarly, a 1993 edited volume published by the World Bank on this subject was titled *The East Asian Miracle* (World Bank 1993). It seems well worth attempting to solve the riddle of the Asian "miracle" to allow others to walk in their footsteps, even if it looks *a priori* difficult to imagine (Banerjee and Duflo 2020).

Figure 1 illustrates the extent of the change in selected economies in comparison to two Asian Miracle economies. The top two lines show the miracle economies of Taiwan and South Korea. The middle lines show another East Asian economy, Malaysia, which has experienced strong growth compared to much of the world but does not quite fit in the miracle category, and one of the fastest-growing economies in South America, Chile. The bottom line, Ghana, is representative of a number of slower-growth countries in much of the developing world.

■ *Reda Cherif and Fuad Hasanov are both Senior Economists, International Monetary Fund, Washington, DC, and affiliated researchers at the Bennett School of Public Policy at the University of Cambridge, United Kingdom. Hasanov is also an Adjunct Professor of Economics at Georgetown University, Washington, DC. Their email addresses are acherif@imf.org and fhasanov@imf.org.*

Figure 1
Uneven Economic Growth across Countries
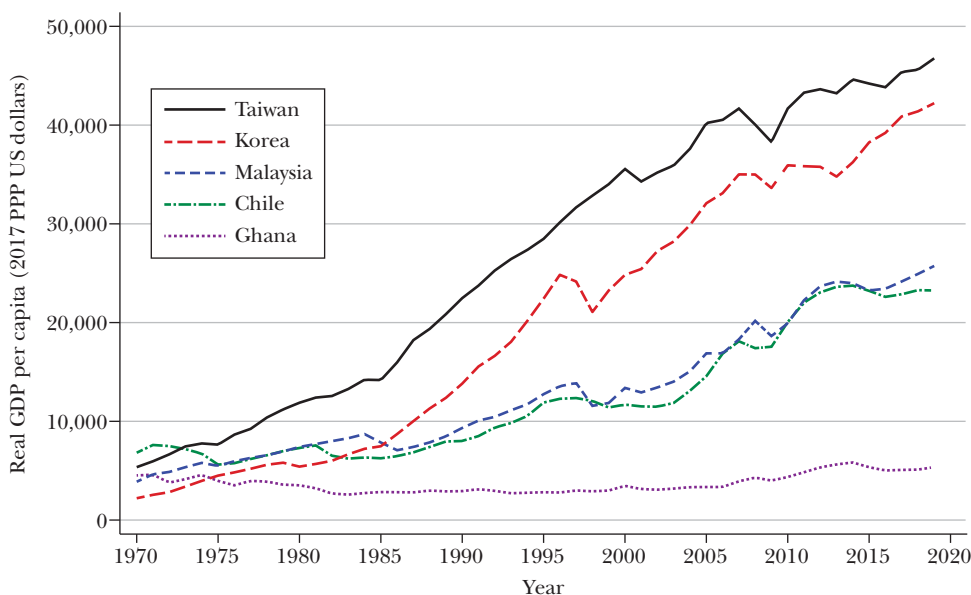
Note: The figure shows three sets of economies: Asian Miracles such as Korea and Taiwan that grew at exceptionally high rates; Malaysia and Chile, which grew at sufficiently high rates to reach upper middle income status; and Ghana, a representative of many other developing countries that barely increased their per capita income. Note also that the Asian Miracles had similar (or lower) levels of per capita income as other countries in 1970, but then their growth experience diverged substantially.

The levels of per capita income in the 1970s indicate the initial conditions of countries and how low the odds were of reaching high-income status in a short period of time. Korea, a low-income country at the time, had about one-half the real income per capita of Ghana, Malaysia, and Taiwan, which had comparable income levels, while it had about one-third of Chile's real income per capita. Translated to 2019 levels, this would be equivalent to the odds that a country much poorer than Ghana at less than one-half of its real income per capita, one-third of India's, and about one-sixth of Colombia's would reach the standards of living of Spain or Italy in about 50 years. In other words, a worker entering the labor market today in one of the poorest countries in the world would retire with the living standards of an Italian or a Spanish worker.

However, most countries representing a large swath of the global population have barely made progress over these decades in closing their huge gap in income per capita; in many cases, they have been falling even further behind the high-income nations (in this journal, Pritchett 1997; see also Johnson and Papageorgiou 2020). In turn, the variation of income per capita across countries explains the

majority of the differences of most of the dimensions that matter for the quality of life, such as poverty, health, education, leisure, and inequality (Pritchett 1995; Sen 1999; Jones and Klenow 2016).

In this paper, we address the following question, paraphrasing a comment from Lucas (1988): Is there some action a government of countries like Ghana or Colombia could take that would lead their economy to grow like the Asian Miracle economies? If so, what, exactly?[1] We argue that the secret of the Asian Miracles lies in the special type of industrial policy they had in common. We suggest three key principles that made the industrial policy of the Asian Miracles stand out compared to the numerous failed attempts at industrial policy in developing economies (Cherif and Hasanov 2019a, 2024): (1) proactive and continuous state action to steer resources into the development of sophisticated industries, with the goal that domestic firms participate fully in technology creation and innovation; (2) exports, exports, and exports; and (3) enforcement of accountability for the support received, both through increased domestic market competition and through goals set based on market signals. We outline a way of putting these principles into action in a large array of contexts, focusing on the institutional set-up and necessary actions.

We take a *pluralistic* perspective in our discussion, drawing from disparate literatures that do not usually speak to each other (in the spirit of Chang 2014). The patterns and commonalities on which we focus draw from a rich non-neoclassical literature in economic history, development studies, and political economy, which studies in detail the policies and contexts of the Asian Miracles. However, we remain fully steeped in the neoclassical tradition inspired by the work of Lucas (1988), as we rationalize the principles of successful industrial policy using neoclassical theories of economic growth and market failures. This framework allows us to synthesize the success into a limited number of replicable principles instead of a vast array of historical and political conditions and factors—say, political consensus built on Cold War geopolitical conditions, the quality of the bureaucracy, culture, or the supportive political system—as is often the case in development studies of the Asian growth miracles.

Our argument emphasizes a role for the state in determining the thrust and direction of growth, by targeting specific sectors and being an integral part of the production process. Thus, our approach is in stark contrast with a more laissez-faire philosophy, in which the state's primary role is to provide a conducive business environment by investing in public goods such as education and infrastructure and ensuring property rights and the rule of law, while regulating only with a light hand.

Growth Traps and Miracles: 50 Years of Economic Development

During the last half-century, the economies of many developing countries either have grown slowly, have stagnated, or have even fallen behind relative to

---

[1] Lucas (1988) wrote: "Is there some action a government of India could take that would lead the Indian economy to grow like Indonesia's or Egypt's? If so, what, exactly?"

Figure 2
50 Years of Economic Development
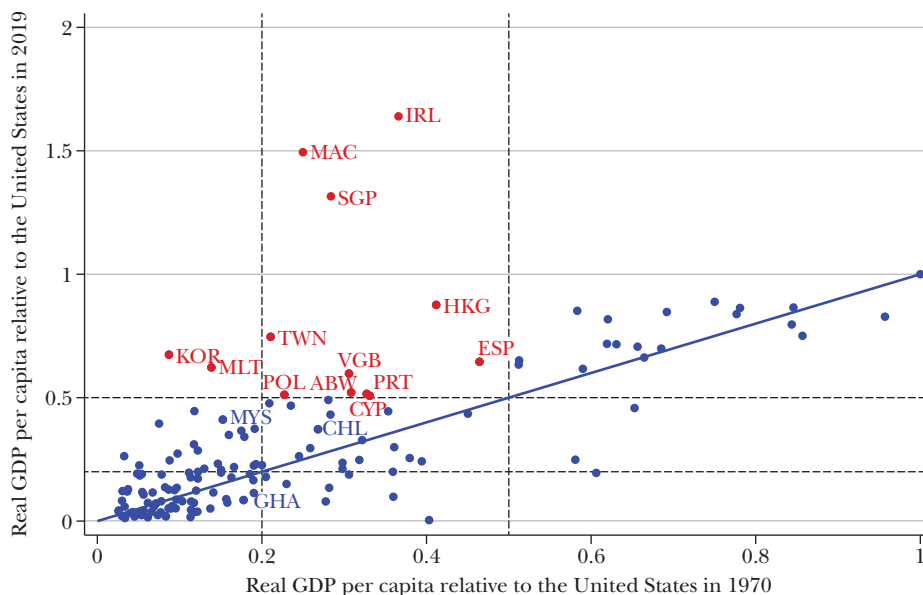
Source: Based on Penn World Tables 10.01 (Feenstra, Inklaar, and Timmer 2015).
Note: The upper-middle income threshold is set at 20 percent of US real GDP per capita (in PPP terms) while that for low-income countries is at 10 percent. The dashed lines at the 20 and 50 percent thresholds approximately correspond to 50th and 75th percentile of the post-2000 income distribution, respectively (Cherif and Hasanov 2019b). These thresholds are more conservative than the absolute thresholds provided by the World Bank (for example, World Development Report 2024), but are more in line with the key percentiles of the global income distribution. We focus on the period from 1970–2019 to avoid including the economic effects of the pandemic (in addition to being constrained by the PWT database), but extending the figure to more recent years should not change the qualitative pattern shown.

the US economy, which is often used as a relevant yardstick for economic development (Cherif and Hasanov 2019b). A good number of countries reached the middle-income status but have not escaped the middle-income region, as illustrated in Figure 2. The horizontal axis shows real per capita GDP relative to the United States in 1970; the vertical axis shows the same measure in 2019. Thus, countries above the 45-degree line are catching up; countries below the 45-degree line are falling further behind, and a large number of countries are quite close to the 45-degree line, remaining roughly the same relative to the US economy, often at levels of 20 percent of the US real per capita GDP (upper-middle income threshold) or less. Figure 2 shows that over 50 years, countries broadly remained within their relative income categories.

Out of the 183 countries in the sample in 2019, only 19 economies forged ahead to reach high-income status from 1970 to 2019 (defined here as above 50 percent of the US real per capita GDP). Our focus here is on the Asian "tiger"

or Asian Miracle economies of Hong Kong, China (HKG), Korea (KOR), Singapore (SGP), and Taiwan, China (TWN). We argue that some of the circumstances of the other countries make them relatively less replicable than the Asian Miracles. Ten others—Poland (POL), Portugal (PRT), Slovenia (SVN), Spain (ESP), Czechia (CZE), Estonia (EST), Ireland (IRL), Lithuania (LTU), Cyprus (CYP), and Malta (MLT)—are periphery European countries, which experienced faster catching up after joining the European Union. Without dismissing the contribution of the institutional reforms conducted in the process, we argue that some of the exceptional benefits of joining the union and their proximity to core European economies make it more difficult to draw practical lessons for the developing world. These countries benefitted from access to large markets, substantial grants for infrastructure, preferential loans, and the integration into regional value chains such as the German automotive industry (Cherif, Hasanov, and Li 2024). The others include the two oil-exporting nations of Bahrain (BHR) and Saudi Arabia (SAU), along with a group of small island nations including Aruba (ABW), Macao, China (MAC), and British Virgin Islands (VGB), which relied on exceptionally large commodity or tourism rents relative to their size. (Incidentally, both Aruba and British Virgin Islands are overseas territories of the Netherlands and the United Kingdom, respectively.)[2]

Given the very low odds of a developing country reaching high-income status over a couple of generations, we argue that the Asian Miracle economies may have important lessons for development for those countries that lack substantial amounts of oil, gas, or other large sources of rents relative to their size, or are not located next to large European high-income countries.

Sustained Productivity Growth: Seeking New Learning Curves

A key fact separating the Asian miracle economies was their sustained productivity growth, driven by process and product innovation, quality upgrading, and moving to sophisticated sectors. Thus, their high sustained growth cannot be traced only to their rise in capital or human capital accumulation (for example, Young 1995; World Bank 1993), which would have faced diminishing returns over time (as in the model of Solow 1956).

Using a standard decomposition of the drivers of economic growth, during the period from 1970 to 2000, the Asian Miracle economies as a group grew by 6.3 percent per year with total factor productivity growth of 1.1 percent, while Korea and Taiwan grew at an annual rate of 7.2 percent with total factor productivity

---

[2]From 1960 to 1970, only Italy (ITA), Israel (ISR), and Japan (JPN) crossed over the high-income threshold. Bahrain (BHR) and Saudi Arabia (SAU) reached the high-income threshold by 1970 but slid back in the 1980s–1990s as oil prices collapsed, crossing back to high income in the 2000s as oil prices recovered. Oman (OMN) also crossed the high-income threshold in the 2000s but fell below it by 2019 (and as a result, it is not counted as a high-income country).

growth of 2.1 percent.[3] Conversely, other high-growing countries such as Malaysia and Chile have not done as well in productivity terms and can trace more than 100 percent of their growth to gains in physical capital, human capital, and the employment/population ratio, averaging negative annual productivity growth over 1970–2019 (Ocampo, Rada, and Taylor 2009).

Starting at about 25 percent of real GDP per capita of the US economy in 1970, on average, the Asian Miracle economies reached about 90 percent of the US level in 2019. On average, it took them about 15 years to cross over 50 percent of the US real GDP per capita, the high-income threshold: from 1970, it took Hong Kong 6 years; Korea, 25 years; Singapore, 9 years; and Taiwan, 18 years.[4]

Korea's path offers an impressive example. Starting as a low-income country at 9 percent of the US per capita income in 1970, Korea took 13 years to reach 20 percent of the US per capita income and only another 12 years to reach our "high-income" threshold of 50 percent of the US income. In contrast, Malaysia, a good-performing country, went from 15 percent to 30 percent of the US per capita income in the same 25 years, muddling through the middle-income trap. Although capital and human capital contribution to growth was high in Korea, total factor productivity growth of about 2 percent a year was instrumental in sustaining a continuous increase in real income, increasing by about 9.1 percent a year.

Lucas (1993) argues that the origin of the Asian Miracles' prodigious productivity growth lies in their ability to continuously enter new tasks, exhausting the productivity gains in those tasks, and quickly move to new ones where the learning curve is steep again. His theory rests on two assumptions: (1) there is learning-by-doing translated into productivity gains, stemming from the cumulative production in each task (Arrow 1962); and (2) for every new task, the productivity gains start at a high level and decline over time. Lucas cites the evidence of US Liberty ships produced during World War II, with learning being steep initially, resulting in high productivity growth and eventually flattening out as cumulative production rises. In this framework, aggregate economic growth would ultimately depend on the economy's ability to learn new tasks, while abandoning others. This is also directly related to the Schumpeterian growth theory of creative destruction,

---

[3]Assuming a Cobb-Douglas production function, real per capita GDP growth can be decomposed into (1) the capital-output ratio, (2) human capital, (3) the employment-population ratio, (4) hours worked, and (5) total factor productivity (TFP). For more details, see Inklaar and Timmer (2013) and Jones (2022).

[4]Distilling lessons for development from country-level experiences is hard given the difficulty of disentangling policies from luck (Banerjee and Duflo 2020; Sachs 2020). Henry and Miller (2009), examining growth experiences of Barbados and Jamaica, argued that policies were more important than legal institutions, geography, culture, and other similar characteristics. Easterly (1995) argued that the Asian Miracles were statistical outliers, which could be the case in a normal distribution. Cherif and Hasanov (2024b) show that the distribution of long-run growth across countries is in fact fat-tailed and could be approximated by a power law distribution. In this case, the East Asian experience holds more information than pure outliers. Using a stylized probabilistic growth model in which both policies and luck play a role in long-run growth, Cherif and Hasanov (2024b) show that there is a myriad of factors that could go wrong, including "bad luck" or negative shocks such as natural disasters, wars, social unrest, or commodity price fluctuations that could even outweigh relatively "good" policies. Only in the long run, as good policies are sustained over time, potentially overshadowing bad luck, an observer could disentangle policies from luck.

in which "quality ladders" are key to growth (Aghion and Howitt 1992). In Lucas (1993), the tasks are aligned on a continuum, which clarifies the formal exposition. However, as economies are instead composed of different industries, each having its own scope of tasks and dynamics, we can use Lucas's (1993) mechanism to infer growth patterns in these economies.

First, industries are unequal in their degree of sophistication, that is, both in terms of the scope of tasks involved and how this scope evolves through time. For example, the tourism industry mostly involves a relatively limited number of tasks, such as cleaning, cooking, and entertaining. Indeed, archeological discoveries show that tourists of ancient Rome, Pompeii, Egypt, or Greece enjoyed broadly similar activities as their twenty-first-century counterparts (Beard 2008; Popkin 2022; Beard 2022). At the other extreme, sophisticated industries such as electronics, aerospace, or automotive involve tens of thousands of tasks, which are perpetually evolving and expanding in quality and scope and entail high and ongoing levels of research and development spending. Indeed, sophisticated industries are commonly defined based on their research and development intensity (Cherif and Hasanov 2019b).

Second, the industrial structure or specialization of economies has often less to do with an inherited comparative advantage than with an acquired one, which arises through the role of the state in tackling a myriad of externalities, coordination failures, and market imperfections that can prevent the emergence of any sophisticated industry. The more sophisticated the industry is, the larger the array of potential market failures. Examples of limitations that need to be overcome include the early-stage barriers and costs to creating industrial clusters and agglomeration economies; the barriers to early production so that it is possible to reap the gains from learning-by-doing; coordination failures (like building an industry-specific infrastructure or training programs for workers); and information asymmetry between the skills and needs of industry and the actions of government.[5] In such cases, a coordinated state intervention is necessary, and it goes beyond providing general education and infrastructure or encouraging research and development in a "horizontal" or across-the-board fashion. There are other types of non-industry-specific externalities and market failures requiring interventions such as antitrust and contract enforcement, patent systems, or a health insurance market. Again, these are not sufficient for the development of the targeted industries (and some may not even be necessary) in the absence of other industry-specific policies. Instead, targeting specific industries and designing a package of industry-specific policies are paramount. There are as many theories justifying state intervention to raise growth as there are externalities and market imperfections (prominent examples include Krugman 1987; Romer 1986; Lucas 1988; Matsuyama 1992; Rodrik 2009; Harrison and Rodríguez-Clare 2010; Cherif 2013; Stiglitz and Greenwald 2014).

---

[5] The aerospace and aircraft industry is an example of a highly sophisticated industry that cannot be developed without a large array of policies, including government procurement, financing and investments, aviation security, airports, and training of pilots and technicians, as shown by the examples of Airbus in Europe, Embraer in Brazil, and the Commercial Aircraft Corporation in China (Fallows 2012).

Third, the extent to which domestic firms are involved in technology creation and innovation is critical to sustaining the mechanism described, especially in escaping the middle-income trap (Cherif and Hasanov 2019c; Lee 2019). The available empirical evidence on the effects of foreign direct investment shows a positive albeit limited effect in terms of productivity gains and spillovers. Multinational corporations can transfer important knowledge, but a host of home-biases and frictions would limit the pace at which new tasks are acquired, domestic firms are created, and homegrown technology is built (Guillouët et al. 2024).

All-in-all, industries or sectors matter for growth, and the development of the sectors that are conducive to achieving high and sustained growth depends on state policies to help create a cluster of competitive domestic firms in these, essentially sophisticated, sectors. But the question remains: how did the Asian Miracles develop these types of industries, and in turn achieve sustained productivity gains, when most other economies failed?

Growth Policies in an Historical Perspective

To make sense of the experience of the Asian Miracles, we consider their experience in the wider context of a quasi-natural experiment that took place on a global scale. In the aftermath of World War II and the wave of decolonization that followed, most developing economies adopted a sort of industrial policy aiming at developing domestic industries, especially heavy industries and manufacturing to replace imports, known as "import-substitution industrialization." Many states intervened heavily in the workings of the economy, subsidizing industries, setting up state-owned enterprises, and more important, protecting domestic markets from international competition. Until the early 1980s, some progress was made in acquiring capabilities in manufacturing and productivity was increasing in many developing economies. Figure 3 shows that this era was characterized by a strong convergence of real manufacturing value-added per capita. Indeed, manufacturing growth from 1965 to 1980, during the import-substitution industrialization era, was much higher than in the subsequent period of 1980–2010, a relatively more laissez-faire period.

The debt crises of the 1980s that hit most of these economies exposed the vulnerabilities of this growth model. During this time, firms were both explicitly and implicitly subsidized, consumers were forced to purchase lower quality goods at higher prices, while governments were accumulating unsustainable levels of external debt to maintain domestic production. The crises led to a blanket rejection of industrial policy in general, summed up by the phrase "you can't pick winners," in favor of what can be described as a standard growth recipe (Krueger 1990; Cherif, Engher, and Hasanov 2024). Subsequently, a large number of these developing economies adopted this standard approach, liberalizing trade and finance and privatizing large swaths of the economy previously under state ownership.

Many economies made great policy strides along all the dimensions of the standard growth recipe—that is, the factors typically considered as explaining growth such as the quality of institutions (Acemoglu and Robinson 2012), education (Barro and Lee 2013) and infrastructure, financial deepening, and openness to trade.

Figure 3
Growth Policy Reversal: Import-Substitution Industrialization versus Laissez-Faire
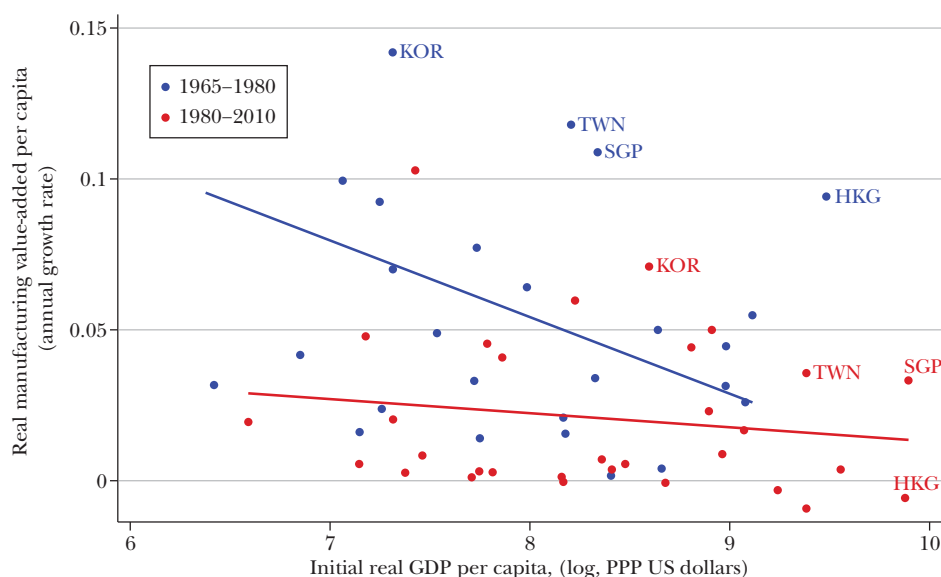
Source: GGDC 10-Sector Database (Timmer, de Vries, and de Vries 2015) and PWT 10.01.
Note: The figure shows real manufacturing value added average annual growth rate for select economies, including Korea, Taiwan, Singapore, and Hong Kong, given their initial level of economic development (real GDP per capita), for two periods: 1965–1980 (import-substitution industrialization) and 1980–2010 (laissez-faire), with respective regression lines. It shows that the growth rates fell from 1965–1980 to 1980–2010.

However, miracles did not result. A comparison of good-performing middle-income countries such as Chile and Malaysia on the one hand, and Korea on the other, along these dimensions is illustrative (Table 1). The three countries had similar levels of real income per capita in 1985. In 2005, the level of income per capita of Malaysia and Chile was reached by Korea back in 1991. Except for schooling, where the difference was relatively small, Chile and Malaysia scored better than Korea along all the dimensions in 1985, while their growth was dwarfed by Korea's growth rate. In 2005, Chile scored better than Korea in 1991 along all the dimensions, and Malaysia scored better along most dimensions. Their growth rates were still one-half of that of Korea at similar levels of real income per capita.

    The diversity of cross-country experiences shows that even good performers in terms of the standard growth recipe did not even come close to achieving the growth performance of the Asian Miracles. In fact, the rate of convergence in manufacturing productivity (measured in per capita value-added) collapsed during the era of standard growth recipe compared to the earlier period of import-substitution industrial policy, while the Asian Miracles remained outliers

Table 1
Key Features of the Standard Growth Recipe

Variables	Chile		Korea		Malaysia	
	1985	*2005*	*1985*	*1991*	*1985*	*2005*
Governance (score)	3	5	2	5	5	2
Law and order (score)	4	5	2	5	5	4
Schooling (years)	7.7	9.7	9.0	9.9	6.7	9.7
Credit (percentage of GDP)	69%	78%	44%	49%	84%	109%
FDI (percentage of GDP)	2%	6%	0%	0%	2%	4%
Trade (percentage of GDP)	52%	70%	62%	52%	106%	201%
Income (US$ in PPP)	6,258	14,571	7,477	15,553	7,856	16,883
Income growth	1%	5%	8%	11%	3%	5%

Source: ICRG scores for governance (corruption) and law and order variables (the lower the measure, the higher the risk); years of schooling from Barro and Lee (2013); private credit, foreign direct investment (FDI), and trade openness from World Development Indicators; and real income per capita and its growth rate (ten-year centered moving average) from PWT 10.01.

in both eras (Figure 3).[6] This evidence does not mean that the ingredients of the standard growth recipe do not matter. For example, the Asian Miracles had high levels of investment in human and physical capital. However, it indicates that the Asian Miracles were doing something beyond the standard growth recipe, which other developing economies did not pursue.

The Three Principles of Technology and Innovation Industrial Policy

The overall strategy that sustained high growth in the Asian Miracles, allowing them to reach high-income status in a couple of generations, was neither import-substitution industrialization nor *sole* reliance on the standard growth recipe. Many economists have argued that a standard growth policy of providing an enabling business environment, like business-friendly regulations, supporting trade liberalization, and investing in education and infrastructure, paved the way for sustained high growth across the developing world, including in East Asia (for example, World Bank 1993; Young 1995; Easterly and Levine 2002; Acemoglu and Robinson 2012). In general, the standard growth formula seeks to address "government failures" such as improving regulations (that is, building "good" institutions) and providing public goods like education and infrastructure.

But although the standard growth recipe may support growth, it is not sufficient for high and sustained growth. "Market failures" such as investment coordination, informational asymmetries, knowledge and learning externalities and agglomeration, and missing finance are not much resolved with standard growth policies.

---

[6] Hong Kong's relatively lower growth coincides with the outsourcing of its manufacturing to the Pearl River Delta.

The constraints generated by a myriad of market failures lead to underinvestment or no investment in innovation, startups and scale-ups, firm growth, sophisticated industries, and homegrown technologies, and essentially result in missing dynamic tradable sectors that are key for sustained growth (Cherif, Hasanov, and Zhu 2016). Seeking to address these constraints is the focus of the industrial policy geared toward technology creation and innovation. For more discussion of government and market failures, a useful starting point is Rodrik (2005).

Economists have emphasized this broader definition of market failures in growth theory, but they have largely been ignored in the policy discussions. Notable exceptions include Murphy, Shleifer, and Vishny (1989) on coordination and "big push" investment and Lucas (1993) on new tasks and learning-by-doing, who provide mechanisms behind the success of the Asian Miracle economies, but without much discussion of policies to implement these mechanisms to generate high, sustained growth. Aghion and Howitt (1992) emphasize creative destruction and moving up the quality ladder in production, suggesting some elements of industrial policy of the Asian Miracles like promoting competition. Recent empirical evidence on assessing the impact of industrial policies in East Asia supports the importance of industrial policies for industry and firm growth (Kim, Lee, and Shin 2021; Choi and Levchenko 2025; Lane 2025).

Moreover, the literature on East Asia's growth experience suggests that there are strong policy commonalities among these high-growth countries (for examples, Amsden 1989, 2003; Wade 1990; Woo 1991; Fallows 1995; Lall 1996; Chang 2002; Chang and Andreoni 2016; Cherif and Hasanov 2019b; Adler 2020). The common thread runs not only through the geographic location of the Asian Miracles but through a set of policies similar across East Asia, Japan before them, and even Germany, the United States, and the United Kingdom in the nineteenth and early twentieth centuries (Chang 2002).

Drawing from a rich literature on the industrial policy of the Asian Miracle economies, we believe that the core elements can be condensed to three principles (Cherif and Hasanov 2019a, 2024a): (1) proactive and continuous state action to steer resources into the development of sophisticated industries, with the goal that domestic firms participate fully in technology creation and innovation; (2) exports, exports, and exports; and (3) enforcement of accountability for the support received, both through increased domestic market competition and through goals set based on market signals. In a nutshell, it was this form of technology and innovation policy, industrial policy Asian Miracle style, sustained over decades, that was the key to developing sophisticated sectors with a focus on domestic firms creating various technologies and products. This policy mix differentiated the Asian Miracles from developing countries that grew at relatively high rates but did not make it to high income. Here, we discuss these three ingredients in turn.

Principle 1: Steering Resources to Domestic Participation in Sophisticated Industries

The Asian Miracle economies all entered highly sophisticated sectors early on, which were far beyond their pre-existing "comparative advantage." For example, in

the 1970s, they explicitly targeted sectors with high research and development levels such as electronics, automotive, and machinery when they had no prior experience. In contrast, if one took seriously their revealed comparative advantage in the 1960s and 1970s—that is, the export products for which these countries had relatively large global market shares—the optimal sectors for the Asian Miracles would have consisted in producing rice, vegetables, or sewing machines. Such a strategy would have been obviously self-limiting (Cherif and Hasanov 2019b; Lin and Chang 2009; Lall 2005).

Their path to more sophisticated technologies often involved intermediate steps. For example, the entry into more sophisticated industries often coincided with a reliance on low-skilled industries like textiles. These industries were a "quick win" in the early stages of the development to generate growth and employment as well as finance their investment in sophisticated industries. Yet not graduating to a higher skilled industry or moving up the quality ladder eventually exhausts the productivity gains and innovation potential (Lucas 1993; Aghion and Howitt 1992; Acemoglu, Aghion, and Zilibotti 2006; Hausmann, Hwang, and Rodrik 2007; Cherif, Hasanov, and Aghion 2023). For instance, Bangladesh has not managed to move away from the textile industry. Although Korea relied on textiles for a sizable share of its exports until the early 1990s, its share dwindled afterward.

Essentially, the role of the state is first to identify certain sophisticated industries as the potential engines of growth for long-term development objectives. This step does not preclude investing in and exporting products in low-tech industries during the transition, to generate jobs and hard currency. Once the decision is taken, however, how policymakers invest in these sectors matters. The extent to which there is a continuous effort to encourage domestic firms to contribute to technology creation and innovation explains the divergence of Korea and Taiwan from Malaysia in electronics, for example.

Contrary to a common belief, the larger Asian Miracle economies, Japan, Korea, and Taiwan, did not much rely on foreign direct investment to develop their sophisticated sectors, although they did commonly adapt and experiment with foreign technologies. They typically developed partnerships with multinational corporations, often licensing critical technologies in the early stages of development but quickly developing their own. In fact, Japan and to some extent Korea had regulations restricting foreign direct investment, rather than encouraging it, for a long time (Johnson 1982).

Malaysia's endeavor in electronics at around the same time as Korea and Taiwan in the 1970s, and their divergence afterward, is illustrative of the limits of relying on foreign direct investment. Despite efforts to increase local content in production of electronics and offering a particularly conducive business environment, Malaysia could not replicate the success of Korea's Samsung. In the 1980s and into the 1990s, Korea and Taiwan began to emphasize domestic research and development spending and to increase their level of patents. In contrast, research and development spending was much lower in Malaysia than in Korea at the same levels of development. For instance, research and development in Korea in the mid-1990s was already above 2 percent of GDP, while in Malaysia it was less than 0.5 percent of GDP (Cherif and

Figure 4
Innovating Early
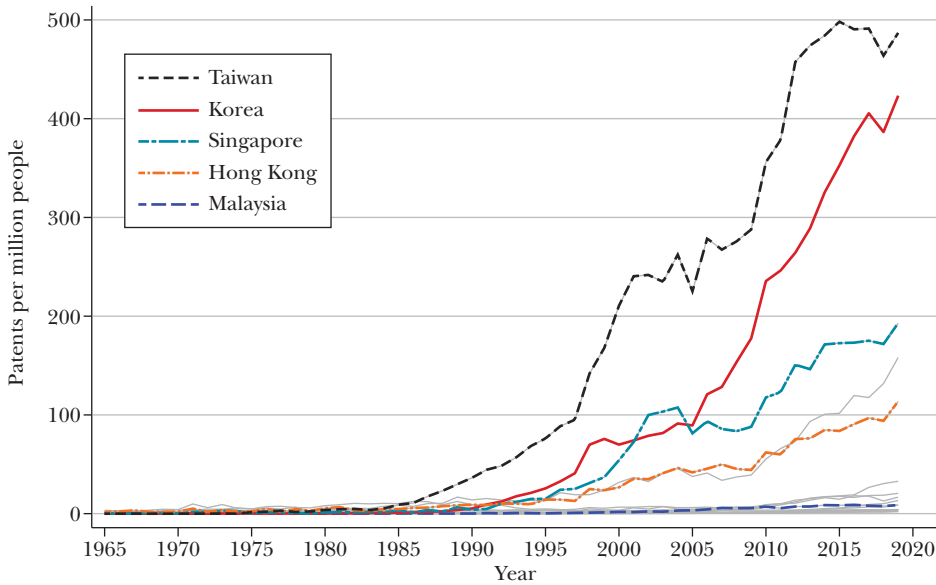
Source: Based on US PTO.
Note: The figure shows patents granted in the United States per million people for Asian Miracles (Korea, Taiwan, Singapore, and Hong Kong) and developing countries (Malaysia, highlighted). Note the substantial divergence in patents with twelve other developing countries shown with the gray lines.

Hasanov 2019b). Multinationals were much less interested in producing innovation in Malaysia while domestic firms in Korea were investing and innovating to compete globally. Even if foreign direct investment could help to jumpstart a certain industry, pivoting toward homegrown technology creation early on when the level of economic development was still low is part of what differentiated Asian Miracles from other relatively fast-growing countries that were not able to sustain their growth. Figure 4 illustrates how patents per million people granted in the US skyrocketed in Korea and Taiwan compared to the rest of the developing world.

Principle 2: Export Orientation

A focus on exports and global markets, or the discipline provided to domestic producers by the need to meet export-related goals (Studwell 2013), substantially increases the odds of an industrial policy leading to sustained growth.

Historically, the industrial policy used by most developing economies in the 1960s and 1970s involved import-substitution industrialization, in which the focus was on replacing imported products with domestic production. The majority of these policies were deemed a failure and halted two to three decades later. This experience contributed to the toxic reputation of industrial policy and to the

maximalist argument that developing economies should not attempt industrial policy (for example, Krueger 1990).

These early efforts at import-substitution industrial policy did conform with the first principle, as economies from Latin America and Africa to Asia poured resources in heavy industries and manufacturing and started acquiring advanced capabilities. As noted earlier, this investment was reflected in productivity gains in manufacturing and an increase in manufacturing output early in this period (Cherif and Hasanov 2019b, 2024a).

However, the import-substitution industrial policies ended up shielding domestic producers from international competition through tariffs and subsidies, while also relying on imported critical inputs and machinery and selling to a captive domestic market. At least in theory, one can justify this approach based on an "infant industry" argument in which learning needs to take place before the industry matures and is ready to compete with others (for example, Krugman 1987; Melitz 2005). But in practice, import-substitution protectionism creates complacency and reduces innovation (Aghion et al. 2015), thus making the industry uncompetitive and unsustainable in the long run. The protected domestic firms have little incentive to innovate, to invest in marketing and design, to integrate vertically and increase scale, or to develop a network of efficient suppliers. More importantly, the developmental state in its support of industries is flying blind in the absence of reliable market signals that exports and global competition provide. Even countries with a large domestic market like India could not or did not create enough competition and innovation in its strongly protected auto industry (Cherif and Hasanov 2024a).

In contrast, the Asian Miracle economies stood as exceptions. Although a few of the initial steps had some flavor of import substitution, they pushed relatively early for their domestic firms to export and continuously adjusted their policies in response to export performance (Johnson 1982; Wade 1990; Chang 1994; Woo-Cumings 1999; Cherif and Hasanov 2024a, 2019b). Figure 5 shows a continuous increase in real (deflated by the US GDP deflator) manufacturing exports per capita compared to other developing countries. Malaysia's manufacturing exports had been growing in tandem with Korea's until the early 2000s, but then Korea had overtaken it, too.

The stark difference in the approaches adopted by Korea and Malaysia in the automotive industry in the 1970s and 1980s illustrates the importance of export-orientation (Cherif and Hasanov 2019c). Malaysia relied on a national champion, Proton, that was barely pushed to export. Because Korea's Hyundai was pushed to export, it invested in large factories far beyond the needs of the Korean market to leverage economies of scale. It also set up its own network of dealerships in the US market, actively helped develop a network of domestic suppliers to cut its costs, and invested heavily in research and development to develop its own engine. By continuously pursuing innovation and investing in the company, Hyundai has become an internationally competitive firm and a major exporter. In contrast, Proton remained dependent on Japanese engines, offering lower quality cars at higher prices compared to foreign competition, and could only survive in the

Figure 5
Strong Export Performance in Asian Miracles
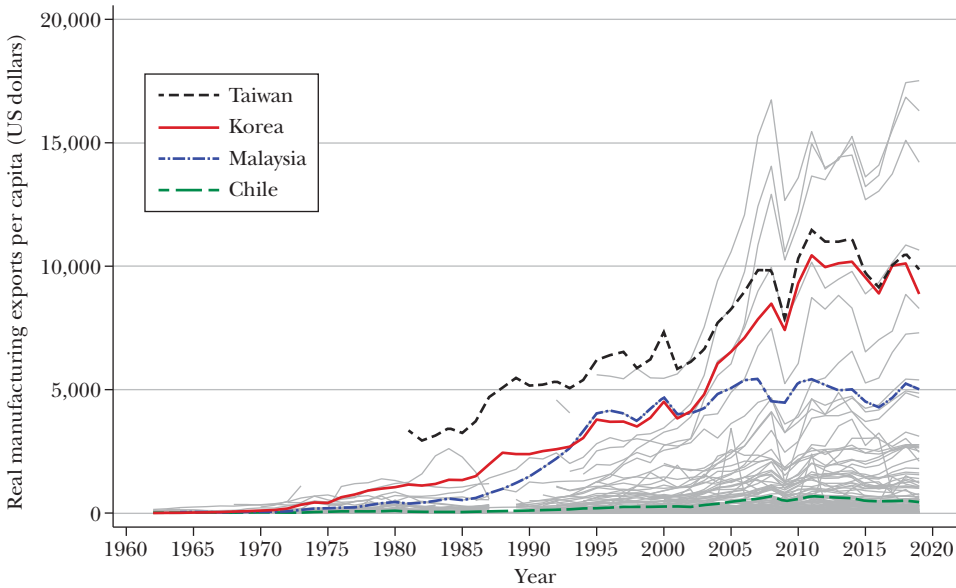
Source: World Development Indicators.
Note: The figure shows real manufacturing exports per capita (deflated by the US GDP deflator) for Korea, Taiwan, Malaysia, Chile, and 120 other developing countries (light gray lines). It indicates that manufacturing exports grew substantially in Asian Miracles. Smaller European economies such as Czechia, Slovakia, and Slovenia had higher manufacturing exports per capita in the 2000s than Korea or Taiwan, as shown in the figure.

Malaysian market because it was protected. After decades of direct and indirect subsidies, it was eventually sold to a Chinese multinational firm.

Principle 3: Accountability for Support Provided

The principle of strict accountability for support provided through technology and innovation industrial policy addresses both efficiency and political economy concerns in conducting industrial policy. Although export orientation provided competition on a global level, domestic competition is also important as state support to industry, rather than to a firm, has a potential to create a competitive industry even when not all firms survive (Aghion 2016). The state support is provided with strings attached, or accountability, to align incentives for growth and innovation. Market competition is an integral part of accountability, and the role of industrial policy is not merely to ensure competition among domestic firms (or with foreign ones), but it is also *responsible* for the emergence of a competitive industry as a whole. As human and financial resources are mobilized, both accountability and market competition act as self-correcting and mutually reinforcing mechanisms.

This principle, along with export-orientation, was another major missing ingredient in most failed industrial policies attempted outside of the Asian Miracles.

The standard approach adopted by the failed industrial policies in many countries emphasizes accountability at the project level and project execution. Thus, it tends to focus on a mix of misleading targets such as quantities produced and local content, cost-benefit analyses often conducted by interested parties, and governance issues, while ignoring dynamic patterns, innovation, spillovers, and opportunity costs. Large infrastructure projects or low-sophistication industries like tourism may be executed well and fulfill these kinds of accountability standards, but may not do much to raise the sophistication of a country's industrial structure. These bets involve sizable public investments but have large opportunity costs, including trade-offs between "quick wins" and long-run transformative changes. Policies also often focus on a single national champion and do not assess the competitiveness of the whole industry, including its value chain structure, its market power in international markets, and its exposure to exchange rate movements (or else many countries would not have pursued import-substitution industrial policy).

In contrast, the strategy of the Asian Miracle economies would often have failed a traditional test of accountability. Corruption and nepotism were endemic, and many projects and investments failed or would have been considered "white elephants" at the onset (Chang, Park, and Yoo 2001; Ang 2020). However, their accountability philosophy focused on assisting the emergence of internationally competitive industries at the frontier of technology and innovation. This meant a continuous benchmarking against international standards and favoring the accumulation of skills and capabilities. The focus was on indicators such as global market shares, exports, and patents, instead of shorter-term signals such as aggregate GDP growth, employment, or profits.

The Asian Miracles encouraged fierce competition among domestic firms and rarely depended on a single champion. Using competition as an evolutionary process, policymakers gave the failing firms a chance to compete, but eventually such firms would be restructured, and a leading firm would often emerge through mergers and acquisition. Studwell (2013) describes this process as "culling," and Cherif and Hasanov (2024) describe how it occurred in the case of Hyundai in Korea where domestic competition among the chaebols in the auto industry was fierce. There is a large literature describing how the institutions in charge of industrial policy were imposing competition and using market signals to enforce accountability (for example, Johnson 1982; Wade 1990; Chang 1994; Woo-Cumings 1999; Cherif and Hasanov 2024a). The growth of US major industrial firms in the early twentieth century followed a similar pattern of culling, including in the auto industry (Chandler 1990).

Granted, there are several instances when the state intervened to manage competition in the early stages of development in the Asian Miracles by limiting the number of entrants, imposing mergers, or forcing firms to follow a cartel behavior (Johnson 1982; Chang 2002). However, these measures were unusual, and they came only *after* domestic competition turned out to be too fierce, potentially threatening

the survival of the industry. Essentially, the state acted as a venture capitalist, which did not know *a priori* which firms would succeed and adopted a portfolio approach to selecting firms that, instead of "picking winners," in the words of Sanjaya Lall (1996), "created winners."

The Role of a Lead Agency in the Conduct of Industrial Policy

The success of the Asian miracles was built on continuous implementation of technology and innovation policy driven by a handful of key government agencies, often described as a "developmental state" (Johnson 1982; Wade 1990; Evans 1995; Woo-Cumings 1999; Amsden 1995; Haggard 1990; Chang 1994; Lin 2012). Broadly speaking, discussions emphasize the role of a powerful and competent bureaucracy actively intervening to spur industrial development through a range of tools such as subsidies, protectionism, and credit. However, certain aspects of this approach are generally considered unique to the Asian Miracles, such as their embedded autonomy (Evans 1995) and their technical capabilities (Haggard 1990), which in turn are the result of a replicable institutional design and the concrete steps they took (Cherif, Hasanov, and Xie 2025). In other words, the Asian Miracles, which started with low-capacity administration as relatively low-income economies, managed to build the institutional structure and capabilities that were adapted to their industrial policy objectives (for example, Chang 1994). There are important lessons to be learned in terms of what constitutes a competent bureaucracy and how to acquire it.

At the heart of the developmental state lies a leading agency tasked with implementing industrial policy—that is, targeting sophisticated sectors, coordinating and executing sector-specific policies, and engaging with the private sector and other government agencies while changing policies when the context and priorities change (for more detail, see Cherif, Hasanov, and Xie 2025). These leading agencies, like Japan's Ministry of International Trade and Industry, Korea's Economic Planning Bureau, or Singapore's Economic Development Board, differed in various ways, but they all implemented the three principles of technology and innovation industrial policy.

The first task for this leading agency is to accumulate a context- and sector-specific knowledge about industries, markets, technologies, and firms, which can be accumulated through learning-by-doing by continuous experimentation and feedback from all the stakeholders involved. This accumulated knowledge is akin to the concept of *metis* described by Scott (1998).

In addition to knowledge accumulation, coordination is needed. The set of policies to be implemented can involve many ministries, agencies, and quasi-governmental organizations, covering a wide array of distinct policy areas such as trade, education, infrastructure, finance, technology, and regulation. This agency must be endowed with a mandate to implement technology and innovation policy and sufficient power to coordinate various stakeholders. To facilitate the

accumulation of knowledge, the East Asian agencies were structured into separate industrial departments, while intersectoral (or horizontal) departments ensured coordination and accountability. Many developing economies have a one-stop shop to attract foreign direct investment or to encourage private investment, but these institutions are not designed to develop the necessary expertise, set priorities, and address the full range of bottlenecks and constraints. This form of one-stop shop can be modestly useful, but would not suffice to achieve high and sustained growth.

Although the leading agencies are often described as "planning agencies" in the literature, it would perhaps be more accurate in practice to call them anti-planning agencies. That is, they did not specify a narrow set of tools to achieve specific targets. They did not strive for what has been called "legibility," seeking a kind of explicit clarity—which was often behind major policy failures in the past (Scott 1998).

Thanks to their embeddedness with the private sector, while remaining autonomous at the same time (Evans 1995), they were able to accumulate a deep knowledge of industries, design the full spectrum of policies, and continuously adapt them.

These agencies had an array of prerogatives they held under their control, including trade policy, purpose-specific investment in infrastructure or skills (Chang 1994), technology diffusion (Wade 1990), and channeling financing to the targeted industries, which was labelled "administrative guidance" in Japan or "rediscounting" in Korea (Studwell 2013). The relations with the private sector were important as knowledge flowed in both directions and policies and firm actions were coordinated and adapted to the changing circumstances (Cherif, Hasanov, and Xie 2021).

For a leading agency to conduct technology and innovation policy effectively, the state needs to apply a "4A model" of the organizational design for the agency to be effective (Cherif, Hasanov, and Xie 2025): (1) *ambition* in the goals and the agency to achieve them, (2) *autonomy* from undue influence in the conduct of technology and innovation policy, (3) *accountability* along several dimensions (vertical vis-à-vis the political leadership, horizontal vis-à-vis firms and other agencies, and internal), and (4) *adaptability* to the changing political and economic context. These characteristics align with features of the leading agencies of the Asian Miracle economies as well as with our reinterpretation of the rich literature describing the "developmental mindset" or the "embedded autonomy" of the institutions that produced economic miracles (see also Wade 1990; Chang 1994; Ohno 2012; Wade 2012). Indeed, modern central banks mandated with managing short-term economic fluctuations have largely followed the same 4A model, shielding the bank from political interference and providing policy continuity, crucial to the success of policies that bear fruit over the long horizon.

Here, we add a few thoughts about each aspect of the 4A model for a leading agency. Its ambition can be described in terms of the degree of sophistication of the sectors the agency targets and the involvement of domestic firms, especially in terms of technology creation. The Asian Miracle economies showed an ambition that was described as "reckless" at the onset of their industrial policies (Woo 1991). They started from a low level of economic development and had no prior experience in any of the sophisticated industries they entered such as shipbuilding, automotive,

and electronics. While they pursued "quick wins" in the short run to maintain growth and provide employment in industries like textiles, their technology and innovation policies constituted a "moonshot" approach to development, investing in it very early, nurturing and growing industries, and generating growth and employment for the next stages of development (Cherif and Hasanov 2019b).

The ambition of the leading agency needs political support. National goals are articulated by the politicians, usually in the form of a long-term "vision," and it must include ambitious aims—that is, development of internationally competitive export-oriented and innovative domestic industries and firms. In the Asian Miracles, it was achieved initially through an alignment on national goals and by having the leading agency's senior staff nurture close personal relations and trust with the executive, which was critical to make these ambitious bets and coordinate policies across various agencies (Cherif, Hasanov, and Xie 2025; Dercon 2022).

Ambition also needs to be reflected in the quality of the staff. The leading agencies for the Asian Miracle economies relied, since their inception, on a small-scale competitive recruitment from the most prestigious institutions, often based on exams. For example, Japan's Ministry of International Trade and Industry and Korea's Economic Planning Bureau were mostly staffed by the best graduates from the University of Tokyo's Economics and Law Departments and Seoul National University, respectively (Johnson 1982; Choi 1987). A competent leading agency can be built even in a low capability environment by hiring the best graduates and/or providing training. Chang (1994) gives the example of Korean civil servants receiving training in the 1970s from more advanced economies at that time such as Egypt and Pakistan.

Autonomy was critical in the success of development projects throughout the world, including East Asia (Evans 1995). The leading agency needs protection from undue interventions, ranging from nepotism to rent-seeking behavior or the protection of special interests. The legal framework adopted by the Asian Miracle economies to endow their leading agencies with autonomy follows the same blueprint followed later in the 1990s and 2000s by most economies to make their central banks independent (Eijffinger and De Haan 1996)—that is, personnel independence, where the agency was given the ability to decide its own recruitment, wage, and promotion policy outside of the generally restrictive general civil service system; financial independence, where the agency had its own budget, which was relatively easy to implement given that these agencies were lean; and policy independence, such that the agency could design policies and enforce its decisions.

To ensure that the leading agency fulfills its mandate and is not captured by politicians or private executives, accountability is paramount (as discussed in Rodrik 2008). Accountability comes in three forms: (1) internal accountability refers to systems within the agency to ensure transparency, efficiency, and adherence to goals; (2) vertical accountability refers to the agency's responsibility to the executive branch in fulfilling its mandate; and (3) horizontal accountability refers to the agency's ability to hold firms accountable for results when they receive support as well as to the engagement with other agencies (Cherif, Hasanov, and Xie 2025).

Being flexible and autonomous enough, while also being accountable for the actions taken, can keep these agencies focused on its key mandate to implement technology and innovation policy.

Adaptability is needed because as the economic and political context changes in terms of technologies, market structure, and macroeconomic and sectoral dynamics (like relative wages), the industry-specific requirements and priorities also eventually change. The leading agencies of the Asian Miracle economies had similar arrangements that led to being proactive in monitoring and understanding these trends and changes, adapt to them, and pivot if needed. This approach to policy formulation steeped in a broad and flexible mandate; monitoring channels to evaluate compliance and adapt their goals accordingly; and an evolving institutional structure reflecting changes in industries targeted and the level of maturity of the industries (Hasanov, Cherif, and Xie 2025, 2021).

Lessons from the Industrial Policy Asian Miracle Style

There is a remarkable difference between the initial conditions of Japan, Korea, Taiwan, Singapore, and Hong Kong (among other differences, in size and political system), which influenced their industrial policies. Japan's Ministry of International Trade and Industry developed a hidden technocracy in the shadow of its democracy, which was tightly knit with the industrial conglomerates, or *zaibatsus*, while Korea's Economic Planning Bureau used its proximity with an authoritarian president to support and discipline the Korean conglomerates, or *chaebols*. Taiwan focused on semiconductors, instead of finished goods, and devised a spin-off system to create startups such as the United Microelectronics Corporation and the Taiwan Semiconductor Manufacturing Company that grew rapidly to become global leaders. A city-state like Singapore relied on state-owned enterprises and leveraged partnerships with multinational corporations, while Hong Kong under British rule deployed an effective industrial policy through a network of councils relying on indirect tools such as support for research and development.

But behind these differences lurk discernible similarities. Ambition, or a developmental mindset, was a common driving force for change. The leading agency for the developmental state also had autonomy, accountability, and adaptability. The leading agencies mandated to implement technology and innovation industrial policy geared toward developing export-oriented sophisticated industries with homegrown technology and innovation and focused on the long run, emphasizing policy continuity and reducing political and economic uncertainty with regard to regulations, tax regimes, availability of infrastructure, and so on.

We interpreted the experience of the Asian Miracles through the lens of neoclassical economics with market imperfections, while challenging the traditional concept of industrial policy. For example, the standard theory of comparative advantage largely ignores how that comparative advantage is developed and evolves. In contrast, considering externalities such as learning-by-doing could provide a

rationale for the type of industrial policies conducted by the Asian Miracles and their success at creating internationally competitive sophisticated industries. In modern research and development–intensive technology industries, a combination or symbiosis of the state and the market raises the odds for success (Cohen and DeLong 2016; Mazzucato 2015).

An old-style industrial policy that only imposes tariffs and disperses subsidies is unlikely to achieve a growth miracle. Indeed, a technology and innovation industrial policy cannot be designed *a priori*. It goes far beyond the idea that there are a few market failures that can be precisely identified and tackled, or that it simply consists in providing a mix of subsidies and trade protection. We emphasize the existence of many coordination failures, learning externalities, and market imperfections.

To formulate the policies to tackle market failures, the leading agencies accumulated practical, specialized knowledge of industries, largely acquired through continuous experimentation and feedback. As they experimented, the Asian Miracles supported industries with a host of policies beyond subsidies or protectionism. Some of these tools such as tariffs may not have been necessary, and they may not be adapted to today's context anymore. Empirical studies on past successes, or failures, at industrial policies are very important to show what worked and what did not. However, research in this area typically focuses on only a narrow set of measurable tools, mostly subsidies, which is not very informative for the technology and innovation industrial policy we are discussing.

The type of institutions and the strategy needed to walk in the footsteps of the Asian Miracle economies do not necessarily fit the standard understanding of "good" institutions in economics (Chang 2011). The definition in the economics literature emphasizes factors like rule of law, respect for private property, absence of corruption, and the general provision of public goods such as education and infrastructure. Although the benefits of "good" institutions in this sense are meaningful, these institutions do little to tackle many market failures constraining the growth of sophisticated industries—and moving up a ladder of sophisticated industries is the heart of a growth miracle. The experience of the Asian Miracles suggests that institutions may not have to be that "good" in the conventional sense if a country adopts a technology-based and export-oriented approach, at least up to a level of development closer to the frontier. In other words, the standards needed in terms of good governance and rule of law to achieve a miracle may not be as high as previously thought.

Although high levels of investment in physical and human capital are an integral part of the Asian Miracle story, the direction given to these investments differs from the traditional growth narrative. Instead of trying to improve all growth factors at once and hoping that a private production would spontaneously emerge, the institutions of the Asian Miracles took a more localized and purpose-specific approach. They focused at first on a leading agency that targeted selected industries and geographic areas. They could thus mobilize their scarce resources and build momentum. It is easier to build competence in a relatively small agency, provide the right infrastructure or skills for a specific industry in a specific area, and tackle

specific types of corruption in a certain context (Ang 2020) than to fix all these challenges everywhere at the same time. A purpose-specific rather than a general approach saves on resources and opportunity costs (Chang 1994).

Trying to emulate the Asian Miracles will necessarily attract skepticism. We argue that we should still strive to understand what and how they did it, for "if we know what an economic miracle is, we ought to *make* one" (Lucas 1993).

References

Acemoglu, Daron, Philippe Aghion, and Fabrizio Zilibotti. 2006. "Distance to Frontier, Selection, and Economic Growth." *Journal of the European Economic Association* 4 (1): 37–74.

Acemoglu, Daron, and James Robinson. 2012. *Why Nations Fail: The Origins of Power, Prosperity and Poverty.* Crown Business.

Adler, David. 2020. "Korean Industrial Policy: From the Arrest of the Millionaires to Hallyu." *American Affairs* 4 (1).

Aghion, Philippe. 2016. "Growth Policy Design for Middle-Income Countries." In *Breaking the Oil Spell: The Gulf Falcons' Path to Diversification*, edited by Reda Cherif, Fuad Hasanov, and Min Zhu, 117–30. International Monetary Fund.

Aghion, Philippe, Jing Cai, Mathias Dewatripont, Luosha Du, Ann Harrison, and Patrick Legros. 2015. "Industrial Policy and Competition." *American Economic Journal: Macroeconomics* 7 (4): 1–32.

Aghion, Philippe, and Peter Howitt. 1992. "A Model of Growth through Creative Destruction." *Econometrica* 60 (2): 323–51.

Amsden, Alice H. 1989. *Asia's Next Giant: South Korea and Late Industrialization.* Oxford University Press.

Amsden, Alice. 1995. "Like the Rest: South-East Asia's 'Late' Industrialization." *Journal of International Development* 7 (5): 791–9.

Amsden, Alice. 2003. *The Rise of "The Rest": Challenges to the West from Late-Industrializing Economies.* Oxford University Press.

Ang, Yuen Yuen. 2020. *China's Gilded Age: The Paradox of Economic Boom and Vast Corruption.* Cambridge University Press.

Arrow, Kenneth J. 1962. "The Economic Implications of Learning by Doing." *Review of Economic Studies* 29 (3): 155–73.

Banerjee, Abhijit V., and Esther Duflo. 2020. "How Poverty Ends: The Many Paths to Progress—and Why They Might Not Continue." *Foreign Affairs* 99 (1): 22–29.

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

Beard, Mary. 2008. *Pompeii: The Life of a Roman Town*. Profile Books.

Beard, Mary. 2022. "I Came, I Saw, I Bought the Souvenir: Evidence of Tourism in the Ancient World." *TLS*, May 20. https://www.the-tls.co.uk/classics/roman-souvenirs-kimberly-cassibry-maggie-popkin-book-review-mary-beard.

Chandler, Alfred D., Jr. 1990. *Scale and Scope: The Dynamics of Industrial Capitalism*. Harvard University Press.

Chang, Ha-Joon. 1994. *The Political Economy of Industrial Policy*. Macmillan.

Chang, Ha-Joon. 2002. *Kicking Away the Ladder: Development Strategy in Historical Perspective*. Anthem Press.

Chang, Ha-Joon. 2011. "Institutions and Economic Development: Theory, Policy and History." *Journal of Institutional Economics* 7 (4): 473–98.

Chang, Ha-Joon. 2014. *Economics: The User's Guide*. Bloomsbury Press.

Chang, Ha-Joon, and Antonio Andreoni. 2016. "Industrial Policy in a Changing World: Basic Principles, Neglected Issues and New Challenges." Paper presented at 40 Years of the *Cambridge Journal of Economics* Conference, Cambridge, July 13.

Chang, Ha-Joon, Hong-Jae Park, and Chul Gyue Yoo. 2001. "Interpreting the Korean Crisis: Financial Liberalization, Industrial Policy and Corporate Governance." In *Financial Liberalization and the Asian Crisis*, edited by Ha-Joon Chang, Gabriel Palma, and D. Hugh Whittaker, 140–55. Palgrave Macmillan.

Cherif, Reda. 2013. "The Dutch Disease and the Technological Gap." *Journal of Development Economics* 101: 248–55.

Cherif, Reda, Marc Engher, and Fuad Hasanov. 2024. "Crouching Beliefs, Hidden Biases: The Rise and Fall of Growth Narratives." *World Development* 173: 106246.

Cherif, Reda, and Fuad Hasanov. 2019a. "Return of the Policy that Shall Not Be Named: Principles of Industrial Policy." IMF Working Paper 19/74.

Cherif, Reda, and Fuad Hasanov. 2019b. "Principles of True Industrial Policy." *Journal of Globalization and Development* 10 (1): 20190034.

Cherif, Reda, and Fuad Hasanov. 2019c. "The Leap of the Tiger: Escaping the Middle-Income Trap to the Technological Frontier." *Global Policy* 10 (4): 497–511.

Cherif, Reda, and Fuad Hasanov. 2024a. "The Pitfalls of Protectionism: Import Substitution vs. Export-Oriented Industrial Policy." *Journal of Industry, Competition and Trade* 24 (1): 14.

Cherif, Reda, and Fuad Hasanov. 2024b. "When Tolstoy Meets Leontief: Luck, Policies, and Learning from Miracles." *Structural Change and Economic Dynamics* 68: 86–97.

Cherif, Reda, Fuad Hasanov, and Philippe Aghion. 2023. "Fair and Inclusive Markets: Why Dynamism Matters." *Global Policy* 14 (5): 686–701.

Cherif, Reda, Fuad Hasanov, and Xun Li. 2024. "A 'True' Industrial Policy for Europe is a Technology and Innovation Policy." In *Investing in the Structural Transformation: 2024 European Public Investment Outlook*, edited by Floriana Cerniglia and Francesco Saraceno. Open Report Series, vol. 13. https://doi.org/10.11647/OBP.0434.

Cherif, Reda, Fuad Hasanov, and Lichen Wang. 2018. "Sharp Instrument: A Stab at Identifying the Causes of Economic Growth." IMF Working Paper 18/117.

Cherif, Reda, Fuad Hasanov, and Gary Xie. 2021. "The Making of East Asia's Electronics Champions." *Revista de Economía Mundial* 59.

Cherif, Reda, Fuad Hasanov, and Gary Xie. 2025. "The Foundations of Economic Development: Institutions for Industrial Policy." Unpublished.

Cherif, Reda, Fuad Hasanov, and Min Zhu, eds. 2016. *Breaking the Oil Spell: The Gulf Falcons' Path to Diversification*. International Monetary Fund.

Choi, Byung-Sun. 1987. "The Structure of the Economic Policy-Making Institutions in Korea and the Strategic Role of the Economic Planning Board (EPB)." *Korean Journal of Policy Studies* 2: 1–25.

Choi, Jaedo, and Andrei A. Levchenko. 2025. "The Long-Term Effects of Industrial Policy." *Journal of Monetary Economics* 152: 103779.

Cohen, Stephen S., and J. Bradford DeLong. 2016. *Concrete Economics: The Hamilton Approach to Economic Growth and Policy*. Harvard Business Review Press.

Dercon, Stefan. 2022. *Gambling on Development: Why Some Countries Win and Others Lose*. Hurst.

Easterly, William. 1995. "Explaining Miracles: Growth Regressions Meet the Gang of Four." In *Growth Theories in Light of the East Asian Experience*, NBER-EASE Vol. 4, edited by Takatoshi Ito and Anne O. Krueger, 267–99. University of Chicago Press.

Easterly, William, and Ross Levine. 2002. "It's Not Factor Accumulation: Stylized Facts and Growth

Models." In *Economic Growth: Sources, Trends, and Cycles*, Vol. 6, edited by Norman Loayza and Raimundo Soto, 61–114. Central Bank of Chile.

Eijffinger, Sylvester C. W., and Jakob De Haan. 1996. "The Political Economy of Central-Bank Independence." Princeton University Special Papers in International Economics 19.

Evans, Peter B. 1989. "Predatory, Developmental, and Other Apparatuses: A Comparatvie Political Economy Perspective on the Third World State." *Sociological Forum* 4 (4): 561–87.

Evans, Peter B. 1995. *Embedded Autonomy: States and Industrial Transformation.* Princeton University Press.

Fallows, James. 1995. *Looking at the Sun: The Rise of the New East Asian Economic and Political System.* Pantheon Books.

Fallows, James. 2012. *China Airborne: The Test of China's Future.* Vintage 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.

Guillouët, Louise, Amit K. Khandelwal, Rocco Macchiavello, Madhav Malhotra, and Matthieu Teachout. 2024. "Language Barriers in Multinationals and Knowledge Transfers." *Review of Economics and Statistics.* https://doi.org/10.1162/rest_a_01487.

Haggard, Stephan. 1990. *Pathways from the Periphery: The Politics of Growth in the Newly Industrializing Countries.* Cornell University Press.

Hale, Galina, and Cheryl Long. 2011. "Are There Productivity Spillovers from Foreign Direct Investment in China?" *Pacific Economic Review* 16 (2): 135–53.

Harrison, Ann, and Andrés Rodríguez-Clare. 2010. "Trade, Foreign Investment, and Industrial Policy for Developing Countries." In *Handbook of Development Economics*, Vol. 5, edited by Dani Rodrik and Mark Rosenzweig, 4039–214. North-Holland.

Hausmann, Ricardo, Jason Hwang, and Dani Rodrik. 2007. "What You Export Matters." *Journal of Economic Growth* 12: 1–25.

Henry, Peter Blair, and Conrad Miller. 2009. "Institutions versus Policies: A Tale of Two Islands." *American Economic Review* 99 (2): 261–67.

Inklaar, Robert, and Marcel P. Timmer. 2013. "Capital, Labor, and TFP in PWT8.0." Unpublished.

Johnson, Chalmers. 1982. *MITI and the Japanese Miracle: The Growth of Industrial Policy, 1925–1975.* Stanford University Press.

Johnson, Paul, and Chris Papageorgiou. 2020. "What Remains of Cross-Country Convergence?" *Journal of Economic Literature* 58 (1): 129–75.

Jones, Charles I. 2022. "The Past and Future of Economic Growth: A Semi-endogenous Perspective." *Annual Review of Economics* 14: 125–52.

Jones, Charles I., and Peter J. Klenow. 2016. "Beyond GDP? Welfare across Countries and Time." *American Economic Review* 106 (9): 2426–57.

Kim, Minho, Munseob Lee, and Yongseok Shin. 2021. "The Plant-Level View of an Industrial Policy: The Korean Heavy Industry Drive of 1973." NBER Working Paper 29252.

Krueger, Anne O. 1990. "Government Failures in Development." *Journal of Economic Perspectives* 4 (3): 9–23.

Krugman, Paul. 1987. "The Narrow Moving Band, the Dutch Disease, and the Competitive Consequences of Mrs. Thatcher: Notes on Trade in the Presence of Dynamic Scale Economies." *Journal of Development Economics* 27 (1-2): 41–55.

Lall, Sanjaya. 1996. *Learning from the Asian Tigers: Studies in Technology and Industrial Policy.* Palgrave.

Lall, Sanjaya. 2005. "Reinventing Industrial Strategy: The Role of Government Policy in Building Industrial Competitiveness." In *The IMF and the World Bank at Sixty*, edited by Ariel Buira, 197–233. Anthem Press.

Lane, Nathane. 2025. "Manufacturing Revolutions: Industrial Policy and Industrialization in South Korea." *Quarterly Journal of Economics* 140 (3): 1683–741.

Lee, Keun. 2019. *The Art of Economic Catch-Up: Barriers, Detours and Leapfrogging in Innovation Systems.* Cambridge University Press.

Lin, Justin Yifu. 2012. *The Quest for Prosperity: How Developing Economies Can Take Off.* Princeton University Press.

Lin, Justin, and Ha-Joon Chang. 2009. "Should Industrial Policy in Developing Countries Conform to Comparative Advantage or Defy it? A Debate between Justin Lin and Ha-Joon Chang." *Development Policy Review* 27 (5): 483–502.

Lucas, Robert E., Jr. 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics* 22 (1): 3–42.

Lucas, Robert E., Jr. 1993. "Making a Miracle." *Econometrica* 61 (2): 251–72.

Matsuyama, Kiminori. 1992. "Agricultural Productivity, Comparative Advantage, and Economic Growth." *Journal of Economic Theory* 58 (2): 317–34.

Mazzucato, Mariana. 2015. *The Entrepreneurial State: Debunking Public vs. Private Sector Myths.* Revised ed. PublicAffairs.

Melitz, Marc J. 2005. "When and How Should Infant Industries Be Protected?" *Journal of International Economics* 66 (1): 177–96.

Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny. 1989. "Industrialization and the Big Push." *Journal of Political Economy* 97 (5): 1003–26.

Ocampo, José Antonio, Codrina Rada, and Lance Taylor. 2009. *Growth and Policy in Developing Countries: A Structuralist Approach.* Columbia University Press.

Ohno, Kenichi. 2012. *Learning to Industrialize: From Given Growth to Policy-Aided Value Creation.* Routledge.

Popkin, Maggie L. 2022. *Souvenirs and the Experience of Empire in Ancient Rome.* Cambridge University Press.

Pritchett, Lant. 2024. "The Case for Economic Growth as the Path to Better Human Wellbeing." https://static1.squarespace.com/static/62e11adc3b05e91cc75bbc10/t/670992bc797e3f7021 cb2d88/1728680642371/Economic+Growth+first+complete+draft.pdf.

Pritchett, Lant. 1997. "Divergence, Big Time." *Journal of Economic Perspectives* 11 (3): 3–17.

Ricardo, David. 1817. *On the Principles of Political Economy and Taxation.* John Murray.

Rodrik, Dani. 2005. "Growth Strategies." In *Handbook of Economic Growth*, Vol. 1A, edited by Philippe Aghion and Steven N. Durlauf, 967–1014. Elsevier.

Rodrik, Dani. 2008. "Normalizing Industrial Policy." World Bank Commission on Growth and Development Working Paper 3.

Rodrik, Dani. 2009. "Industrial Policy: Don't Ask Why, Ask How." *Middle East Development Journal* 1 (1): 1–29.

Romer, Paul M. 1986. "Increasing Returns and Long-Run Growth." *Journal of Political Economy* 94 (5): 1002–37.

Sachs, Jeffrey D. 2020. "Trials and Tribulations: A Response to 'How Poverty Ends.'" *Foreign Affairs* 99 (3): 186–89.

Sen, Amartya. 1999. *Development as Freedom.* Oxford University Press.

Scott, James. 1998. *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed.* Yale University Press.

Solow, Robert. 1956. "A Contribution to the Theory of Economic Growth." *Quarterly Journal of Economics* 70 (1): 65–94.

Stiglitz, Joseph E., and Bruce Greenwald. 2014. *Creating a Learning Society: A New Approach to Growth, Development, and Social Progress.* Columbia University Press.

Studwell, Joe. 2013. *How Asia Works: Success and Failure in the World's Most Dynamic Region.* Grove Press.

Timmer, Marcel, Gaaitzen J. de Vries, and Klaas de Vries. 2015. "Patterns of Structural Change in Developing Countries." In *Routledge Handbook of Industry and Development*, edited by John Weiss and Michael Tribe, 65–83. Routledge.

Wade, Robert. 1990. *Governing the Market: Economic Theory and the Role of Government in East Asian Industrialization.* Princeton University Press.

Wade, Robert H. 2012. "Return of Industrial Policy?" *International Review of Applied Economics* 26 (2): 223–39.

Wade, Robert H. 2018. "The Developmental State: Dead or Alive?" *Development and Change* 49 (2): 518–46.

Williamson, John, ed. 1990. *Latin American Adjustment: How Much Has Happened?* Institute for International Economics.

Woo, Jung-en. 1991. *Race to the Swift: State and Finance in Korean Industrialization.* Columbia University Press.

Woo-Cumings, Meredith. 1999. "Introduction: Chalmers Johnson and the Politics of Nationalism and Development." In *The Developmental State*, edited by Meredith Woo-Cumings, 1–30. Cornell University Press.

World Bank. 1993. *The East Asian Miracle: Economic Growth and Public Policy.* Oxford University Press.

Young, Alwyn. 1995. "The Tyranny of Numbers: Confronting the Statistical Realities of the East Asian Growth Experience." *Quarterly Journal of Economics* 110 (3): 641–80.

The World Bank's *East Asian Miracle*: Too Much a Product of Its Time?

Nancy Birdsall

A renewed interest in industrial policy—in which governments promote certain industries seen as critical to long-run growth of the economy—has arisen in the last decade. In high-income countries, the focus is on green energy, strategically important products such as chips, and a resurgence of investment in manufacturing in general. For developing economies, there is the question of whether "industrial policy" might stimulate catch-up growth, and whether it might help deal with "premature deindustrialization" (Rodrik 2016)—that is, with the global decline of the manufacturing jobs that led the growth of today's high-income countries.

It is an apt moment to look back to the 1993 World Bank publication, *The East Asian Miracle* (Page et al. 1993). The "miracle" of the title referred to the extraordinarily rapid growth in the 1960s through the 1980s of Japan and the four "tiger" economies (Korea, Taiwan, Hong Kong, and Singapore) of northeast Asia, and the three economies of southeast Asia (Indonesia, Malaysia, and Thailand) that had experienced economic growth of a lesser but still-impressive degree. From 1960 to 1985, this set of countries had growth "more than twice as fast as the rest of East Asia, three times as fast as Latin America and South Asia, and five times as fast as sub-Saharan Africa (pp. 2–3). In fact, along with China, the East Asian economies represent the greatest triumph of global development in the past century—a triumph built in part on "industrial policy." Yet the World Bank, the largest lender and a major adviser to developing countries, has not encouraged other countries

■ *Nancy Birdsall is Senior Fellow and President Emerita, Center for Global Development, Washington, DC. Her email address is nbirdsall@cgdev.org.*

to implement industrial policy in the three decades since. The *Miracle* report can be read, after the fact, as an effort to explain and reconcile the resulting tension.

What was the recipe for this remarkable growth in countries of East Asia, and what role was played by industrial policy relative to other factors? The *Miracle* book was the first full exploration by the World Bank of what could be viewed as "industrial policy" in a developing country setting. The report was clear that these East Asian countries grew via success in industry or manufacturing and in some cases were characterized by considerable "government efforts to alter industrial structure to promote productivity-based growth" (p. 304). However, in no way was the book an endorsement of industrial policy by the World Bank for other developing regions and countries—indeed, quite the contrary. The book argued that East Asia's success was grounded in the "fundamentals": macroeconomic stability, high rates of investment in human as well as physical capital, free trade, policies that ensured growth was shared or inclusive, and government institutions and practices that minimized corruption. Even in the book's discussion of industrial policy, the emphasis was on the strong institutional framework that supported the interventions and the unusual capacity of governments to maintain and sustain that underlying support.

The *Miracle* book got largely positive attention in the business and financial press (in outlets like the *Economist*, the *Financial Times*, and the *Wall Street Journal*), where coverage focused on the Bank's emphasis on the market-friendly "fundamentals" versus state-led industrial policy as central to the region's economic success. In the World Bank itself, the book was well-received; the use of "miracle" in the title gave it a certain elan. For all practical purposes, industrial policy did not subsequently make it onto the Bank agenda, nor onto the agenda of other development sponsors and supporters.

But history tells a long tale, and in this third decade of the twenty-first century, more than 30 years later, industrial policy is back on the agenda of development and other economists, including some at the World Bank—this time with a growing number of high-profile academic papers, using significantly improved empirical methods, finding evidence of positive effects of industrial policy in advanced economies (Juhász, Lane, and Rodrik 2024).

I spent 14 years from 1980 to 1993 at the World Bank; I was the head of the Policy Research Department, reporting to then Bank Chief Economist Lawrence Summers, when the Government of Japan offered to finance a study by the Bank of its own and its neighbors' notable postwar economic success. In that role, I appointed John Page, a senior World Bank economist, to head the team of World Bank staff and consultants that prepared the report, and he was its main author. I am listed as one of the book's eight authors because I did the background work on education, income distribution, and labor market issues; but with little experience or expertise on industrial policy and related macroeconomic, finance, and trade issues, and no experience in East Asia, I was a student more than a contributor on the issue of industrial policy in those countries. When Summers left the World Bank in late 1992 to join the incoming Clinton administration, I became the Acting Chief Economist with respect to the (virtually completed) East Asia report. In that role I

was the designated defender of the book-to-be for its in-house prepublication review and got some insight into the resistance inside the Bank to any seeming endorsement of industrial policy for other developing countries.

In this essay, I begin with a reminder of the origins of the report in the early 1990s, both within the economic and political context of that era and within the World Bank itself. I then review the main themes of the *Miracle* story explaining the success of the East Asian economies—high investment rates in human and physical capital, relatively open trade policies, inclusive growth, and a form of industrial policy that focused on exports (the latter seen as the key to these economies' unusually high rates of productivity growth as a result of "technological learning"). I consider central criticisms made of the report both at the time and since. In the conclusion, I consider the extent to which lessons of the early 1990s may be relevant, or not, for developing countries today.

A Product of Its Time

The *Miracle* publication was very much a product of its time with its emphasis on market fundamentals—sound macroeconomic management, agricultural policies "that did not tax the rural economy excessively" (p. 5), financial policies that made banking accessible to nontraditional savers, and largely free and open trade policies—as critical to successful growth. It was prepared in 1992 and published in 1993. This was just a few years after the fall of the Berlin Wall in 1989; by the end of 1991, the Union of Soviet Socialist Republics had dissolved itself as a sovereign state. In trade policy, President George Bush signed the North American Free Trade Agreement (NAFTA) in 1992 on behalf of the United States with representatives of Canada and Mexico; the treaty was then approved by Congress and signed into law by President Clinton in 1993. The World Trade Organization came into existence at the beginning of 1995 as the successor to the post–World War II General Agreement on Tariffs and Trade (GATT). In the realm of development policy, the so-called "Washington Consensus" was first published in 1989, in which John Williamson had set out ten policies to promote growth in developing countries—policies that were widely (and incorrectly) viewed as reflecting an exaggerated "Washington-based international organizations" and US government view that developing countries should adhere strictly to free markets and open trade. (For a more sophisticated view of what the Consensus said and how it was applied, a useful starting point is Spence (2021) and the other three papers in the "Symposium on the Washington Consensus" in the Spring 2021 issue of this journal.)

Taking these events together, they seemed to many at the time to vindicate both democracy as a political system and free, open markets as an economic system. The early 1990s are reasonably remembered as a time when the view that market-driven economies and free trade (now often called "neoliberalism") were key to successful growth in developing countries was so widely shared it was not debated. In the World Bank at the time, free markets were taken for granted.

It was at about this time that the Government of Japan offered to finance the marginal cost (for example, staff travel and consultants' time) of a study by the World Bank of the postwar economic success stories of East Asia, with a focus on Korea, Taiwan, Hong Kong, Singapore, Malaysia, Thailand, and Indonesia, as well as Japan itself. Their success stories were built at least partly on policies and programs that were enacted in addition to and sometimes in violation of the singular pro-market orientation that had come to dominate the World Bank and the larger development community by the early 1990s.[1] We can surmise that the Japanese wanted World Bank economists to unpack the keys to Japan's own dramatic post–World War II economic growth—including the role of its powerful Ministry of Trade and Industry (MITI) in supporting Japan's impressive industrial growth since the end of that war—as well as the keys to the extraordinary growth success of Japan's neighbors in East Asia in the three decades beginning in the 1960s. What lessons from East Asia might guide the World Bank (and the other multilateral banks and the IMF) in their future advice and support to developing countries in other less successful regions?

The Japanese gift was savvy (as well as generous) in its recognition that the way to influence the views of World Bank economists about of the "miracle" of fast growth in East Asia in the prior three decades was to sponsor a study, thus ensuring that at least some of the authors would internalize the potential role of industrial policy as a contributor to long-run growth—and in that way expand their perspectives beyond the basic recipe of market-driven growth based on accumulation of human and physical capital and productivity gains driven by healthy market competition. The resulting *Miracle* book took a big step in the direction the Japanese presumably envisioned, with its emphasis on the "export push" across all the high-performing East Asian economies, including its acknowledgement of the "modest" (always the adjective used) import protection on behalf of exporters and the mild financial repression that financed governments' credit support to exporting firms. (The next section will spell out the *Miracle* recipe for growth in some detail.)

Overall, did the (possible) objective of the Japanese—to put some form of industrial policy as practiced in Japan and its near neighbors on the World Bank map of good options for other developing countries in other regions—succeed?

Yes and no. The requirements and the benefits of state support for growth of manufacturing and especially of manufactured exports were laid out clearly in the *Miracle* report. But the report is also replete with cautionary points for other developing countries on the steep institutional demands of successfully implementing such an "export push" (and managing other programs that diverged from the neoclassical market model): "The institutional demands are not compatible with other developing countries where the fundamentals are not in place" (p. 26); "the use of contests in Japan and Korea required competent and insulated civil servants" (p. 26); "the prerequisites for success were so rigorous that policymakers seeking

---

[1] Wade (1990, p. xviii) says Japan was "angry" at the "Bank's criticisms of its aid program in support of capitalist transformation . . . " and wanted more research on the region as a whole.

to follow similar paths in other developing economies have often met with failure" (p. 6); "the ability of economic technocrats to formulate and implement policies with a minimum of lobbying for special favors from politicians and interest groups . . . is critical" (p. 167).

I wonder whether the Japanese sponsors of the report were not in the end disappointed by its tone and what Wade (1996) refers to as its "paradigm maintenance." Ultimately, the 1993 *Miracle* book did not endorse "industrial policy" for other developing country regions, at least not in the immediate future. Indeed, the report did not even endorse the approach of Japan's own vehicle of industrial policy. Japan's powerful Ministry of Trade and Industry and the Korean Heavy and Chemical Industry Drive were deemed to be largely ineffective (p. 293 and Appendix 6.1).

Looking back, it is hard to evaluate whether these cautions about an industrial policy approach in the Miracle countries influenced subsequent hesitation of staff inside the World Bank and in the larger development community to encourage developing country governments to be activists in some form, or whether the cautions simply reflected real and already understood constraints to successful industrial policy in most developing countries (then if not now). Either way, industrial policy did not make its way onto the World Bank policy map for developing countries in other regions in the more than three decades that followed, with one short-lived exception noted below.

In the nigh 30 years following the *Miracle* report, industrial policy (at the least in the form of the "export push") did not have legs inside the Bank—not among the small team of research economists in the Bank, and not among World Bank country economists working in other regions. Nor was it much explored in development aid programs of the high-income countries, nor by developing countries themselves, with one exception noted below. In the practice of economic development, a singular market-oriented approach reigned for another three decades.

The Miracle Book Recipe: Neoliberalism Plus Exports Plus Institutions of Market Coordination

The *Miracle* volume set out three key ingredients for the rapid growth of the high-performing Asian economies: market fundamentals, shared growth, and the export-push industrial policy.

Market Fundamentals

The Miracle economies hewed closely to the "market fundamentals" for growth. As one would expect of any report in the early 1990s, following the "lost decade" of development in the 1980s due to hyperinflation and debt crises in many developing countries, good macroeconomic and financial management were repeatedly underlined as the *sine qua non* of healthy growth.

Good management encouraged high rates of accumulation of physical capital, with investment and savings rates (the latter supported by "postal" savings systems

operated through the national postal service in several of the economies) almost twice those of other regions (in the report, Figure 1.5). It also included unusually rapid increases in human capital (from what was already a head start in the 1960s compared to other regions; pp. 15, 43), aided by, in a virtuous circle, declining fertility associated with increases in girls' education.

What distinguished the Miracle economies from most of the developing world (and at the time from the former republics of the Soviet Union), however, was the ever-increasing efficiency in the use of these inputs—that is, high total factor productivity growth—including by adoption of foreign knowledge and technology from high-income countries, or "catching up." The *Miracle* authors attributed the very high levels of total factor productivity growth in these East Asian economies to the "technological learning" associated with exports; they saw the continuous process of technological learning as key to understanding the growth miracle.

Equitable and Shared Growth

The *Miracle* report also put considerable emphasis on steps the East Asian countries took to ensure equitable or shared growth. For example, there was land reform in Korea and Taiwan (in the case of Korea, initiated by the US military in the immediate post-Korean War period), in which large estates were broken up and distributed to farmers. Education spending was concentrated at the primary and secondary levels—thus resisting the political pressure in most developing countries to favor university education for the elite in the use of public resources for education—with a resulting rapid elimination of initial gender gaps in education. These countries generally avoided the high indirect taxes on agriculture typical in other regions (especially Latin America), which were commonly associated with food price controls and chronically overvalued currencies that benefited urban consumers at the cost of rural producers. Such policies were part of what Lipton (1977) had called the "urban bias" in policies of developing countries. Financial policies like the postal savings accounts were instituted that made banking accessible to nontraditional savers.

In the early 1990s, this argument that lower levels of inequality could improve growth prospects in developing countries was new to the World Bank. It would not be until 2013 that the World Bank announced it would pursue what it labeled "shared prosperity," making it an explicit goal for all member countries, and began to publish books under that name and with that measure (for example, World Bank 2014; World Bank 2016). Since then, a book on shared prosperity has been published annually.

What led the high-performing Asian economies policies to carry out policies and programs of shared growth? The *Miracle* report refers to the need for these countries to establish their political viability and legitimacy in the face of recent war and political threats: the aftermath of the Korean War, the conflict as Taiwan separated politically from mainland China, and for some, local Communist movements. Similarly, the political leadership of Japan needed to earn public confidence after the losses and destruction the population had endured during World War II.

The *Miracle* report includes country-by-country approaches to "achieving legitimacy through shared growth" (in addition to the heavy investments in education and health that yielded the notable declines in poverty and inequality): from Malaysia's National Economic Plan to increase the wealth of the Bumiputra (which refers to Malays and indigenous people) relative to the ethnic Chinese; to postwar land reform in Japan, Taiwan, and Korea; to large, public investments in housing targeted to low-income households in Hong Kong and Singapore; and to support for small and medium-sized enterprises. Mechanisms that involve permanent asset transfers (land reform is an obvious example), by promising a future stream of income to beneficiaries, make the latter more likely to support growth-promoting policies and inhibit them from disrupting the economy—and do so over the long run.

Several years after the *Miracle* report, Campos and Root (1996) authored *The Key to the Asian Miracle: Making Shared Growth Credible.* In an indirect response to the *Miracle* report, they emphasize shared growth as a legitimating principle central and fundamental to the success of the Miracle economies—shared growth in the form of land reform; support to small and medium-sized industries, including in rural areas; and large public investments in rural infrastructure (in Indonesia especially). As they point out (p. 45): "Asset transfers (skills and property) represent a permanent stream of income and give beneficiaries a stake in the economy's long-term growth."[2]

Export-Push as Industrial Policy

The *Miracle* volume highlights a version of industrial policy in the implementation of an "export push" strategy for manufactured goods—"the broadest interventions were generous incentives for manufactured exports" (p. 88). The export push can be seen as a strategy to speed the structural transformation of economic development that Lewis (1954) had set out—essentially the transfer of labor from agriculture to higher-productivity industry. The Kuznets (1955) curve, showing that economic development was associated with an increase and then a decline in inequality, was documented based on this pattern of structural transformation and increasing productivity in economic development for the United Kingdom, the United States, and Prussia some decades earlier.

However, the focus was not just on a shift to manufacturing, but specifically on manufactured goods for export to the international market, justified in the report as a matter of economic theory by referring to the information-related externalities benefiting all domestic firms in "imperfect world technology markets" (p. 23) and by the coordination problems inherent in building an industrial economy. The report does not actually treat export-push as an "industrial policy"; instead, the report defines industrial policy as "government efforts to alter industrial structure to promote productivity-based growth": "We find, on the one hand, that government

---

[2] They cite an earlier working version of a paper later published as Alesina and Perotti (1996), an econometric study across 71 countries showing that improvements in the share of income of the third- and fourth-income quintiles (meant to roughly represent the middle class) reduce political instability.

efforts to promote specific industries generally did not increase economywide productivity. On the other hand, the evidence shows that broad government support for exports was a highly effective way of enhancing absorption of international best-practice technology, thus boosting productivity and output growth" (p. 293).

The export push took the form of various nonmarket policies in support of exports of manufactured products, especially in the northern tier of East Asia (Japan itself, Korea, Taiwan, Singapore, and Hong Kong). All the high-performing Asian economies, for example, practiced some form of mild import protection, while ensuring automatic access for their exporters to intermediate imported inputs (always at international prices). Korea and Taiwan provided exporters with credit subsidies (performance-based, tied to firms' export orders), financed by mild financial repression. Also in Korea and in Thailand, export targets "provided the basis for awarding access to foreign exchange, investment licenses or credits" (p. 89). In Hong Kong, public housing programs helped keep labor costs of industry down. All the high-performing Asian economies maintained flexible labor markets. Also, to safeguard the cost-competitiveness of exports they resisted creation of industry-wide unions and minimum wage demands, especially during any pressure or shock to macroeconomic stability.

The *Miracle* report argues that the support for industrial exporters enabled "technological learning" on the part of the private firms; that is, they could move rapidly up the scale of technological sophistication as they learned from equipment suppliers and importing firms abroad, from buyers who would transfer tacit knowledge, and from occasional access to nonproprietary technology—and eventually via the return of nationals from study and experience abroad. The authors attributed the very high levels of total factor productivity growth in the Miracle economies to the "technological learning" associated with exports, a benefit also emphasized by others (for example, Studwell 2013; Campos and Root 1996); they saw the continuous process of technological learning as the key to understanding the high rates of productivity and thus of growth in the countries of East Asia.

A perpetual question for advocates of industrial policy is not whether such a policy can possibly work, but whether the political process can oversee such a policy without giving in to rent-seeking special interests. The success of industrial policy in East Asia is described in the *Miracle* report as associated with two critical institutional traits: insulation of a well-compensated economic and bureaucratic technocracy from political pressure and the use of "deliberation councils."

On insulation of economic bureaucrats, for example, the report noted that in Thailand, the Budget Bureau "has tight control of the budget drafting process" (p. 171), and in Indonesia, "parliamentary rules restrict the legislature's discussion of the budget to "broad policy issues" (p. 173). Moreover, these insulated bureaucrats were honest—the successful representatives of a merit-based and highly competitive system—and were "amply rewarded." Their resulting prestige provoked the only explicit reference to culture in the *Miracle* book: "Prestige is of course partly a function of culture . . . Confucian thought, with its veneration of scholars and preoccupation with written tests, especially civil service exams, remains a powerful force

in Hong Kong, Japan, Korea, Singapore and Taiwan, China. Not surprisingly, these societies have produced strong bureaucracies" (p. 178).

Deliberation councils were made up of business and labor members chosen by their respective sectors to allow for interaction of the technical bureaucrats with private business players; they were meant to address the high coordination costs that all countries face between and among business actors from different sectors, and between the private sector and the government, as well as to reduce rent-seeking by either private firms or powerful bureaucrats. The report's authors took care to put the councils in the context of addressing a market failure: the councils "supplemented the market's information transmission function."

Such councils were especially well-established in Japan and Korea, but still "evolving" in Malaysia at the time. In the report's discussion of these councils, there is a rare reference to them as a specific experience other developing countries might follow: "Malaysia's experience with public-private cooperation through deliberation councils has particular relevance for developing countries. Unlike Japan or Korea, Malaysia is a multi-ethnic society; and unlike Singapore it has a relatively large land area and population" (p. 186). Several years after the report, deliberation councils were described and discussed in much greater detail in Campos and Root (1996).

The critical role of deliberation councils in East Asia has not subsequently received much attention in the World Bank, nor at the other multilateral banks. One possibility is that they have been seen as unlikely to work in other regions in the absence of such key preconditions as secure macroeconomic stability and a strong cadre of honest and well-compensated technocrats in government. Nor has creation of a core of technocratic bureaucrats received much attention as a critical "policy" in itself.

Summarizing the Miracle Recipe

How one interprets the core message of the *Miracle* report becomes a matter of emphasis among the three main ingredients. The report emphasized such market fundamentals as sound macroeconomic management and free and open trade policies as critical to successful growth. It also emphasized an agenda of shared prosperity, including sustained investment in improving educational attainment, agricultural policies that "did not tax the rural economy excessively" (p. 5), and others. At the same time, it was clear that market fundamentals did not tell the whole story in East Asia. Indeed, some of the steps the governments of the Miracle economies took to push exports are still discouraged in other developing countries, such as (modest) protection of domestic industry substitutes, subsidized credit to selected industries, interest rates paid on private savings kept low, ceilings on borrowing rates to increase firm profits—and most central, using all of the above to support exporting industries.

In short, the report constituted a quiet endorsement of a limited form of "industrial policy" that encouraged exports but was seen by the *Miracle* authors as neutral to the type of industry. For example, it did not endorse the specific efforts in Japan and Korea—and later in Indonesia—to develop "heavy" industries, assessing these as ineffective in the long run. It was, in short, a careful and qualified endorsement,

for some developing countries under some conditions, of some aspects of "industrial policy," but not of industrial policy itself. The report did not usher in consideration of advice or support for industrial policy in and for other developing countries by World Bank (or International Monetary Fund) economists in the subsequent three decades, even on a limited basis.

Reaction Inside the World Bank

The general reception of the *Miracle* report was positive inside the World Bank, similar to the broader public reception described earlier. However, the reception was less positive in the East Asia region of the World Bank.

My recollection is that the Vice President of the East Asia region objected to use of the word "strategic" in the *Miracle* team's initially proposed subtitle of the book, presumably because it could be taken to suggest a heavy government role as the key to the region's success; the final subtitle became "Economic Growth and Public Policy." In particular, staff in the East Asia region were worried about anything in the text of the book that might undermine the Bank's difficult, then-ongoing position with the government of Indonesia: the Bank was pushing back hard on that government's first steps with an industrial policy in support of developing a commercial airline industry—on the grounds it could never be competitive—and correctly so as it eventually turned out (as reported by McBeth 2023).

At least some former World Bank staff now say this divergence in reactions was no surprise. After all, the operational departments of the World Bank directly interacting with governments will naturally emphasize different perspectives than that of the Policy and Research Department where the book was prepared. For example, when the text of a World Bank report makes repeated use of caveats warning of the challenges of carrying out any version of industrial policy in countries where macroeconomic stability is an ongoing challenge and/or governance is weak, the operational departments face the unenviable task of telling governments that these caveats apply directly to them.

Here's an epilogue: Fifteen years later, in 2008, Justin Lin, a citizen of China born in Taiwan, became Chief Economist of the World Bank. He developed a good relationship with the government of Ethiopia, advising the government on an approach he called New Structural Economics (Lin 2011). The approach, which emphasized governments building on the initial comparative advantage of their economies, was described by a then-former World Bank economist as "a relatively activist—and controversial variant of industrial policy" (Gelb, Clark, and Meyer 2012) and generated considerable mumbling among Bank economists. My impression from friends among Bank staff at the time was that Lin's approach was seen as a version of misconceived industrial policy, including by the research staff that reported to him. After Lin left the Bank in 2012, he continued to advise the government of Ethiopia from his base at the University of Peking in China, until the political transition in Ethiopia in 2018, when Prime Minister Abiy Ahmed came to power. At the time, I had left the World Bank many years earlier and, in the interval, had become President of the Center for Global Development. But from what I knew

of inside–World Bank views among economists, Lin and his approach were viewed with deep skepticism.

In retrospect, the breadth of this skepticism seems overdone. After all, Lin's focus in Ethiopia was on building up light manufacturing in a special export zone, with tax incentives and more attractive regulatory conditions for local manufacturers. Even at that time, such special export zones were encouraged and supported by the Bank.

Critiques

The fundamental critique of the *Miracle* report has had to do with its definition and treatment of industrial policy, and with its implicit conclusion that for most developing countries, beyond a carefully delimited export push, active involvement of the state in their economies would be risky and potentially counterproductive. But before discussing this issue, consider some other concerns.

The Not-Enough-On Critiques

Some critics of the report wanted more on the three southeast Asian countries that were considered part of the *Miracle*—Indonesia, Malaysia and Thailand—with attention to their impressive success (if lesser than that of the northeast "tigers"). The criticism was fair. But it is worth noting that the World Bank also published a contemporaneous book edited by Leipziger and Thomas (1993) with more detailed studies of each of the seven Miracle countries.

Some wished for inclusion of other countries such as the Philippines, although from the perspective of the time the Philippines did not fit the *Miracle* framework; its low growth and persistently overvalued exchange rate made it similar to many countries of Latin America.[3] In retrospect, it is too bad Vietnam was not included or even mentioned, given its economic success in recent decades, but its growth had not yet taken off by the early 1990s. Also, formal relations between Vietnam and the World Bank started only in 1993, meaning data on Vietnam's economy were limited at the time of the book preparation.

From a modern perspective, it is natural to wish that China had been discussed more explicitly, but although its economic reforms started in the early 1980s, China was still very much a low-income country in the early 1990s. The *Miracle* report does name fear of China on the part of its smaller neighbors, and fear of China's potential influence on Communist insurgencies at home, as giving leaders in the Miracle economies ample reason to focus on growth, and especially on growth that was shared (Campos and Root 1996, pp. 30–44)—helping to fuel, for example, the

---

[3] The *Miracle* report did include a box entitled "Vested Interests Doom Philippine Land Reform," which explains the failure of land reform as the outcome of a weak bureaucracy and poor enforcement of the reform. Landlords found ways to avoid losing control of their land, "confident they had plenty of time to lobby a future administration for more favorable provisions" (p. 169).

extensive investment in public housing in Hong Kong and Singapore (though they were also "a response to the massive influx of refugees and migrants from China", p. 163). (Note that in deference to China, by then a member of the World Bank and International Monetary Fund, Taiwan is referred to throughout the text as "Taiwan, China"; China had become a member of the World Bank in 1980.)

What About Agriculture?

The critical positive role of agriculture was not much explored in the book, except indirectly to the extent that agriculture was not penalized as in many other developing countries by overvalued exchange rates (as in much of Latin America) or by other forms of "urban bias" that kept food prices low for urban workers (as in much of sub-Saharan Africa).

In contrast, Studwell (2013) emphasizes the centrality of agricultural success. He focuses on "gardening"-style agriculture—by which he means households maximizing labor input to maximize output on family farms—to explain the high output of agriculture in the postwar northeast Miracle economies. He argues that high domestic agricultural output minimized use of foreign exchange for food imports in urban areas and *released* labor for work in urban industry, in contrast to the Lewis/Kuznets model of structural transformation pulling labor to higher-wage industry. High agricultural product and productivity may have been the key to those economies avoiding the phenomenon common to most developing countries of a high proportion of urban workers moving out of low-productivity agriculture to the huge informal sectors in developing countries' urban areas—where they can earn more, but not much more, than they could in agriculture (Gollin, Lagakos, and Waugh 2014). As does the *Miracle* report, Studwell emphasizes for Korea and Taiwan the importance of the land reforms supported by the US military in these countries and refers favorably to subsequent work at the World Bank of Deininger (2003) on land distribution and land reform in East Asia and other developing countries. Studwell's (2013) book was published 20 years after the *Miracle* report and does not criticize the report directly. However, his thesis suggests that agriculture and "gardening" was too fundamental to be left out of the discussion.[4]

There is also the question whether the nature of crop cultivation affects social norms in ways that matter for the effects of policy on growth. The *Miracle* report generally avoids any reference to "culture" mattering, laying out instead the "institutional basis" for shared growth, as in the deliberation councils. However, noneconomists have pointed out the likely roots of the high-cooperation societies of the northeast countries of East Asia, including China, in the demands of growing rice—which for example requires collaboration among households on irrigation

---

[4]Referring to the history of implicit support for industry in already high-income countries, Studwell (2013) writes that the southeast Asian economies, in contrast to Japan and Korea, "had to wait until the post-colonial, post–World War era, when international agencies (notably the World Bank) . . . offered up a new 'approved' set of ideas . . . and more abstract economic theory took over, predicated on a belief that all economies . . . are subject to similar rules . . . the 'science' of economics" (pp. 82–83).

systems—compared to wheat and maize (for example, see Talhelm et al. 2014; Talhelm and Dong 2024). With one exception noted earlier, the *Miracle* report, written by economists, abjures any invocation of a "Confucian" culture.

The Total Factor Productivity Dispute

Young (1992, 1995) and Krugman (1994) argued that most of the unusually high growth in the Miracle economies was due to "perspiration" (Krugman's evocative term for accumulation of physical and human capital) rather than to high total factor productivity growth—or gains in the apparent efficiency with which those inputs were deployed. In short, the miracle was more about increases in investment than in hard-to-measure "efficiency."

This argument was in marked contrast to the (largely) contemporaneous view taken in the *Miracle* report. The report, in its analysis and discussion of what accounted for the high growth of these East Asian economies, argued that while high inputs of physical and human investment certainly mattered compared to other developing regions, those underlying factors could not explain the total differences in overall growth: "Controlling for their superior rates of accumulation, the HPAEs [high-performing Asian economies] still outperform the statistical relationship between accumulation and growth . . . . We are able, in the end, to predict only about 17 percent of the actual difference in growth between Latin America and the HPAEs" (p. 54). The conclusion was that the Miracle economies achieved unusually high levels of total factor productivity growth; they did much better at allocating their resources to high-productivity activities and "in mastering and adopting catch-up technologies" than other developing countries (for example, see Figure 1.10 in the report).[5]

Young (1992) singled out Singapore to make his point that productivity growth had been minimal relative to accumulation of physical and human capital. Singapore seemed a good example for him at the time; the *Miracle* authors' estimate of total factor productivity growth (1960–1989) for Singapore of less than 2 percent per year (1960–1989) was indeed low compared to the estimate for the northeast countries (Korea, Taiwan, Japan, and Taiwan) of more than 3 percent (Figure 1.10, p. 56; Figure A1.3, p. 66). However, 30 years later, Singapore (along with its neighbors to the northeast) has escaped the so-called middle-income trap, becoming a high-income country according to the World Bank categories.

The Fundamental Critique: Was Industrial Policy Central?

The fundamental critique of the *Miracle* report was that, although it credited the export-push approach with fostering growth and technological advances ("a

---

[5] Stiglitz (2001, p. 512) later, when he was Chief Economist at the World Bank, called the total factor productivity debate "much ado about nothing," noting there had been "a narrowing of the technology gap . . . " and referring to "deliberate efforts to enhance the transfer of technology . . . . [T]he successful countries made deliberate efforts to enhance the transfer of technology . . . most notably their support for technical education."

winning mix of fundamentals and interventions", p. 22), it greatly understated the centrality of industrial policy to the economic success of the region.

In a long and stinging critique of the report, Rodrik (1994) wrote that the thesis of the book is divided against itself—that the book overdoes the argument throughout that it was the market approach to growth that worked, despite the extensive interventions in markets that the book amply describes. He was particularly critical of the definition of "industrial policy" in the book as a policy meant to change the "sectoral composition of industry." The *Miracle* authors argued there was no such change in Korea—thus writing off any effect of Korea's heavy and chemical industry drive or of Japan's similar program led by its Ministry of Trade and Industry on the buildup of heavy industries (steel, shipbuilding, and others) in those countries. They argued instead that in Korea the most successful sector (with the highest total factor productivity growth, consistent with the book's alternative definition of industrial policy as pushing total factor productivity growth beyond its natural course) was the textile and apparel industry, a "market-conforming outcome"—and thus that "industrial policy" defined as meant to change the sectoral composition of industry in favor of heavy industries had failed. (This argument about the "failure" of industrial policy based on the sectoral composition definition has become harder to defend with the passage of time—should not some effects of Korea's heavy and chemical industry drive be credited with the global success today of Hyundai and Samsung?[6])

In addition, based on Rodrik's own analysis of available data, he criticized the report's lack of attention to the favorable initial conditions for growth in East Asia at the beginning of the 1960s—for example, the good education levels and the relatively equal distribution of land due to postwar reforms, especially in Korea and Taiwan—in comparison to other developing countries and regions. This was largely overlooked in the *Miracle* analysis, though alone it is not necessarily a sufficient explanation for their subsequent "miracle" growth rates measured starting in 1960.

Along similar lines, two well-known experts on Korea and on Taiwan were highly critical of the *Miracle* report's dismissal of industrial policy as effective in those countries. In the 2003 introduction to the paperback edition of his 1990 book on Taiwan, Wade, like Rodrik, dismisses growth of total factor productivity in an industry the state supported as the correct measure of the impact of industrial policy (for example, Wade 1990, p. xx). He argues that the question of whether sectoral policies were effective ought to be based on the effectiveness of the specific policy instruments used to promote the targeted industries (pp. xix, xxii). He points to Taiwan's "below-the-radar" nudging policies, for example to encourage foreign firms to switch supplies from imports to domestic providers (p. xxi), as sectoral rather than across-the-board.[7]

---

[6]Thanks to Danny Leipziger for making this point in a recent conversation.
[7]Wade also cites apparent success in Brazil, Mexico, and Argentina with "sectoral promotion policy" in the auto industry (p. xxi).

Amsden (1989) had written a book on Korea's successful industrialization, in which she told the story of the considerable state role in supporting the country's successful conglomerations of industries (the "chaebols"). In her review of the *Miracle* report (Amsden 1994), she argued that the Bank's plaudits for good macroeconomic policy in Korea overlooked the central role of strong state institutions even in that success. Amsden (p. 627) was also eloquent on the failure of the report to extract lessons for other developing countries from the success of East Asia's interventionist approach: "The greatest disappointment of the *Report's* market fundamentalism is a failure to study seriously how elements of the East Asian model can be adapted to suit conditions in other countries."

Such a study would have been and still could be a useful follow-up to the *Miracle* report, given the World Bank has a key role as adviser as well as lender to developing countries. Yet it is not surprising that no such study has been done, given the skepticism about industrial policy inside the Bank reflected in the report.

Was the Miracle Study Too Skeptical of Industrial Policy for Most Developing Countries?

My own judgment is that it remains unclear that the *Miracle* authors went too far in discouraging other developing countries in other regions from state-led interventions in support of industry, or even from such limited, performance-based support as the East Asian export push—with its modest import protection, subsidized credit, and other violations of free-market religiosity.

The bottom line was not far off given the risks that industrial policy programs could only make sense if and where macroeconomic stability was secure (the *Miracle* book includes several examples of East Asian countries abandoning industrial policy initiatives to preserve macroeconomic stability) and where rent-seeking—both by government bureaucrats selling and by firms buying privileges such as better access to foreign exchange, to credit subsidies, or to place-specific investments in infrastructure and so on—is not a major issue. In fact, macro stability in many developing countries, including East Asia, continued to be vulnerable to domestic as well international shocks leading to financial crises; even Korea and other East Asian economies proved vulnerable to the global financial shocks of the late 1990s and 2008–2009, though relative to other regions of the developing world they recovered more quickly.

On rent-seeking and other aspects of governance, what evidence there is suggests minimal if any improvement in most countries across the world. Kaufmann and Kraay (2024), who have reported on measures of economic and political governance for most developing countries for several decades, find "little evidence of trends in global averages of governance."

Moreover, the worry about the institutional capacity to manage such policies in the developing world persists today in the increasingly sophisticated literature that has emerged in recent years. For example, Juhász, Lane, and Rodrik (2023) argue that there is a "generic and powerful economic case for industrial policy" and conclude that certain types of industrial policy were "powerful in driving structural

change in countries such as Japan, South Korea, and China." But they also conclude that it is "very difficult to derive broad generalizations for other countries and time periods from this experience *without taking institutional differences into consideration*" (my italics), consistent with the cautions for other developing countries three decades ago in the *Miracle* study. Similarly, Barteska and Lee (2024) conclude that the effect of industrial policy during Korea's growth miracle was entirely dependent on bureaucratic capacity—with the key mechanism involving the transmission of information about market conditions—rather than on direct or indirect government subsidies.[8]

In short, there is still no agreed-upon bottom line on industrial policy for developing countries, among economists outside or inside the World Bank. That suggests the *Miracle* book, though in many ways a product of its neoliberal time, and probably overly wedded to the "neoclassical" growth explanations for East Asia's success, was not completely off the mark in its caution about industrial policy for other regions of the developing world.

Does the *Miracle* Report Have Lessons for Today?

That the *Miracle* report's caution about industrial policy for other developing countries was not completely off the mark in 1992–1993 does not mean, however, that 30 years later there are not some developing countries today where governance is in fact above average and some forms of industrial policy could be tried—including special export zones for industry and support for innovation in the service as well as manufacturing industries. And there are many others where the prerequisites of success with industrial policy, such as a laser focus on the quality of primary and secondary education, a well-organized and well-paid civil service, and introducing or strengthening some form of consultative interaction between government officials and businesses, should now be high priorities.

The report itself includes one clear statement about the applicability of lessons from East Asia for other developing regions: "The export push strategy appears to hold great promise for other developing countries" (p. 25). (However, even that conclusion is hedged by concern that the key tactic of credit subsidies to exporters paid for by financial repression would violate international trade rules—still true today under World Trade Organization rules—and by the requirement of a "high level" of institutional capacity to manage such a program.) Three other lessons of the report are worth noting before turning to the question of industrial policy built, as in the Miracle countries, for export success.

A first lesson is the benefit of "shared" growth policies. The World Bank (2014) took up that lesson, not in response to the *Miracle* report, but two decades later in response to a growing body of academic work on the negative effects of inequality

[8]Jaramillo and Kim (2025) discuss positive effects of Korea's first "mission-oriented" research and development program, instituted in 1992 when the *Miracle* report was under preparation.

on growth,[9] when it instituted an annual publication on "shared prosperity" in developing countries.

A second lesson is the potential benefits to an economy of creating a consultative mechanism between the public and private sectors, as in the deliberation councils in many countries of East Asia. Many more developing countries now have the prerequisite of a well-compensated, technically competent cadre of civil servants, and many also have some system of dialogue with private businesses that could be strengthened and institutionalized.

A possible third lesson, even in the digital age, is the value of a "postal savings" scheme. The two key characteristics of such a scheme are that it should be widespread and easily available, thus bringing in substantial amounts of savings across the population, and also that this ease and convenience should make it possible to pay depositors a lower interest rate—a mild form of financial repression. The *Miracle* report credited postal savings in some countries with the high private savings rates that enabled high rates of public and private investment in physical and human capital.

Still, the most important lesson of the *Miracle* report may be the potential of export-oriented industrial policy to spur growth in those developing countries with adequate governance—which nowadays would include many of the countries the World Bank classifies as lower-middle-income (currently defined as with Gross National Income per capita in 2025 from $1,136–$4,495) and upper-middle-income ($4,496–$13,935) (Metreau, Young, and Eapen 2025).

Bangladesh and Vietnam are examples of countries that have become successful exporters of industrial goods in this century. But for most developing countries, expanding that category of exports is increasingly challenging due to "premature deindustrialization" (Rodrik 2016)—that is, developing countries are running out of opportunities to industrialize at lower levels of income than did the advanced economies and the East Asian economies, as robots and other technological advances substitute for low-skilled jobs in manufacturing and as services take up an increasing portion of global consumption. Of course, it has not helped that China in the last few decades has beaten most developing countries to what are declining opportunities for exports of manufactured goods.

With premature deindustrialization, more new jobs everywhere arise nowadays in services: education, health care, elderly care, retail, finance, insurance, logistics, and so on. Perhaps ironically, the objective of "industrial" policy (defined as interventions by the state) in many developing countries in the third decade of the twenty-first century should be to find ways to increase jobs and productivity in the modern "services" sectors. But productive jobs in modern service industries require more education and training—more than those of workers in the earlier period of low-skilled manufacturing jobs in the last century. That makes improving the

---

[9] An early draft of an International Monetary Fund paper, Ostry, Berg, and Tsangarides (2014), may have been particularly influential. See also Lane (2025); using newly assembled data on the introduction and termination of HCI drive policies in Korea, Lane concludes that the drive was in fact successful.

quality of education at the secondary level in middle-income countries, particularly in "technical" skills (mathematics, pre-engineering) an abiding lesson of the *Miracle* report.

Indeed, the *Miracle* authors concluded that the key to the unusually high total factor productivity growth of the East Asian economies, which distinguished them from most other developing countries (as much as their high investment rates at the time), was the "technological learning" associated with their export push. The report credits the export push for that learning, as exporters learned from and through interactions with importing firms abroad.

But behind technological learning at the factory level was the high level of good-quality schooling of managers and workers.[10] In other words, a key lesson from East Asia is the emphasis going back decades on high-quality primary and secondary education and on secondary and post-secondary technical education.

The likely importance of broad-based, educational achievement is worrying for many middle-income countries—including India and Nigeria[11]—where recent tests of primary school graduates ten years after they have completed school find many cannot read a sentence; more worrying still, the results of such testing over the last decade in many countries suggest that the quality of education in those and other countries is actually declining (Le Nestour, Moscoviz, and Sandefur 2022).

Beyond these lessons, the *Miracle* study raises "why?" questions for students and scholars of development. Was fear of China the main reason behind the early emphasis of the Miracle countries on creating and maintaining legitimacy through shared growth? Why did the need for newly formed governments to create legitimacy not encourage shared growth in other regions such as postcolonial Africa; was a foreign threat necessary? Why did shared growth never take hold in Latin America—except perhaps in Costa Rica and Uruguay, where mountains (and smallholder coffee) in the former and grasslands in the latter meant they avoided the plantation-type agriculture associated with colonialism and long-run inequality? Is it a coincidence that Costa Rica and Uruguay have the largest middle classes in the region today? Is a large-enough middle class the key to good governance (Birdsall 2016)? Why were the nations of East Asia politically able to take the sometimes difficult steps needed to maintain macroeconomic stability? Why were the Miracle countries able to institute and strengthen postal savings systems? Why did the system to woo big business—including deliberation councils—emerge in the East Asian region but not elsewhere in the developing world? How important overall was the influence of the successful Japanese example, including on industrial policy?

---

[10] See Appendix 1.2 in the report. Still today, students in Korea, Taiwan, and other East Asian Miracle countries typically do much better than students in other advanced economies on, for example, the internationally comparable PISA (Programme for International Student Assessment) test (for example, see https://worldpopulationreview.com/country-rankings/pisa-scores-by-country).

[11] India and Nigeria are both classified as lower-middle-income economies by the World Bank.

The relative silence of the *Miracle* report on lessons for other developing regions was unfortunate, but forgivable. How much could be said in one book? What matters now is that in the last few years a methodologically sophisticated literature on industrial policy has appeared among economists, with lessons for developing countries (for example, Juhász, Lane, and Rodrik 2023).[12] Economists in the research arm of the World Bank are also contributing. McKenzie (2025) counters many concerns about direct government support for private firms, noting that the market failures that industrial policy targets "have long been at the core of what development economists study" and arguing that recent research in developing countries shows industrial policy "can deliver." Reed (2024b), in a paper on "embedded autonomy" (Evans 1995) in Korea's late twentieth-century computer sector, refers to "public-private dialogue" in Cambodia, Peru, and Tunisia—examples akin perhaps to the deliberation councils of East Asia.

In this journal, a symposium in the Fall 2024 issue described practical issues in industrial policy: how a government might choose among the available industries (Reed 2024a) and the issues of political economy needed to support such a choice (Juhász and Lane 2024). There are also questions of what specific industries a government might choose in a world economy where export-push manufacturing is unlikely to be a widespread answer, what industrial policies a government might choose, how the government might evaluate whether those policies are working, and whether the government will be able to back away from efforts at industrial policy that are not working. In short, the details of what industrial policy would help which developing countries catch the train of technological learning remain unclear. However, the issues, after three decades, are back on the agenda, and the 1993 *Miracle* report may find a new generation of students and scholars eager to explore again whether and how "industrial policy," however defined, can advance the cause of development.

---

[12] See also the papers at this conference at Columbia University in November, 2024: https://ipdcolumbia.org/publication/new-thinking-in-industrial-policy-perspectives-from-developed-and-developing-countries; and this policy brief summarizing some issues at that conference https://ipdcolumbia.org/wp-content/uploads/2025/02/conference_policybrief_v4.pdf.

References

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review* 91 (5): 1369–401.

Alesina, Alberto, and Roberto Perotti. 1996. "Income Distribution, Political Instability, and Investment." *European Economic Review* 40 (6): 1203–28.

Amsden, Alice H. 1989. *Asia's Next Giant: South Korea and Late Industrialization.* Oxford University Press.

Amsden, Alice H. 1994. "Why Isn't the Whole World Experimenting with the East Asian Model to Develop? Review of the *East Asian Miracle.*" *World Development* 22 (4): 627–33.

Barteska, Philipp, Jay Euijung Lee. 2025. *Personnel Is Policy (Implementation): Bureaucrats and the Korean Export Miracle.* Center for Development Economics and Policy.

Birdsall, Nancy. 2016. "Middle-Class Heroes: The Best Guarantee of Good Governance." *Foreign Affairs.* February 15.

Campos, Jose Edgardo, and Hilton L. Root. 1996. *The Key to the Asian Miracle: Making Shared Growth Credible.* Brookings Institution.

Deininger, Klaus. 2003. *Land Policies for Growth and Poverty Reduction.* World Bank.

Evans, Peter. 1995. *Embedded Autonomy: States and Industrial Transformation.* Princeton University Press.

Gelb, Alan, Julia Clark, and Christian Meyer. 2012. "New Structural Economics: Industrial Policy 2.0 or Same Old Structuralism?" Center for Global Development (blog), April 26. https://www.cgdev.org/blog/new-structural-economics-industrial-policy-20-or-same-old-structuralism.

Ghezzi, Piero Eduardo, and Steve Denis Jean Utterwulghe. 2017. "Public-Private Dialogue for Modern Industrial Policies: Towards a Solutions-Oriented Framework." World Bank Working Paper 117468.

Gollin, Douglas, David Lagakos, and Michael E. Waugh. 2014. "The Agricultural Productivity Gap." *Quarterly Journal of Economics* 129 (2): 939–93.

Haggard, Stephan. 2000. *The Political Economy of the Asian Financial Crisis.* Peterson Institute for International Economics.

Initiative for Policy Dialogue. 2024. "New Thinking in Industrial Policy: Perspectives from Developed and Developing Countries." Columbia University. https://ipdcolumbia.org/wp-content/uploads/2025/02/conference_policybrief_v4.pdf.

Jaramillo, Luis F., and Chan Kim. 2025. "Innovation Spurred: Evidence from South Korea's Big R&D Push." Center for Development Economics and Policy Working Paper.

Juhász, Réka, and Nathan Lane. 2024. "The Political Economy of Industrial Policy." *Journal of Economic Perspectives* 38 (4): 27–54.

Juhász, Réka, Nathan Lane, and Dani Rodrik. 2023. "The New Economics of Industrial Policy." *Annual Review of Economics* 16: 213–42.

Kaufmann, Daniel, and Aart Kraay. 2024. "The Worldwide Governance Indicators: Methodology and 2024 Update." World Bank Policy Research Working Paper 10952.

Krugman, Paul. 1994. "The Myth of Asia's Miracle." *Foreign Affairs* 73 (6): 62–78.

Kuznets, Simon. 1955. "Economic Growth and Income Inequality." *American Economic Review* 45 (1): 1–28.

Lane, Nathan. 2025. "Manufacturing Revolutions: Industrial Policy and Industrialization in South Korea." *Quarterly Journal of Economics* 140 (3): 1683–741.

Le Nestour, Alexis, Laura Moscoviz, and Justin Sandefur. 2022. "The Long-Run Decline of Education Quality in the Developing World." Center for Global Development Working Paper 608.

Leipziger, Danny M., and Vinod Thomas. 1993. *The Lessons of East Asia: an Overview of Country Experience.* World Bank.

Lewis, W. Arthur. 1954. "Economic Development with Unlimited Supplies of Labor." *Manchester School* 22 (2): 139–91.

Lin, Justin Yifu. 2011. "New Structural Economics: A Framework for Rethinking Development." *World Bank Research Observer* 26 (2): 193–221.

Lipton, Michael. 1977. *Why Poor People Stay Poor: A Study of Urban Bias in World Development.* Maurice Temple Smith Ltd.

McBeth, John. 2023. "Long, Slow Death of Indonesia's National Plane Dream." *Asia Times*, July 13. https://asiatimes.com/2023/07/long-slow-death-of-indonesias-national-plane-dream/.

McKenzie, David. 2025. "Is There Still a Role for Direct Government Support to Firms in Developing Countries?" *New Zealand Economic Papers* 59 (1): 24–29.

Metreau, Eric, Kathryn Elizabeth Young, and Shwetha Grace Eapen. 2025. "Understanding Country

Income: World Bank Income Classifications for FY26 (July 1,2025–June, 2026)." World Bank Blog.

Ostry, Jonathan D., Andrew Berg, and Charalambos G. Tsangarides. 2014. *Redistribution, Inequality, and Growth*. International Monetary Fund.

Page, John. 2016. "The *East Asian Miracle* and Development Policy: A Twenty-Year Retrospective." In *Japan's Development Assistance*, edited by Hiroshi Kato, John Page, and Yasutami Shimomura, 105–19. Palgrave MacMillan.

Page, John, Nancy Birdsall, Ed Campos, W. Max Corden, Chang-Shik Kim, Howard Pack, Richard Sabor, Joseph E. Stiglitz, and Marilou Uy. 1993. *The East Asian Miracle: Economic Growth and Public Policy*. Oxford University Press.

Reed, Tristan. 2024a. "Export-Led Industrial Policy for Developing Countries: Is There a Way to Pick Winners?" *Journal of Economic Perspectives* 38 (4): 3–26.

Reed, Tristan. 2024b. "Is There a Way to Do Industrial Policy without Strong Institutions?" World Bank (blog), February 4. https://blogs.worldbank.org/en/developmenttalk/is-there-a-way-to-do-industrial-policy-without-strong-institutio.

Rodrik, Dani. 1994. "King Kong Meets Godzilla: The World Bank and the East Asian Miracle." CEPR Discussion Papers 944.

Rodrik, Dani. 2016. "Premature Deindustrialization." *Journal of Economic Growth* 21 (1): 1–33.

Spence, Michael. 2021. "Some Thoughts on the Washington Consensus and Subsequent Global Development Experience." *Journal of Economic Perspectives* 35 (3): 67–82.

Stiglitz, Joseph E. 1996. "Some Lessons from the East Asian Miracle." *World Bank Research Observer* 11 (2): 151–77.

Stiglitz, Joseph E. 2001. "From Miracle to Crisis to Recovery: Lessons from Four Decades of East Asian Experience." In *Rethinking the East Asian Miracle*, edited by Joseph E. Stiglitz and Shahid Yusuf, 509–26. World Bank and Oxford University Press.

Studwell, Joe. 2013. *How Asia Works: Success and Failure in the World's Most Dynamic Region*. Grove Press.

Talhelm, Thomas, and Xiawei Dong. 2024. "People Quasi-Randomly Assigned to Farm Rice Are More Collectivistic than People Assigned to Farm Wheat." *Nature Communications* 15: 1782.

Talhelm, Thomas, Xuemin Zhang, Shigehiro Oishi, Chen Shimin, Dongyuan Duan, Xuehao Lan, and Shinobu Kitayama. 2014. "Large-Scale Psychological Differences within China Explained by Rice Versus Wheat Agriculture." Science 344 (6184): 603–8.

Wade, Robert. 1990. *Governing the Market: Economic Theory and the Role of Government in East Asian Industrialization*. Princeton University Press.

Wade, Robert. 1996. "Japan, the World Bank, and the Art of Paradigm Maintenance: The East Asian Miracle in Political Perspective." *New Left Review* 217: 3–36.

World Bank. 2014. "WB-IMF Report Gauges Progress on Development Goals, Including New Target of Promoting Shared Prosperity of Bottom 40 Percent." Press Release, October 8. https://www.worldbank.org/en/news/press-release/2014/1%8/wb-imf-report-progress-development-goals-promoting-shared-prosperity.

World Bank. 2016. *Poverty and Shared Prosperity 2016: Taking on Inequality*. World Bank.

Young, Alwyn. 1992. "A Tale of Two Cities: Factor Accumulation and Technical Change in Hong Kong and Singapore." In *NBER Macroeconomics Annual*, Vol. 7, edited by Olivier Jean Blanchard and Stanley Fischer, 13–54.

Young, Alwyn. 1995. "The Tyranny of Numbers: Confronting the Statistical Realities of the East Asian Growth Experience." The Quarterly Journal of Economics 110 (3): 641–80.

Credit, Debt-Deflation, and the Great Depression Revisited

Ben S. Bernanke

I t would be difficult to overstate the economic, social, and political impact of the Great Depression of the 1930s. At the worst of the slump, in 1933, as many as one in four workers were unemployed, with many more on short hours and reduced pay. In the absence of an adequate social safety net, millions of Americans contended with poverty, hunger, and homelessness. And yet, even as the jobless searched desperately for work, many of the nation's factories lay idle, with industrial production falling by half [Cb31] and real gross domestic product declining by more than a quarter [Ca6] between 1929 and 1933.[1] Depression-era unemployment was not caused by a shortage of productive capacity.

For economists, the coexistence for more than a decade of idle labor and unused capital—want in the midst of plenty—presented a troubling paradox, and a satisfactory explanation of the Depression proved elusive for many years. An important breakthrough came with the work of Friedman and Schwartz (1963), who blamed the economic contraction on a steep decline in the money supply—the result of misguided monetary policies and waves of bank failures that the Federal Reserve failed to arrest. Subsequent research put the Friedman-Schwartz narrative in an international context (Eichengreen, 1992, is the definitive reference). This later work supported the monetarist thesis by documenting that countries (including the United States) that tied their currencies to a flawed international gold standard, and that consequently adopted tighter monetary policies than appropriate, experienced

■ *Ben S. Bernanke is a Distinguished Senior Fellow at the Brookings Institution, Washington DC. His email address is BBernanke@brookings.edu.*

[1] Bracketed codes refer to data series drawn from the *Historical Statistics of the United States: Millennial Edition* (Carter et al. 2006).

deeper and more extended slumps than countries that left the gold standard earlier. Although no event as complex as the Depression has a single cause, most economists now accept the monetarist *cum* gold standard story as an essential piece of the explanation. Indeed, this framework underlies much of my own work on the period (collected in Bernanke 2024).

However, despite its strengths, the standard monetarist explanation of the Depression has significant gaps. Notably, it provides limited insight into the *monetary transmission mechanism*, the channels through which the fall in the money supply putatively led to *large* and *persistent* declines in US output and employment. Moreover, the monetarist account views a distinctive feature of the Depression—widespread financial distress and disrupted credit markets—through a narrow lens, focusing only on banking panics and their effects on the money stock. But the broader credit crisis of the 1930s was no sideshow. It devastated both bank and nonbank lenders, paralyzed bond and stock markets, and pushed millions of borrowers—including farmers, homeowners, and businesses—into default, bankruptcy, or foreclosure.

In Bernanke (1983), I argued that the credit crisis of the 1930s was, in fact, the primary mechanism through which monetary contraction debilitated the economy. The prelude to the crisis was a substantial buildup in private debt during the 1920s, reflecting, among other factors, post–World War I optimism about US economic prospects. That optimism was shattered when the collapse in the money supply helped trigger an economic downturn in the early 1930s. Because debt contracts are normally written in nominal (dollar) terms, the declines in prices and nominal incomes that accompanied the downturn made it increasingly difficult for many borrowers to service their debts, resulting in escalating rates of delinquency and default.

Given preexisting vulnerabilities, including structural weaknesses in the US banking system and chaotic financial conditions abroad, deteriorating credit quality in turn posed existential threats to banks and other lenders. In particular, many weaker banks succumbed to depositor runs that forced them to close their doors, while surviving banks girded themselves against possible future withdrawals of deposits by stockpiling safe, liquid assets, calling outstanding loans, and reducing new lending. As borrowers struggled and lenders failed or retrenched, refinancing existing debts or obtaining new credit became extremely difficult, with one consequence being reduced private spending on goods and services. Declining demand further depressed production and prices, helping to propagate the economic slump and adding to the financial pressures on both borrowers and lenders. Government policies aimed at normalizing credit markets—rescuing delinquent borrowers, resolving failed institutions, strengthening surviving lenders, and implementing financial reforms—took considerable time to have their full effect, helping to explain why full employment was not restored until the United States entered World War II.

These ideas were not new when I first wrote about the Depression but descended from two traditions in economics. The older tradition goes back at least

to John Maynard Keynes and Irving Fisher. Fisher (1933) believed that the effects of deflation on the real value of nominal debts—a phenomenon he dubbed *debt-deflation*—was by far the most important source of the Depression, an argument he made in testimony before the US House Ways and Means Committee in April 1932. Even earlier, Keynes (1931) had predicted that deflation, through its effects on the ability of borrowers to repay, would force the global banking system into insolvency, with dire economic consequences.[2]

The second, newer tradition on which I drew was the then-evolving literature on the role of asymmetric information in markets, including credit markets (Stiglitz and Weiss 1981). Because lenders have imperfect information about potential borrowers' creditworthiness, they must incur costs (which I called the *costs of credit intermediation*) to screen and monitor borrowers. Intermediation costs are lower when potential borrowers are financially strong. For example, a borrower who is able to make a large down payment or offer adequate collateral will have both greater incentive and greater ability to avoid defaulting on the loan, reducing the need for costly screening and monitoring by the lender. Moreover, as lending institutions are also borrowers, obtaining funds from depositors and other creditors, the financial health of lenders also affects the costs of extending credit. For example, a well-capitalized bank, being safer, is able to borrow more cheaply; it is also generally more willing to make risky loans, because any losses it sustains are less likely to threaten its solvency. Debt-deflation erodes the financial strength of both borrowers (the real burden of their debts increases) and lenders (their asset quality deteriorates but their nominal liabilities are unchanged). If the losses to either borrowers or lenders, or both, are severe enough, the cost of intermediating credit may become so high that private credit markets function poorly, if at all. A contribution of my paper was to relate these two traditions and show how together they help us better understand the Great Depression. In later work with Mark Gertler and others (see references in Bernanke 2024, chap. 1), I explored these ideas more formally and showed how they could be incorporated into standard macroeconomic analysis.

This article revisits the proposition that the credit crisis was a key monetary transmission mechanism during the Depression. I see the evidence for this story today as far stronger than when I advanced it 40 years ago, for two broad reasons.

First, in recent decades a growing literature has focused on the macroeconomic implications of so-called "financial frictions," including asymmetric information, the real costs of bankruptcy and illiquidity, the effects of "fire sales" on asset prices, and other departures of financial markets from the frictionless ideal. Indeed, Akerlof (2019) has argued in this journal that the exclusion of financial frictions from the standard Keynesian-monetarist framework would prove to be that framework's

---

[2] Other economists who discussed the roles of debt and financial distress in the Depression include Hart (1938), Chandler (1970), and Mishkin (1978). In this journal, Calomiris (1993) compared the roles of monetary and financial (that is, credit) factors in the Depression, finding greater support for the latter. Similarly, Hamilton (1987) concluded that credit-market effects best explain the real effects of monetary contraction in the 1930s.

greatest weakness. Not surprisingly, the global financial crisis of 2007–2009 and the subsequent recession accelerated this line of research by demonstrating that credit-market dysfunction can have large macroeconomic consequences, even absent price deflation (Bernanke 2018).

Second, there has been remarkable improvement in recent decades in the data and techniques used in empirical studies of the economic effects of disrupted credit markets, including during the Depression. Advances include the development of detailed microeconomic datasets that allow comparisons across geographic areas and over time, the inclusion of better controls to help isolate credit effects, and the application of new econometric techniques. Most important, I believe, is the emphasis of recent work on identifying causal relationships. Much of the evidence in my original paper was anecdotal and historical, with the only formal econometric tests based on timing relationships among aggregate variables. More recent empirical work on the Depression has made substantial strides in identifying and quantifying causal linkages, using clever instruments or natural experiments applied to microeconomic data. This article can draw on only selected examples from this burgeoning empirical literature. Taken as a whole, however, the new empirical research provides significant support for the view that credit-market dysfunction was a key transmission mechanism during the Depression, and in some other episodes (including the 2007–2009 crisis) as well.

The next section provides a closer look at the credit boom of the 1920s and the crisis that followed. I focus on the financial distress of Depression-era borrowers—who are largely ignored in the standard monetarist narrative—as well as the troubles of banks and other lenders. I then discuss the government's response to the crisis, documenting that contemporary policymakers viewed the normalization of credit flows as a top priority in the fight against the Depression. Finally, I briefly consider some alternative monetary transmission mechanisms that have been proposed and offer a short conclusion.

Historical Perspectives on the Credit Crisis of the 1930s

According to what has become a conventional taxonomy, credit crises have their roots in periods of heavy borrowing and rapid increases in asset prices that make the economic system more vulnerable to external shocks or changes in sentiment (for comprehensive cross-country evidence, see Jordà, Schularick, and Taylor 2013). Not all credit booms end in crises, but when they do, the increased difficulty of refinancing old debts or obtaining new credit depresses the willingness of households and businesses to spend, potentially leading to extended recessions. In severe crises, even households and businesses with no immediate need for credit will hunker down, knowing that credit might not be available if needed in the future.

Private-sector debts did increase rapidly in the 1920s, relative to prewar levels—a development that some observers noted with concern at the time (for example, Persons 1930). Economic optimism motivated much of this borrowing.

Indeed, following a relatively brief recession in 1920–1921, America's economic prospects seemed bright. With Europe rebuilding from World War I, the United States was the world's dominant economy, with a growing middle class eager to purchase automobiles and other mass-produced consumer durables. Short recessions in 1923 and 1926 did not quench this optimism, as they suggested that temporary downturns would not impair longer-term growth. Indeed, real GDP would grow at a 4.9 percent annual clip between 1921 and 1929 [Ca9].

However, the downside of post–World War I optimism was insufficient caution. Lenders loosened credit standards and ran ad campaigns to encourage borrowing. The period also saw rapid financial innovation, including the popularization of new forms of credit (installment loans, personal loans), market-based instruments (mortgage securitizations), and quantitative credit evaluation tools. There were episodes of intense if ill-fated speculation, including land-price booms and busts in Florida and elsewhere and the breathtaking ascent of stock prices at the end of the decade.

By the end of the 1920s, the buildup in private debt had made the economy vulnerable to declines in the price level (deflation) and incomes. Indeed, the deflation, when it came, would be severe: Consumer prices fell by about 27 percent between December 1929 and April 1933 [Cb74], consistent with a 25 percent decline in the M1 money stock [Cj42] over the same period. Together with the broader slowing of the economy that began in 1929, falling prices and incomes left many debtors struggling to make payments, which in turn pushed financial institutions closer to insolvency. The worsening credit crisis helped to convert what initially seemed to be a normal downturn into a deep and persistent depression. The rest of this section takes a closer look at how the crisis played out in some key economic sectors: agriculture, the household sector, nonfarm businesses, and banks.

Agriculture

In 1929, about one-quarter of Americans lived on farms and more than 40 percent lived in rural areas [Aa712, Ac414, Aa699]. Developments in agriculture thus had substantial economic and political impact.

American farmers had already experienced a boom and bust in the 1920s. During World War I and its immediate aftermath, demand for US food exports led farm prices to more than double [Da1337], with farmland values rising sharply [Da27]. To ramp up capacity, between 1914 and 1921 farmers in turn more than doubled their outstanding mortgage and nonmortgage debts, from $6.3 billion to $14.1 billion (*Historical Statistics* 1975 edition, K361, K376). The reliance of farmers on debt-financed expansion continued, with the share of farms with mortgages rising from 35 percent in 1920 to 42 percent in 1930 [Da578]. For farms with mortgages, mortgage debt as a percentage of the value of land and buildings rose from 29 percent in 1920 to 40 percent in 1930, then to 50 percent by 1935 as farm valuations fell [Da579].

However, within a few years after the end of World War I, crop prices began to soften, reflecting the recovery of global food production, increasing protectionism,

productivity gains, and other factors. Lower prices for their products pressured financially overextended farmers and, despite the general prosperity of the 1920s, the decade saw a significant rise in farm defaults and foreclosures. Collateral damage included the failure of thousands of small agricultural lenders over the decade (Alston, Grove, and Wheelock 1994).

The advent of the Great Depression further worsened farmers' finances. Between 1929 and 1932, crop prices fell 58 percent [Da1338]. In 1932, a farmer would have needed to produce 263 bushels of wheat to meet a fixed $100 obligation, compared to 96 bushels in 1929 (Chandler 1970, p. 60). In addition to mortgages, many farmers relied on short-term credit to finance planting and harvesting even as they continued to make payments on debts incurred in the 1920s to expand acreage under tillage and buy new equipment. Farm owners also faced property tax payments, which typically adjusted only slowly to declines in farmland values. A contemporary estimate found that farm income, net of operating expenses, taxes, and interest costs, fell from $4.7 billion in 1929 to only $1.3 billion in 1932 (Forster and Weldon 1934, Table 5). Compounding farmers' woes, the Dust Bowl ravaged the western plains in the middle of the decade, destroying crops and eroding the topsoil, leading farmland values to drop further (Hornbeck 2012).

As farm prices fell and the debt crisis worsened, the foreclosure rate per 1000 farms rose from 14.7 in 1929 to 38.8 in 1933, the latter figure corresponding to more than 200,000 foreclosures within the year. For comparison, decadal averages between 1950 and 1980 were well under 2 foreclosures per year per 1000 farms (Alston 1983, Table 1). Moreover, the foreclosure data for the 1930s almost certainly understate the financial distress in agriculture, as these data refer only to legally completed foreclosure actions and exclude cases in which the farmer ceded the property without contest or filed for bankruptcy. In addition, some potential foreclosures were avoided or delayed by lender forbearance, government credit programs, and state-ordered foreclosure moratoria. A more comprehensive indicator of farm distress is the rate of delinquency. At the beginning of 1933, a remarkable 52 percent of farm mortgage debt was in arrears, often by several years (Bean 1934, p. 4). In part, high delinquency rates reflected the fact that many farmers, unable to meet the down-payment requirements of standard mortgage loans, had taken out second mortgages, usually at high interest rates.

The usual terms of farm mortgages were themselves a source of stress. Except for a limited amount of longer-term credit provided through federal programs, most farm mortgages had a short term—typically five years or less—and were non-amortizing, requiring a balloon payment at maturity. When farmland values were rising in the early 1920s, farmers typically had little difficulty refinancing when their mortgage contracts expired, but when farm values began to fall, refinancing the principal amount became increasingly difficult.

Crushing farm debts and tax burdens led to unrest, political agitation, and occasional violence. In the Midwest, foreclosures and tax sales provoked what amounted to a farmers' strike. Farmers destroyed crops and livestock, picketed crop

deliveries, disrupted auctions of foreclosed farms, and harassed or threatened officials attempting to enforce foreclosure judgments.

By closing off access to credit and reducing discretionary income, the agricultural debt crisis hurt the broader economy by limiting the ability of farmers to invest in their farms and to buy consumer goods. For example, after rapid increases in mechanization in the 1920s, net investment by farmers in tractors and motorized trucks was negligible in the early 1930s [Da623, Da627]. Hausman, Rhode, and Wieland (2019) have recently examined the effects of the crisis on farmers' consumption spending. Following an insight of Temin and Wigmore (1990), these authors looked at the period after the 1933 dollar devaluation, which pushed up crop prices (as discussed further below). As predicted by theory, they found that higher crop prices in a particular county or state were associated with increased local auto purchases (an indicator of overall consumption), with purchases increasing more in areas in which local crop prices had risen the most and in which agricultural debt burdens were particularly large. Hausman, Rhode, and Wieland calculated that increased consumption spending by farmers in 1933 contributed significantly to the rise in aggregate demand that followed the devaluation.

Households

In the Depression era, like today, home mortgages made up the bulk of nonfarm household liabilities. Overall, the 1920s were a banner decade for residential real estate, reflecting rising affluence, rural-to-urban migration, increasing rates of automobile ownership (which promoted the growth of suburbs), and limited housing construction during World War I. Housing starts rose from 247,000 in 1920 to a peak of 937,000 in 1925 [Dc510]. Starts then moderated, declining to 509,000 in 1929, a drop that contributed to the recession that began in August. Starts collapsed to 93,000 by 1933, crushing employment in construction and ancillary industries, including building materials, furniture, and transportation. House prices fell about 30 percent between 1925 and 1933 [Dc826]. Because of the concurrent fall in the general price level, the decline in house prices was not large in real terms, but was much more significant relative to fixed mortgage obligations.

In earlier times, residential housing had typically been acquired for cash. By the 1920s, in contrast, obtaining a mortgage was relatively easy for middle-class families. The most active mortgage lenders were the so-called buildings and loans (B&Ls), the antecedent of savings and loans. In 1929 there were more than 12,000 B&Ls making close to 40 percent of new residential mortgage loans (Fishback et al. 2020). Commercial banks were *not* a major source of residential mortgage credit at the time [Dc983-Dc989], making less than 15 percent of new mortgage loans in the peak lending year of 1928. Until the late 1920s, nationally-chartered banks faced restrictions on their holdings of real estate. However, some state-chartered banks, including the notorious Bank of United States (whose failure in December 1930 triggered a panic) loaded up on mortgages in the 1920s, which would prove disastrous when real estate prices plummeted. Other sources of mortgage credit included mutual savings banks, life insurance companies, and individuals.

Like farm mortgages, residential mortgages in the 1920s typically were not amortizing and thus had to be refinanced at the end of their terms. Down payments on first mortgages were usually in the range of 40–50 percent of the house price. About three-quarters of home mortgage borrowers, unable to cover the down payment, took out a second mortgage as well, usually at a high interest rate (Postel-Vinay 2017).

When the incomes of many homeowners fell in the early 1930s, residential foreclosures rose sharply, from 68,000 in 1926 to roughly 240,000 per year from 1932 through 1935 [Dc1255]. Available evidence suggests that delinquency rates were correspondingly high. A survey of 22 cities, reported in Hart (1938, p. 164), found that, as of January 1, 1934, rates of homeowner default on interest or principal ranged from 21 percent to 62 percent across the cities surveyed. For the nation as a whole, by 1933 probably between one-third and one-half of American homeowners with mortgages were in default. Renters also faced significant financial stress, and major cities saw sharp rises in eviction rates, which in turn led to rent strikes, mass protests, and vigilante action to block evictions. Declining rent collections in turn affected the ability of landlords to make their own mortgage payments. Indeed, the survey of mortgage defaults cited by Hart found that default rates on mortgages on rental properties were as high or slightly higher than those on homeowners' mortgages.

The high rates of mortgage delinquencies and foreclosures, together with the advent of government programs discussed later, virtually eliminated private mortgage lending after 1932. New mortgage loans by private financial institutions in 1933 were about one-fifth what they were in 1928, as banks and mutual savings banks cut lending severely and life insurance companies largely left the mortgage lending business [Dc983, Dc984-9]. Buildings and loans saw their portfolios become stuffed with foreclosed properties, reducing their capacity to make new loans (Fishback et al. 2020). Building and loans ultimately were hit harder even than banks: by 1941, more than half of the B&Ls operating in 1929 had failed. The process of liquidation was slow and painful, with many troubled B&Ls not being wound up until late in the decade (Rose 2014). The B&Ls had begun as cooperatives, and many of their shareholders, often middle-class homeowners, lost their entire stake when their B&L failed.

Another source of mortgage funding, which came into its own in the 1920s, was real estate bonds, an early form of securitization. These bonds helped finance both residential construction and commercial property development. However, with the mortgage crisis, the real estate bond market collapsed. Issuance of new bonds had been $684 million in 1928, near the peak reached in 1925. By 1932, however, issuance was nil as bond values fell sharply (Johnson 1936). The chaos created by home foreclosures, delinquencies, and the failures of lending institutions left a long trail of legal uncertainties about ownership and property rights that took many years to resolve, further delaying recovery in the housing sector (Field 1992).

An additional factor affecting consumer finances was the rapid growth in consumer credit during the 1920s. According to Olney (1999, Table 1), between

1921 and 1929 household nonmortgage debt, including installment debt for mass-market durable goods (automobiles, furniture, appliances) as well as personal loans, rose from 5.5 percent to 9.3 percent of household income. To reduce default risk, installment lenders typically required high down payments and short terms. The penalty for a missed payment was repossession of the item, with no credit for built-up equity. Consumers thus had a strong incentive to keep current with their payments, often prioritizing them over even mortgage payments. The need to service installment debts further restricted consumers' discretionary spending and access to new credit—enough, according to Olney's calculations, to account for a sharp drop in consumption spending in 1930 that was greater than could be explained by declines in income and wealth alone (Temin 1976).

Businesses

The nonfarm business sector also borrowed heavily during the 1920s. Long-term corporate debt increased from $31.8 billion in 1926, the first year for which the *Historical Statistics* includes data, to $50.3 billion [Ch518] in 1930, remaining near the latter figure for the rest of the decade.

Data on the debts of noncorporate businesses are limited. Nevertheless, it seems clear that smaller businesses faced the greatest difficulties in servicing debt and obtaining credit, given that small- and medium-sized firms of the Depression era were generally heavily dependent on banks for working capital, while the largest firms were often net creditors of banks (Richardson and Del Angel 2024). While large firms increased their holdings of liquid assets during the Depression, small firms became very illiquid, affecting their ability to carry on normal business and to service debt obligations (Hunter 1982).

One indication of concern about credit availability for smaller business in the 1930s was the proliferation of both private and public surveys of lenders and businesses. One large survey, US Bureau of the Census (1935), sent questionnaires to 16,500 smaller manufacturers, collectively representing 97 percent of manufacturing firms in 1929, receiving a 46 percent response rate. The survey found that 71 percent of the respondents were or had been regular bank borrowers, and that 45 percent of those reported that they now had difficulty in obtaining working capital. Further, of the firms able to obtain working capital, 40 percent reported that they had no available source for long-term credit. To determine whether the problem was a lack of creditworthy borrowers, the Census also analyzed the financial conditions of reporting firms, ranking them by net-worth-to-debt ratios and Dun and Bradstreet ratings. This exercise revealed that at least 20 percent of the manufacturers reporting difficulty in obtaining credit from customary sources were in excellent financial shape. As the survey was conducted after the stabilization of the banking system, its results suggest a persistent effect of banking problems on credit availability for smaller firms.

Even with better access to banks, larger firms were not necessarily free of credit problems. The issuance of new equity was difficult and expensive; by 1934, new stock issues had fallen to a minimal $35 million [Cj837]. The corporate bond

market continued to function, albeit also at low volumes, with issuance falling from $4.2 billion in 1927 to $570 million in 1933 (Hickman 1953). The reluctance of investors to lend to all but the strongest firms was evident in the widening of credit spreads. In a study of the pricing of high-yield bonds (rated B or lower), which at their 1940 peak made up more than 40 percent of rated bonds, Basile et al. (2015) documented that the cost of capital for firms of low to intermediate credit quality (as measured by the interest-rate spread between high-yield and government bonds) rose substantially early in the Depression and remained high for the rest of the decade.

In further evidence that the corporate bond market became inaccessible for all but the financially strongest corporations, Benmelech, Frydman, and Papanikolaou (2019) used data for 1928–1933 for about 1,000 large firms, separating their sample into firms with long-term debt maturing in 1930–1934 and those without maturing debt in that period. The maturity dates of long-term debt were presumably determined well before the Depression, so splitting the sample in this way provides a natural experiment. If firms with maturing debt have difficulty refinancing, or find doing so prohibitively costly, they might choose instead to reduce costs, including their wage bills. Benmelech, Frydman, and Papanikolaou found quite large differences in employment between firms with maturing debt and those without, implying that even relatively sizable firms found bond issuance difficult and costly.

Banks

Although almost all types of lenders were adversely affected by the Depression, research on lending institutions, following the lead of Friedman and Schwartz (1963), has mostly focused on the banking system and banking panics.

Bank failures had long been common in the United States, but the banking panics of the early 1930s were especially intense, with roughly one-third of all US banks suspending operations between 1929 and 1933. Bank failures or suspensions, which dissipated the local knowledge and personal relationships that facilitated many loans, doubtlessly affected credit availability. Those effects would have been particularly devastating in remote and rural areas served by few banks, or for smaller borrowers dependent on a particular bank. In addition, the freezing or loss of deposits in suspended banks hampered the ability of businesses to make payrolls and of households to service debts. But importantly, bank failures did not capture the full impact of the crisis on economic activity. The altered behavior of surviving banks in the face of heightened run risk was at least as important.

Banks that survived the initial panics understood that some triggers of depositor runs were out of their control. For example, expectations of dollar devaluation likely contributed to the nationwide banking panic of 1933 (Wigmore 1987), while Friedman and Schwartz (1963) point to the confidence-destroying effects of Britain's departure from the gold standard in 1931 and similar events. But if national or international developments were the main source of depositor runs, then panics would have tended to be national in scope. Detailed studies of panic episodes have found instead that most panics were local or regional in nature and could be traced either

to well-publicized problems of individual banks or banking groups, such as Caldwell and Company or the Bank of United States in 1930 (Wicker 1996), or to adverse conditions facing all banks in an area. Even the nationwide 1933 banking panic began with the highly visible collapse of prominent banking concerns in Detroit.

Rather than macroeconomic events, the evidence is strong that, at least before 1933, the condition of each bank's own balance sheet, or the balance sheets of its close affiliates, was the primary determinant of the bank's run risk (Calomiris 2007). This link is supported by careful studies using detailed data on individual banks to understand how features of banks and their portfolios affected the risk of depositor runs in the 1930s. For example, Calomiris and Mason (2003b) found that the composition of bank assets and structural features (like size) that affected banks' ability to diversify effectively predicted failures in the early 1930s. Similarly, using micro-level data, White (1984) identified bank-specific loan losses as a key source of failures in the panic of 1930, and Postel-Vinay (2016) found that large real estate holdings (which in the absence of secondary markets were highly illiquid) were an important predictor of bank failures in Chicago in the 1932 panic. Using a different identification strategy, Mitchener (2005) found that variations in pre-Depression regulatory oversight, and thus in the quality of banks' lending and portfolio management, also helped predict failure rates in the 1930s.[3]

Understanding the link between credit losses and asset illiquidity on the one hand and the risk of depositor runs on the other, surviving banks tried to protect themselves by "scrambling for liquidity," shifting their holdings as quickly as possible toward safer, more liquid assets, like government securities, while reducing new lending and calling or failing to renew existing loans.[4] In general, the value-added of the banking system is its ability to convert short-term liquid funding into long-term, often illiquid loans and investments. In the early 1930s, most banks could no longer fulfill that function.

How important were bank failures and the retrenchment of surviving banks for credit availability and economic activity in the Depression? As mentioned earlier, a strength of the more recent research is the use of detailed microeconomic datasets that allow for sharper comparisons across time and space. For example, using bank-level data, Calomiris and Mason (2003a) found that factors that predicted bank fragility also predicted declines in economic activity at the state and county levels in 1931 and 1932. Mitchener and Richardson (2025), also using bank-level data as well

---

[3] The conclusion that bank runs usually reflect weak bank balance sheets has also been confirmed by studies covering longer periods. For example, Correia, Luck, and Verner (2024), in an analysis covering most US commercial banks from 1865 to 2023, found that, especially in the period before deposit insurance, individual bank failures were highly predictable by bank-specific factors such as rising losses, deteriorating solvency, and an increasing reliance on noncore funding. As mentioned earlier, Alston, Grove, and Wheelock (1994) tied 1920s regional bank failures to credit losses in the agricultural sector.

[4] Interestingly, the link between credit losses and bank failures implies a two-way causal link between credit and money. Monetary contraction affected credit quality through debt-deflation, but credit losses helped bring down banks or made them more conservative, reducing the money supply. This perspective suggests that distressed credit not only amplified the effects of monetary tightness but also led to further declines in the money stock.

as controls for local economic activity, found that the reactions of surviving banks to depositor withdrawals during panic periods accounted for 39 percent of the decline in bank lending between July 1929 and December 1932.

Several researchers have taken advantage of the pyramidal structure of the Depression-era banking system to achieve identification of credit supply effects. At the time, the large so-called money-center banks in New York or Chicago both provided credit locally and acted as correspondents to smaller banks around the country, meaning that they stood ready to accept small-bank deposits during periods of excess liquidity and to lend to their correspondents when they needed extra cash. Money-center banks, with the help of the Federal Reserve, were used to dealing with the cash inflows and outflows associated with the annual agricultural cycle. However, they were not well prepared to deal with cash outflows caused by banking panics in the interior of the country in the 1930s, when regional and rural correspondents tried to reinforce their liquidity buffers. The cash outflows from New York and Chicago banks arising from regional panics were presumably not closely related to economic developments in New York or Chicago. Nevertheless, money center banks reacted to these liquidity drains and the associated increase in the risk of runs by reducing local business loans and shifting their portfolios toward more-liquid assets (Mitchener and Richardson 2013). Likewise, Calomiris and Wilson (2004) found that unexpected outflows of deposits at individual, publicly traded banks in New York City led the money-center banks to de-risk their portfolios, including by reducing local lending.

Several papers, following a strategy due to Rajan and Zingales (1998), identify the effects of Depression-era bank distress on lending and economic activity by considering differences among firms or industries in the extent of their historical dependence on external finance, as opposed to cash reserves built up internally. These papers differ in their details but agree that bank suspensions and the risk-averse behavior of surviving banks had substantial effects on lending, output, and employment (Lee and Mezzanotti 2017; Mladjan 2019; Gorton, Laarits, and Muir 2023). Using a 1934 Federal Reserve survey of both banks and Chambers of Commerce in various localities, as well as county-level measures of bank distress, Carlson and Rose (2015) found that bank failures were a dominant source of business concerns about credit availability, with funding constraints and deposit outflows also playing a role.[5]

---

[5] Studies of the global financial crisis of 2008 provide additional evidence for the effects of banking crises on credit and economic activity; see, for example, Chodorow-Reich (2014) and Duygan-Bump, Levkov, and Montoriol-Garriga (2015). The effect of the global crisis on lending and output in Europe has also been the subject of much research, some of it benefiting from access to micro data, for example on individual loan applications, that is not available in the United States; for examples, see Cingano, Manaresi, and Sette (2016) and Jiménez et al. (2012). Like most American studies, the European research consistently finds large effects of the credit crisis on lending and real activity. In an international context, Dell'Ariccia, Detragiache, and Rajan (2008) study banking crises in 41 developed and emerging-market countries and find that crises cause significant declines in credit and economic growth, especially in emerging markets.

Although government intervention stabilized the US banking system in 1933, bank lending remained stagnant, creating a drag on the recovery. In particular, local bank-borrower relationships were rebuilt only gradually after 1933, slowing the recovery of lending and economic activity in rural and isolated areas (Cohen, Hachem, and Richardson 2021). The lengthy legal process associated with rehabilitating and opening suspended banks also helps explain the slow recovery of bank lending (Anari, Kolari, and Mason 2005).

The Government Response to the Credit Crisis

The most important policy interventions during the Great Depression can be divided into two categories. First, many government programs, such as the Works Projects Administration and the Civilian Conservation Corps, aimed to provide relief and jobs to the unemployed. Although these programs did help many in need, they probably had limited macroeconomic effects because they were small relative to the size of the problem and were accompanied by substantial tax increases under both Hoover and Roosevelt, both of whom advocated balanced budgets (Brown 1956). In an open letter to Roosevelt in the *New York Times* (December 31, 1933), Keynes criticized the president for not understanding the need for deficit-financed fiscal policy in a depression.

My focus here is on the second category of interventions, in which policymakers sought to rehabilitate the financial system and restore normal credit flows. Credit-related policies included not only efforts to stabilize banks and other lenders, federal lending subsidies, and direct government lending to the private sector, but also policies that aimed to raise prices and wages, with the objective of reducing the pressure of debt service on farmers and households. The range and scope of credit-related policies and the accompanying public rationales are evidence that Depression-era policymakers saw the breakdown of the credit system as a major reason for the depth and persistence of the slump.

Stabilization of Financial Institutions

The stabilization of financial institutions and the restoration of normal lending were key objectives of both the Hoover and Roosevelt administrations. Hoover (1952, chap. 3) assigned heavy blame for the Depression to the banking sector, complaining that "commercial banks were permitted to loan excessive amounts of demand deposits on long-term mortgages and to invest in long-term bonds. When stress came, long-term assets could not be quickly liquidated and depositors' demands had to be met by calling in short-term business loans. That paralyzed business and employment." Addressing agriculture, Hoover noted that local banks borrowed from large money-center banks to serve both the seasonal and longer-term credit needs of farmers. But "the moment depression came," he wrote, "the city banks drew in funds from their country correspondents or refused to furnish their usual needs. The insurance companies abruptly ceased lending on mortgages, and

foreclosed wholesale those in default. As a result, farmers and small businessmen were the immediate victims."

Legislation during the Hoover administration to revive private credit flows included the Banking Act of 1932, which expanded the types of assets the Federal Reserve could take as collateral at its discount window, and the 1932 Federal Home Loan Bank Act, which created a mechanism for providing liquidity to mortgage lenders. Of most importance was the Reconstruction Finance Corporation (RFC) Act, signed in January 1932. Financing itself by issuing government-guaranteed bonds, the RFC was authorized to lend to essentially any solvent financial institution as well as to railroads (whose bonds were a key financial asset), against any collateral its board found acceptable. Hoover would later sign legislation to allow the RFC to lend to states and municipalities for the purpose of financing public works.

In its first months, the effectiveness of the Reconstruction Finance Corporation was hampered by its conservative lending practices and allegations of favoritism. In July 1932, the concerns about favoritism led Congress to require that RFC loans be publicly disclosed, which stigmatized the program. However, the deeper problem was that, by 1932, many financial institutions were insolvent or nearly so and consequently needed more capital, not just an infusion of liquidity. In March 1933, Congress authorized the RFC to purchase preferred stock in financial institutions, giving it a chance to stabilize even capital-deficient lenders. By March 1934, the RFC had purchased stock in nearly half of US banks (Mason 2000).[6]

The watershed for the banking system was the March 1933 banking holiday, designed largely by Hoover Treasury staff but proclaimed by Roosevelt in his first days in office. All banks were temporarily closed (most had already been closed by state authorities), supposedly to be reopened only when certified sound by regulators. Bank supervisors could not have inspected thousands of banks in a week or two, so the remarkable success of the holiday—bank panics ended and currency flowed back into the system—no doubt owed much to the implicit government guarantee of banks, which could have been implemented if needed by the purchases of bank stock by the Reconstruction Finance Corporation or by the authority of the Federal Reserve under the banking act to issue emergency currency. The implied government guarantee of the banking system would be reinforced by the creation of temporary federal deposit insurance in June 1933 and the prohibition of interest payments on most deposits, which provided a windfall to banks. Other New Deal legislation aimed at reforming the financial system included the Securities and Exchange Act of 1934, which created the Securities and Exchange Commission, and the Banking Act of 1935, which reorganized the Federal Reserve System and made deposit insurance permanent.[7]

---

[6] The RFC's journey from liquidity provision to purchases of preferred stock seems quite analogous to the sequence of policies used by the government to rescue financial institutions in the 2007–2009 crisis.

[7] Many of the abuses addressed by these pieces of legislation were revealed by the much-publicized "Pecora investigations" of 1932. These originated with hearings by the US Senate Banking Committee, which hired a New York District Attorney named Frederick Pecora to write a report on banking industry abuses. Pecora pursued the topic avidly, leading to an additional month of hearings, and his name

Government Credit Assistance and Lending

Despite the stabilization of the banks and other financial institutions, private lending remained stagnant and foreclosures continued. In response, the federal government created programs to stimulate private lending, to modify existing troubled loans, and to provide credit directly. For example, the Emergency Farm Mortgage Act of May 1933 created programs that made new loans or refinanced existing farm mortgages on preferential terms, including below-market interest rates, extended maturities, and delayed principal payments, with losses to be made up by the Treasury. At their peak, these programs held 1.1 million loans, or about two-fifths of farm mortgages outstanding (Rose 2013). Despite the government's forbearance, however, continued financial stress in agriculture led even the modi-fied loans ultimately to suffer high rates of delinquency (20–30 percent, according to Rose 2013) and foreclosure (9 percent). Supporting these efforts was the Frazier-Lemke Farm Bankruptcy Act of 1934, which increased barriers to farm fore-closure by private lenders; supplementary government lending, including by the Reconstruction Finance Corporation; and the moratoria on both farm and home foreclosures imposed by many states. As government-guaranteed loans replaced troubled loans on their balance sheets, sometimes but not always with principal reductions, private lenders as well as farmers were beneficiaries of these programs.

A parallel lending program was created to help homeowners. The Home Owners' Loan Corporation (HOLC), created in June 1933, issued government-guaranteed bonds and used the proceeds to purchase over a million troubled home mortgages from lenders. As in the farm mortgage program, the HOLC then refinanced the mortgages at lower interest rates and extended maturi-ties. The program's activity peaked in 1934, when it accounted for $2.26 billion out of $3.17 billion new mortgage loans [Dc983, Dc988]. Like the farm mort-gage program, the HOLC was a boon to troubled private lenders who received government-guaranteed mortgages in exchange for impaired loans. However, also like the farm program, the HOLC ultimately did not avert all foreclosures, as the HOLC itself foreclosed on 20 percent of its loans by 1940 (Rose 2011).

New Deal legislation included numerous other programs to support housing and mortgage lending, some of which would lead to permanent changes in the struc-ture of residential mortgage markets and the standard mortgage contract (Green and Wachter 2005). To encourage private lending, the government offered mortgage guarantees through the Federal Housing Administration and the Veterans Adminis-tration. The FHA guaranteed 482,000 mortgages during 1935–1939 [Dc1105]. The National Housing Act of 1934, which created the FHA, also extended deposit insur-ance to savings and loans. A bill appropriating funds for public housing passed in

---

became identified with the resulting report. Pecora's revelations triggered public anger at banks and other financial institutions. One indication of this anger may have been the folk hero status of bank robbers like John Dillinger and Pretty Boy Floyd, both of whom were active in the mid-1930s. It has been alleged that Floyd burned mortgage documents in the banks he robbed, releasing those borrowers from their debts.

1937, and the Federal National Mortgage Association ("Fannie Mae") was created in 1938 with the mission of developing a liquid, national secondary market for home mortgages, the lack of which had proved a major problem for lenders.

Direct government lending to nonfarm business during the New Deal was initially limited, as policymakers expected the newly stabilized banking system to meet business credit needs. But wary bankers retained their preference for safe, liquid assets. Surveys, like the Census survey mentioned earlier, found unmet needs for credit, especially by smaller firms. Congress responded in June 1934 by authorizing both the Federal Reserve and the Reconstruction Finance Corporation to lend directly to businesses. The RFC had a particularly wide mandate, as its loans could be used to expand existing businesses, start new businesses, or even pay off existing debt. By 1939 the RFC, led by Houston banker Jesse Jones, authorized more than 7300 loans to businesses, totaling about $450 million (Klemme 1939)—a significant but not game-changing amount. Interpreting its mandate even more widely, the RFC lent to promote rural electrification, railroad modernization, commercial building, municipal finances (including a loan to Chicago to help pay teachers' back salaries), public works, and more.

Raising Wages and Prices

Both Hoover and Roosevelt understood the adverse effect of falling prices on borrowers. Raising wages and prices, especially crop prices, was thus a policy goal of both administrations. What neither president seemed to understand was that price increases could be an effective remedy for debt-deflation only if they were the result of increased demand, not mandates or restrictions on supply—a point that Keynes also made in his open letter mentioned earlier. Hoover, rejecting unorthodox approaches such as deficit spending or the devaluation of the dollar, focused mostly on supply-side interventions. He set up a Federal Farm Board in 1929, which over the next few years tried, with limited success, to stabilize farm prices by buying up and storing surplus crops. He also harangued private corporations to maintain nominal wages in the face of deflationary pressures.

Roosevelt was much more willing to experiment. Shortly after taking office, he began devaluing the dollar (that is, reducing its official value in terms of gold) in irregular increments—so irregular that Keynes, in his 1933 open letter, characterized Roosevelt's dollar policy as "the gold standard on the booze." Congress approved the devaluation and went further by ending the enforceability of the gold clause in debt contracts—a common clause that, if retained, would have allowed lenders in some cases to demand repayment in the dollar equivalent of gold priced at the old parity of $20.67 per ounce. Uncertainty about the gold clause remained until 1935, however, when the Supreme Court approved Congress's action by a narrow 5-4 vote. The devaluation of the dollar, ultimately to $35 per ounce, together with the stabilization of the banks, helped end deflation and promoted a surge in economic activity.

In retrospect, and consistent with the gold standard theory of the Depression, dollar devaluation has come to be viewed as one of Roosevelt's most

successful initiatives. By increasing the value of the Federal Reserve's gold hold-ings, it increased the scope for expansionary monetary policy. In addition, some economists have argued that the dramatic shift in the monetary framework jolted the public's inflation expectations upward, reducing the perceived real interest rate (Temin and Wigmore 1990; Eggertsson 2008). These interpreta-tions may be substantively correct, but they misconstrue Roosevelt's motivations. For the president, the rationale for the devaluation was to improve the ability of debtors (especially farmers) to service their debts—which in turn, Roosevelt believed, would both help normalize financial conditions and increase demand for manufactures and other goods. The president made this argument frequently. For example, in his second fireside chart (May 7, 1933), Roosevelt noted that millions of people had borrowed money on property "in terms of dollars which had an entirely different [lower] value from the level of March, 1933 . . ." More-over, debtors' problems had implications for the credit system; as Roosevelt put it, before the devaluation "prices for basic commodities were such as to destroy the value of national institutions such as banks, savings banks, insurance compa-nies and others. These institutions, because of their great needs, were foreclosing mortgages, calling loans, refusing credit."

Roosevelt, in his pursuit of higher prices, also appeared indifferent to whether prices rose because of increased demand or reduced supply (or simple mandates). To support commodity prices, for example, he initiated extensive efforts to restrict agricultural production. The Agricultural Adjustment Act (May 1933) paid farmers to take land out of production, while the Commodity Credit Corporation (October 1933) put an effective floor on crop prices. Moving beyond agriculture, as part of the National Industrial Recovery Act of 1933, Roosevelt sought the cartelization of industry and the development of "codes" to set floors on prices and wages. Labor legislation, such as the National Labor Relations Act of 1935, commonly known as the Wagner Act, had broader goals, but the perception that increased labor power would help avoid nominal wage cuts was a factor in the administration's support.

Overall, the policy response to the Depression was a mixed bag. Stabilization of the banking system and devaluation were undoubtedly constructive (although the full rehabilitation of lending institutions would take time), and government lending and refinancing of mortgages and other debts was helpful if not completely successful in restoring normal credit flows. Many structural changes, including deposit insurance, enhanced protection of union negotiations, improvements in the mortgage market, and Social Security, had effects that were mostly positive in the longer term. On the other hand, efforts to raise commodity and other prices by restricting agricultural production and cartelizing industry were likely counter-productive, and fiscal stimulus was too small and offset by tax increases. What these extensive and unorthodox policies make clear, however, is that political leaders, their advisors, and, as best as we can tell, the general public, saw the collapse of private credit markets as a critical issue that had to be addressed if the economy was to recover fully.

Alternative Transmission Mechanisms

This article has argued that debt-deflation and credit-market disruption were key transmission mechanisms through which declines in the money supply and, possibly, other deflationary forces depressed output and employment during the Great Depression. However, economists have proposed at least two alternative transmission mechanisms: (1) high real interest rates, induced by anticipations of deflation; and (2) sticky nominal wages. I discuss each briefly.

If the massive 1930–1933 deflation had been anticipated, then firms and households would have perceived very high real interest rates (nominal interest rates plus expected deflation) early in the downturn, which would have dissuaded them from spending on long-lived goods like machinery and consumer durables.

Was the deflation anticipated? Nelson (1991) and Romer and Romer (2013), based on readings of the business press, found some support for the expected deflation hypothesis, while Cecchetti (1992) used time series analysis to argue that some degree of deflation was statistically predictable at a horizon of three to six months. On the other side, several sources have documented the relative optimism of forecasters and businesspeople in the early phases of the contraction. For example, Klug, Landon-Lane, and White (2005), drawing on real-time economic forecasts of railroad shippers, found that they were surprised by the depth of the downturn and expected a relatively rapid recovery of output and prices. Temin (1976) also documented the optimism of many business leaders. Hamilton (1992) and Evans and Wachtel (1993) presented evidence against the view that deflation was anticipated based on the behavior of commodity futures prices and interest rates respectively.

In this journal, Calomiris (1993) provided further references and a good discussion of the issue. Overall, his assessment that, in the early stages of the Depression, the 1930–1933 deflation could not have been forecasted, if at all, more than a few months ahead—and even in that case only with great uncertainty—seems most plausible. Even less likely is that a deflation as severe and extended as ultimately occurred would have been widely expected. Indeed, the fact that there had been four previous recessions since 1918, all of which proved relatively brief, no doubt led many to believe in 1930 and 1931 that, in a phrase attributed to Hoover early in the Depression, "prosperity was just around the corner."[8]

"Stickiness" in nominal wages is a second possible transmission mechanism. If nominal wages adjust only slowly, then real wages increase when the price level falls, presumably leading firms to reduce employment. Why would firms hesitate to

---

[8] In any case, whether the deflation was anticipated a few months in advance or not is largely irrelevant to the debt-deflation story. Private debt was already high and growing quickly in the 1920s, a period in which inflation and (presumably) inflation expectations were relatively low (but positive) and stable (Fackler and Parker 2005). Much of the debt that was incurred before 1930, including mortgages and corporate bonds, had maturities of five years or more. When the low inflation of the 1920s flipped to deep deflation, debtors were still bound by the terms of their agreements. Thus, debt-deflation effects would have been powerful even if people had suddenly become aware of the threat of severe deflation in the early phases of the downturn.

cut wages in the face of falling demand for their products? In the early stages of the Depression, a possible motivation was the then-prevailing view that—by reducing workers' purchasing power and hence final demand—wage cuts would worsen the downturn and thus be counterproductive for employers as a class.[9] President Hoover was a strong proponent of this doctrine.

A good deal of research has addressed the role of sticky wages in the Depression. In an influential article, Bordo, Erceg, and Evans (2000) found that a small macro model with sticky wages captures the observed decline in output and employment well through early 1932, though it works less well subsequently. They also cite a survey of 1,718 firms by the National Industrial Conference Board (1932) that found that two-thirds of the responding firms had not adjusted their wage scales between December 1929 and December 1931, suggesting significant nominal rigidity in practice.

Despite this and other work, there are reasons to be skeptical of the sticky-wage story.

First, evidence for wage stickiness in the Depression is drawn mainly from manufacturing industries, which may not be representative. In the 1932 NICB survey, for example, 1503 of the 1718 firms surveyed were manufacturers, with most of the remainder in heavy industries such as railroads, utilities, and mining. Wages outside of manufacturing and other heavy industries were much more flexible. For example, average hourly earnings in agriculture fell sharply relative to those in manufacturing (Alston and Hatton 1991). Amaral and MacGee (2017) argue that appropriate accounting for nonmanufacturing industries and for smaller firms with owner-operators eliminates much of the countercyclicality of aggregate real wages in the early 1930s.

Second, even if official wage scales were not changed, firms had many other ways to reduce the effective cost of labor—for example, by demoting more highly-paid workers without changing their duties, by increasing job requirements, or even by firing workers and then rehiring them with lower-paid job titles. Interestingly, a long list of such measures as reported by firms themselves is discussed in the same NICB survey cited by Bordo, Erceg, and Evans (2000) as providing evidence for wage stickiness.

Third, to convert nominal wages to real wages, the usual practice in industry-level studies is to deflate by the wholesale price index for that industry. However, as also pointed out by Amaral and MacGee (2017), wholesale price indices reflect prices for gross output, not value added, which is conceptually closer to the marginal product of a given firm's workers. As the prices of most commodities and some other nonlabor inputs were falling rapidly during this period, properly measured

---

[9] Why maintaining nominal wage schedules would be more important for maintaining workers' purchasing power than avoiding layoffs or reductions in workweeks, both common in the Depression, is not clear. There is also an obvious free-rider problem with this doctrine. O'Brien (1989) argued that implicit and even explicit collusion, together with fears of adverse publicity, kept larger firms from cutting wages unilaterally in the early 1930s.

price indexes for value added in most industries fell less—implying that real wages rose less—than implied by the use of wholesale price indexes to deflate wages. Using a model based on this insight, Amaral and MacGee find that wage stickiness accounts for at most one-quarter of the decline in employment in the early stages of the Depression.

Finally, perhaps the greatest challenge for the sticky-wage hypothesis is identifying causal relationships. If nominal wages are sticky, then by hypothesis most of the variation in real wages over time will be due to changes in the price level. But without additional identifying assumptions, we cannot know whether an observed correlation between price deflation and unemployment reflects deflation's effect on real wages or some other mechanism. Consider the limiting case in which nominal wages are perfectly rigid. Finding a negative correlation between real wages and employment in that case would only reconfirm that deflation was associated with falling employment and would not validate the sticky-wage hypothesis specifically.

Alternatively, one could seek identification by focusing on the relationship of real wages and employment across firms in cross-section. If the sticky-wage hypothesis is true, then firms that were slow to cut nominal wages should have shed more workers, all else equal. If we again turn to the survey cited by Bordo, Erceg, and Evans (2000), however, we find that the opposite result holds, and quite strongly; the survey reports that industries with the largest proportion of firms cutting wages also report the largest declines in employment (National Industrial Conference Board 1932, p. 16). Rather than being a driver of falling employment, it appears that the maintenance of nominal wage scales in the early stages of the Depression was a "luxury good," available only to the largest and most profitable firms in the most concentrated markets, who could afford both to retain their workforces and to delay wage cuts (see also Hanes 2000). For those firms, possible motivations for this behavior included avoiding negative publicity, concerns about worker morale (alternatively, fears of worker resistance or unionization), and the desire to retain and motivate the best workers. In contrast, smaller and less profitable firms came under increasing financial pressure to cut both wages and employment.[10] In short, the heterogeneity of firms' responses to the worsening downturn seems a better explanator of the joint behavior of employment and wages than the sticky-wage hypothesis.

Overall, neither the anticipated deflation nor the sticky-wage hypothesis seem able to account for the powerful impact of deflation on the real economy. In contrast, as discussed in this article, the message of considerable recent empirical work is that credit factors had substantial quantitative impact in the 1930s.

---

[10] Also, firm exits were an important source of employment declines in the downturn. In a study of the motor vehicle industry, Bresnahan and Raff (1991) find that roughly one-third of the decline in employment in that industry between 1929 and 1933 was the result of plant closures, rather than decisions of continuing plants to use less labor. The exits of less profitable firms or plants would bias average wage measures upward even as they reduced employment.

Conclusion

Economic historians today generally agree that monetary factors helped set off the Depression. The question remains why this collapse, which would have no real effect in the textbook macro model with fully flexible prices, created such a large and extended fall in output and employment. This article has revisited the idea that the credit boom of the 1920s, together with various weaknesses of the financial system, made the US economy vulnerable to the declines in prices and nominal incomes of the early 1930s. As predicted by Fisher's (1933) debt-deflation hypothesis, falling prices and incomes severely constrained the ability of farmers, households, and businesses to service existing debts or to obtain new credit, squeezing discretionary income and spending. Meanwhile, worsening credit quality, together with a general loss of confidence in financial institutions, led to failure of or retrenchment by lenders, reducing their capacity to supply credit to the private sector. Government actions to stabilize the financial system and restore credit flows stopped the deflation and initiated a recovery. However, full normalization would take time, and some degree of credit-market dysfunction likely impeded recovery until America's entry into World War II.

Drawing out the lessons of the Depression experience for current economic policy would take another paper. Suffice to say that macroeconomic stability requires financial stability, especially in credit markets. Policymakers should monitor the financial system, particularly in periods of unusually strong credit growth, to ensure that it is sufficiently robust to withstand even large and unexpected shocks.

References

Akerlof, George A. 2019. "What They Were Thinking Then: The Consequences for Macroeconomics during the Past 60 Years." *Journal of Economic Perspectives* 33 (4): 171–86.

Alston, Lee J. 1983. "Farm Foreclosures in the United States during the Interwar Period." *Journal of Economic History* 43 (4): 885–903.

Alston, Lee J., Wayne A. Grove, and David C. Wheelock. 1994. "Why Do Banks Fail? Evidence from the 1920s." *Explorations in Economic History* 31 (4): 409–31.

Alston, Lee J., and Timothy J. Hatton. 1991. "The Earnings Gap between Agricultural and Manufacturing Laborers, 1925–1941." *Journal of Economic History* 51 (1): 83–99.

Amaral, Pedro S., and James C. MacGee. 2017. "Monetary Shocks and Sticky Wages in the U.S. Great Contraction: A Multi-sector Approach." *Journal of Monetary Economics* 92: 112–29.

Anari, Ali, James Kolari, and Joseph Mason. 2005. "Bank Asset Liquidation and the Propagation of the U.S. Great Depression." *Journal of Money, Credit and Banking* 37 (4): 753–73.

Basile, Peter F., Sung Won Kang, John Landon-Lane, and Hugh Rockoff. 2015. "Towards a History of the Junk Bond Market, 1910–1955." NBER Working Paper 21559.

Bean, Louis H. 1934. *Facts Relating to the Agricultural Situation in 1934.* US Bureau of Agricultural Economics.

Benmelech, Efraim, Carola Frydman, and Dimitris Papanikolaou. 2019. "Financial Frictions and Employment during the Great Depression." *Journal of Financial Economics* 133 (3): 541–63.

Bernanke, Ben S. 1983. "Nonmonetary Effects of the Financial Crisis in the Propagation of the Great Depression." *American Economic Review* 73 (3): 257–76.

Bernanke, Ben S. 2018. "The Real Effects of Disrupted Credit: Evidence from the Global Financial Crisis." *Brookings Papers on Economic Activity* 49 (2): 251–342.

Bernanke, Ben S. 2024. *Essays on the Great Depression.* Princeton University Press.

Bordo, Michael D., Christopher J. Erceg, and Charles L. Evans. 2000. "Money, Sticky Wages, and the Great Depression." *American Economic Review* 90 (5): 1447–63.

Bresnahan, Timothy F., and Daniel M. G. Raff. 1991. "Intra-Industry Heterogeneity and the Great Depression: The American Motor Vehicles Industry, 1929–1935." *Journal of Economic History* 51 (2): 317–31.

Brown, E. Cary. 1956. "Fiscal Policy in the 'Thirties: A Reappraisal." *American Economic Review* 46 (5): 857–79.

Calomiris, Charles W. 1993. "Financial Factors in the Great Depression." *Journal of Economic Perspectives* 7 (2): 61–85.

Calomiris, Charles W. 2007. "Bank Failures in Theory and History: The Great Depression and Other 'Contagious' Events." NBER Working Paper 13597.

Calomiris, Charles W., and Joseph R. Mason. 2003a. "Consequences of Bank Distress during the Great Depression." *American Economic Review* 93 (3): 937–47.

Calomiris, Charles W., and Joseph R. Mason. 2003b. "Fundamentals, Panics, and Bank Distress during the Depression." *American Economic Review* 93 (5): 1615–47.

Calomiris, Charles W., and Berry Wilson. 2004. "Bank Capital and Portfolio Management: The 1930s 'Capital Crunch' and the Scramble to Shed Risk." *Journal of Business* 77 (3): 421–55.

Carlson, Mark, and Jonathan D. Rose. 2015. "Credit Availability and the Collapse of the Banking Sector in the 1930s." *Journal of Money, Credit and Banking* 47 (7): 1239–71.

Carter, Susan B., Scott Sigmund Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright, eds. 2006. *Historical Statistics of the United States: Millennial Edition.* Cambridge University Press.

Cecchetti, Stephen G. 1992. "Prices during the Great Depression: Was the Deflation of 1930–1932 Really Unanticipated?" *American Economic Review* 82 (1): 141–56.

Chandler, Lester Vernon. 1970. *America's Greatest Depression, 1929–1941.* HarperCollins.

Chodorow-Reich, Gabriel. 2014. "The Employment Effects of Credit Market Disruptions: Firm-Level Evidence from the 2008–9 Financial Crisis." *Quarterly Journal of Economics* 129 (1): 1–59.

Cingano, Federico, Francesco Manaresi, and Enrico Sette. 2016. "Does Credit Crunch Investment Down? New Evidence on the Real Effects of the Bank-Lending Channel." *Review of Financial Studies* 29 (10): 2737–73.

Cohen, Jon, Kinda Hachem, and Gary Richardson. 2021. "Relationship Lending and the Great Depression." *Review of Economics and Statistics* 103 (3): 505–20.

Correia, Sergio A., Stephan Luck, and Emil Verner. 2024. "Failing Banks." NBER Working Paper 32907.

Dell'Ariccia, Giovanni, Enrica Detragiache, and Raghuram Rajan. 2008. "The Real Effect of Banking Crises." *Journal of Financial Intermediation* 17 (1): 89–112.

Duygan-Bump, Burcu, Alexey Levkov, and Judit Montoriol-Garriga. 2015. "Financing Constraints and Unemployment: Evidence from the Great Recession." *Journal of Monetary Economics* 75: 89–105.

Eggertsson, Gauti B. 2008. "Great Expectations and the End of the Depression." *American Economic Review* 98 (4): 1476–516.

Eichengreen, Barry. 1992. *Golden Fetters: The Gold Standard and the Great Depression, 1919–1939.* Oxford University Press.

Evans, Martin, and Paul Wachtel. 1993. "Were Price Changes during the Great Depression Anticipated? Evidence from Nominal Interest Rates." *Journal of Monetary Economics* 32 (1): 3–34.

Fackler, James S., and Randall E. Parker. 2005. "Was Debt Deflation Operative during the Great Depression?" *Economic Inquiry* 43 (1): 67–78.

Field, Alexander James. 1992. "Uncontrolled Land Development and the Duration of the Depression in the United States." *Journal of Economic History* 52 (4): 785–805.

Fishback, Price, Sebastián Fleitas, Jonathan Rose, and Ken Snowden. 2020. "Collateral Damage: The Impact of Foreclosures on New Home Mortgage Lending in the 1930s." *Journal of Economic History* 80 (3): 853–85.

Fisher, Irving. 1933. "The Debt-Deflation Theory of Great Depressions." *Econometrica* 1 (4): 337–57.

Forster, G. W., and W. C. Weldon. 1934. "The Agricultural Problem." *Social Forces* 12 (3): 359–68.

Friedman, Milton, and Anna Jacobson Schwartz. 1963. *A Monetary History of the United States, 1867–1960.* Princeton University Press.

Gorton, Gary, Toomas Laarits, and Tyler Muir. 2023. "1930: First Modern Crisis." *Financial History Review* 30 (3): 277–307.

Green, Richard K., and Susan M. Wachter. 2005. "The American Mortgage in Historical and International Context." *Journal of Economic Perspectives* 19 (4): 93–114.

Hamilton, James D. 1987. "Monetary Factors in the Great Depression." *Journal of Monetary Economics* 19 (2): 145–69.

Hamilton, James D. 1992. "Was the Deflation during the Great Depression Anticipated? Evidence from the Commodity Futures Market." *American Economic Review* 82 (1): 157–78.

Hanes, Christopher. 2000. "Nominal Wage Rigidity and Industry Characteristics in the Downturns of 1893, 1929, and 1981." *American Economic Review* 90 (5): 1432–46.

Hart, Albert Gailord. 1938. *Debts and Recovery: A Study of Changes in the Internal Debt Structure from 1929 to 1937 and a Program for the Future.* Twentieth Century Fund.

Hausman, Joshua K., Paul W. Rhode, and Johannes F. Weiland. 2019. "Recovery from the Great Depression: The Farm Channel in Spring 1933." *American Economic Review* 109 (2): 427–72.

Hickman, W. Braddock. 1953. *The Volume of Corporate Bond Financing since 1900.* Princeton University Press/NBER.

Hoover, Herbert. 1952. *The Memoirs of Herbert Hoover: The Great Depression, 1929–1941.* Macmillan.

Hornbeck, Richard. 2012. "The Enduring Impact of the American Dust Bowl: Short- and Long-Run Adjustments to Environmental Catastrophe." *American Economic Review* 102 (4): 1477–507.

Hunter, Helen Manning. 1982. "The Role of Business Liquidity during the Great Depression and Afterwards: Differences between Large and Small Firms." *Journal of Economic History* 42 (4): 883–902.

Jiménez, Gabriel, Steven Ongena, José-Luis Peydró, and Jesús Saurina. 2012. "Credit Supply and Monetary Policy: Identifying the Bank Balance-Sheet Channel with Loan Applications." *American Economic Review* 102 (5): 2301–26.

Johnson, Ernest A. 1936. "The Record of Long-Term Real Estate Securities." *Journal of Land and Public Utility Economics* 12 (1): 44–48.

Jordà, Òscar, Moritz Schularick, and Alan M. Taylor. 2013. "When Credit Bites Back." *Journal of Money, Credit and Banking* 45 (S2): 3–28.

Keynes, John Maynard. 1931. *Essays in Persuasion.* Macmillan.

Klemme, Ernest M. 1939. "Industrial Loan Operations of the Reconstruction Finance Corporation and the Federal Reserve Banks." *Journal of Business of the University of Chicago* 12 (4): 365–85.

Klug, Adam, John S. Landon-Lane, and Eugene N. White. 2005. "How Could Everyone Have Been So Wrong? Forecasting the Great Depression with the Railroads." *Explorations in Economic History* 42 (1): 27–55.

Lee, James, and Filippo Mezzanotti. 2017. "Bank Distress and Manufacturing: Evidence from the Great Depression." Unpublished.

Mason, Joseph R. 2000. "Reconstruction Finance Corporation Assistance to Financial Intermediaries and Commercial and Industrial Enterprise in the US, 1932–1937." Preprint, SSRN. http://dx.doi.org/10.2139/ssrn.1337171.

Mishkin, Frederic S. 1978. "The Household Balance Sheet and the Great Depression." *Journal of Economic History* 38 (4): 918–37.

Mitchener, Kris James. 2005. "Bank Supervision, Regulation, and Instability during the Great Depression." *Journal of Economic History* 65 (1): 152–85.

Mitchener, Kris James, and Gary Richardson. 2013. "Shadowy Banks and Financial Contagion during the Great Depression: A Retrospective on Friedman and Schwartz." *American Economic Review* 103 (3): 73–78.

Mitchener, Kris James, and Gary Richardson. 2025. "Bank Lending and Deposit Crunches during the Great Depression." *Journal of Economic History* 85 (2): 442–74.

Mladjan, Mrdjan. 2019. "Accelerating into the Abyss: Financial Dependence and the Great Depression." Preprint, SSRN. http://dx.doi.org/10.2139/ssrn.2366291.

National Industrial Conference Board. 1932. *Salary and Wage Policy in the Depression*. National Industrial Conference Board.

Nelson, Daniel B. 1991. "Was the Deflation of 1929–1930 Anticipated? The Monetary Regime as Viewed by the Business Press." *Research in Economic History* 13: 1–65.

O'Brien, Anthony Patrick. 1989. "A Behavioral Explanation for Nominal Wage Rigidity during the Great Depression." *Quarterly Journal of Economics* 104 (4): 719–35.

Olney, Martha L. 1999. "Avoiding Default: The Role of Credit in the Consumption Collapse of 1930." *Quarterly Journal of Economics* 114 (1): 319–35.

Persons, Charles E. 1930. "Credit Expansion, 1920 to 1929, and Its Lessons." *Quarterly Journal of Economics* 45 (1): 94–130.

Postel-Vinay, Natacha. 2016. "What Caused Chicago Bank Failures in the Great Depression? A Look at the 1920s." *Journal of Economic History* 76 (2): 478–519.

Postel-Vinay, Natacha. 2017. "Debt Dilution in 1920s America: Lighting the Fuse of a Mortgage Crisis." *Economic History Review* 70 (2): 559–85.

Rajan, Raghuram G., and Luigi Zingales. 1998. "Financial Dependence and Growth." *American Economic Review* 88 (3): 559–86.

Richardson, Gary, and Marco Del Angel. 2024. "Business and Borrowing during the Roaring '20s and at the Onset of the Great Depression." NBER Working Paper 32918.

Romer, Christina D., and David H. Romer. 2013. "The Missing Transmission Mechanism in the Monetary Explanation of the Great Depression." *American Economic Review* 103 (3): 66–72.

Rose, Jonathan D. 2011. "The Incredible HOLC? Mortgage Relief during the Great Depression." *Journal of Money, Credit and Banking* 43 (6): 1073–107.

Rose, Jonathan D. 2013. "A Primer on Farm Mortgage Debt Relief Programs during the 1930s." FEDS Working Paper 2013-33.

Rose, Jonathan D. 2014. "The Prolonged Resolution of Troubled Real Estate Lenders during the 1930s." In *Housing and Mortgage Markets in Historical Perspective*, edited by Eugene N. White, Kenneth Snowden, and Price Fishback, 245–84. University of Chicago Press.

Stiglitz, Joseph E., and Andrew Weiss. 1981. "Credit Rationing in Markets with Imperfect Information." *American Economic Review* 71 (3): 393–410.

Temin, Peter. 1976. *Did Monetary Forces Cause the Great Depression?* W. W. Norton and Company.

Temin, Peter, and Barrie A. Wigmore. 1990. "The End of One Big Deflation." *Explorations in Economic History* 27 (4): 483–502.

US Bureau of the Census. 1935. *Survey of Reports of Credit and Capital Difficulties*. US Government Printing Office.

White, Eugene Nelson. 1984. "A Reinterpretation of the Banking Crisis of 1930." *Journal of Economic History* 44 (1): 119–38.

Wicker, Elmus. 1996. *The Banking Panics of the Great Depression*. Cambridge University Press.

Wigmore, Barrie A. 1987. "Was the Bank Holiday of 1933 Caused by a Run on the Dollar?" *Journal of Economic History* 47 (3): 739–55.

Comparing Experimental and Nonexperimental Methods: What Lessons Have We Learned Four Decades after LaLonde (1986)?

Guido W. Imbens and Yiqing Xu

I n 1986, Robert LaLonde published a paper based on part of his PhD thesis (LaLonde 1986), which has had profound impact on both the methodological and empirical literatures on estimating causal effects. As of May 2025, this paper has been cited over 3,000 times, a number that only partially reflects its tremendous influence on the field of causal inference and the credibility revolution (Angrist and Pischke 2010).

The context for LaLonde's (1986) paper was the National Supported Work demonstration program. This program targeted individuals with extremely poor employment prospects: for females, recipients of Aid to Families with Dependent Children; for males, ex-drug addicts, ex-criminal offenders, and high school drop-outs. Although about 10,000 people participated across 15 cities, the evaluation focused on approximately 6,600 participants in 10 cities. Approximately half were randomly assigned to a control group, while the other half received up to twelve months of training in small groups, with a supervisor and a counselor, followed by efforts to place them with outside employers. According to the Manpower Development Research Corporation (MDRC), which operated the program and conducted

■ *Guido W. Imbens is an Applied Econometrics Professor and a Professor of Economics at the Graduate School of Business and the Department of Economics, Stanford University, Stanford, California. He is also a Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. Yiqing Xu is an Assistant Professor in Political Science at the Department of Political Science and a W. Glenn Campbell and Rita Ricardo-Campbell National Fellow at the Hoover Institution at Stanford University, Stanford, California. Their email addresses are imbens@stanford.edu and yiqingxu@stanford.edu.*

the evaluation, the National Supported Work program substantially increased 1978 earnings for female participants, particularly those without prior job experience; in contrast, the effects for male participants were smaller and, for some subgroups, essentially nonexistent (MDRC 1980).

LaLonde (1986) used this experimental evaluation to address a broader methodological question: whether the state-of-the-art nonexperimental evaluation methods of that time could replicate experimental benchmarks—that is, estimates obtained from randomized controlled trials. By nonexperimental methods, we mean statistical approaches that estimate a program's impact using naturally occurring variation in treatment assignment, without randomization or experimental control. LaLonde's conclusion was sharply negative regarding the credibility of the nonexperimental methods he examined. The abstract (p. 604) noted: "This comparison shows that many of the econometric procedures do not replicate the experimentally determined results, and it suggests that researchers should be aware of the potential for specification errors in other nonexperimental evaluations." He concluded (p. 617): "Policymakers should be aware that the available nonexperimental evaluations of employment and training programs may contain large and unknown biases resulting from specification errors."

Around the time LaLonde's paper was published, several other studies made similar points about the credibility of nonexperimental methods, including work by LaLonde's thesis advisers Ashenfelter and Card (1985), as well as Fraker and Maynard (1987) and Heckman, Hotz, and Dabos (1987). One reason LaLonde's study became so influential is that the full data underlying the original analysis later became publicly available. As part of a research project initiated in a 1996 graduate class taught by Imbens and Rubin at Harvard, Dehejia and Wahba (1999a, 2002) retrieved LaLonde's original male subsample data—stored on tapes—by locating a tape reader capable of reading the files. The data are now publicly available at Dehejia and Wahba (1999b). We refer to a subsample of these data that is most commonly used in the literature as the LaLonde-Dehejia-Wahba (LDW) data. More recently, Calónico and Smith (2017) also reconstructed the female samples. These public datasets are highly valuable for both teaching and future research and are provided alongside this paper in the Supplemental Appendix and Imbens and Xu (2025), with code to replicate the estimates reported here.

The methodological literature on nonexperimental evaluation has advanced significantly. Nearly four decades after LaLonde's paper, the answer to his original question—whether nonexperimental methods can successfully replicate experimental benchmarks—is more nuanced than his initial conclusion: Sometimes they can, and we now have better tools to help assess when they are likely to succeed.

Here are five key lessons that have emerged from this literature since the publication of LaLonde (1986). First, the unconfoundedness assumption has become central to modern methods for nonexperimental data. Unconfoundedness means that, conditional on pretreatment variables or covariates, treatment is as if randomly assigned. This assumption elegantly separates the underlying

identification assumptions from functional form considerations. It emphasizes design—how treatment is assigned—over full specification of the data-generating process.

Second, inspecting and ensuring overlap is critical in the process of estimating treatment effects. Overlap means that the covariate distributions in treatment and control groups have common support. Improved overlap reduces sensitivity to the choice of estimation strategy and improves the robustness of estimates.

Third, the use of propensity scores, defined as the probability of receiving treatment given observed covariates, has become widespread. The propensity score, introduced in Rosenbaum and Rubin (1983a), had only just entered the economics literature when LaLonde (1986) was writing his thesis. Since then, propensity scores have become a core component of many analyses that rely on unconfoundedness, including for assessing overlap, but also for estimation.

Fourth, researchers have increasingly taken heterogeneity in treatment effects seriously. This includes both methods for estimating average treatment effects in the presence of heterogeneity, but also going beyond average treatment effects to examine estimates of treatment effect heterogeneity, such as conditional treatment effects and quantile treatment effects (for example, Wager and Athey 2018). These approaches provide a deeper understanding of how the treatment affects different individuals or segments of the outcome distribution.

Finally, validation exercises—particularly placebo tests—are essential for assessing key assumptions and evaluating the credibility of causal claims. In a placebo test, researchers use an outcome that should not be affected by the treatment. For example, LaLonde (1986) used 1975 earnings as a placebo outcome, because it predates the training program. A non-zero estimated effect on the placebo outcome indicates potential unobserved confounding. Placebo tests have been important diagnostic tools in modern nonexperimental research since the credibility revolution.[1]

To illustrate these lessons, we reexamine the original and reconstructed LaLonde datasets. We show that, once overlap is ensured, various modern methods yield similar estimates. However, these estimates lack a causal interpretation if unconfoundedness is violated. Validation exercises such as placebo tests are critical for assessing the credibility of unconfoundedness. For the LaLonde datasets, placebo estimates generally do not support the unconfoundedness assumption.

---

[1] We focus on five issues specific to the setting studied in LaLonde (1986): the evaluation of an individual-level intervention using detailed background information on participants. While the broader literature on causal inference has expanded substantially over the past four decades (for an illustration of the changes in methodologies used over this period, see Currie, Kleven, and Zwiers 2020), we do not attempt to cover that broader landscape here. For more comprehensive overviews, see surveys such as Imbens and Wooldridge (2009) and Abadie and Cattaneo (2018), as well as a number of textbooks (Angrist and Pischke 2008; Imbens and Rubin 2015; Cunningham 2021; Huntington-Klein 2021; Huber 2023; Ding 2024; Wager 2024; Chernozhukov et al. 2024).

LaLonde's Data and Findings

We begin by describing the data used in LaLonde (1986) and outlining the main econometric approaches and results presented in the paper. We then introduce the LaLonde-Dehejia-Wahba data and examine some of the contemporary methodological responses to LaLonde's findings.

LaLonde's Data

Although the evaluation by the Manpower Development Research Corporation included 6,600 individuals, the experimental samples LaLonde used to evaluate nonexperimental methods were much smaller. For males, the sample included 722 individuals (297 treated and 425 controls), and for females, 1,158 individuals (600 treated and 585 controls). This substantial reduction in sample size stemmed from three main factors. First, many participants failed to complete follow-up interviews needed to collect post-program earnings data. Second, due to budget constraints, the MDRC team randomly selected subsamples for the 27- and 36-month follow-up interviews. Third, LaLonde excluded male participants who entered the program before January 1976 or were still enrolled in January 1978.

To conduct a nonexperimental analysis of the effects of the training program, LaLonde combined the experimental treatment group with comparison groups drawn from external, nonexperimental datasets used as controls. For both female and male participants, he used two main sources. The first, CPS-SSA-1, is based on Westat's Matched Current Population Survey–Social Security Administration File and includes all females or males under age 55 who met Westat's eligibility criteria. The second, PSID-1, comes from the Panel Study of Income Dynamics and includes all female or male household heads under age 55 from 1975 to 1978 for males and 1979 for females, excluding those who identified as retired in 1975. The age cutoff was intended to improve comparability between the comparison group and the experimental sample. In Table 1, columns 1–4 present summary statistics for the male samples in the experimental data and the comparison groups LaLonde used. Respondents in both the CPS and PSID comparison groups were, on average, older, more educated, substantially less likely to be high school dropouts, more likely to be married, and less likely to be Black or Hispanic than participants in the National Supported Work experiment. They also had much higher earnings in the years before the program. For example, average 1975 earnings were $3,000 in the experimental data, compared to $14,000 in the CPS-SSA-1 sample and $19,000 in the PSID-1 sample. Notably, the comparison groups are much larger: the male CPS-SSA-1 sample includes 15,922 observations, and the male PSID-1 sample contains 2,490 observations. To improve comparability with the experimental sample, LaLonde dropped from these datasets individuals based on employment status, time of survey, and poverty status, creating four additional, smaller comparison groups.

Columns 5 and 6 of Table 1 present summary statistics for the experimental treated and control units in the LaLonde-Dehejia-Wahba data (discussed further

Table 1

Descriptive Statistics: LaLonde and LaLonde-Dehejia-Wahba Male Samples

	LaLonde experimental		LaLonde comparison groups		LaLonde-Dehejia-Wahba experimental	
	Treated	*Control*	*CPS-SSA-1*	*PSID-1*	*Treated*	*Control*
	(1)	(2)	(3)	(4)	(5)	(6)
Age	24.63	24.45	33.23	34.85	25.82	25.05
	(6.69)	(6.59)	(11.05)	(10.44)	(7.16)	(7.06)
Years of school	10.38	10.19	12.03	12.12	10.35	10.09
	(1.82)	(1.62)	(2.87)	(3.08)	(2.01)	(1.61)
Proportion high school dropouts	0.73	0.81	0.30	0.31	0.71	0.83
	(0.44)	(0.39)	(0.46)	(0.46)	(0.46)	(0.37)
Proportion married	0.17	0.16	0.71	0.87	0.19	0.15
	(0.37)	(0.36)	(0.45)	(0.34)	(0.39)	(0.36)
Proportion Black	0.80	0.80	0.07	0.25	0.84	0.83
	(0.40)	(0.40)	(0.26)	(0.43)	(0.36)	(0.38)
Proportion Hispanic	0.09	0.11	0.07	0.03	0.06	0.11
	(0.29)	(0.32)	(0.26)	(0.18)	(0.24)	(0.31)
Real earnings in 1975 (thousand)	3.07	3.03	13.65	19.06	1.53	1.27
	(4.87)	(5.20)	(9.27)	(13.60)	(3.22)	(3.10)
Proportion unemployed in 1975	0.37	0.42	0.11	0.10	0.60	0.68
	(0.48)	(0.49)	(0.31)	(0.30)	(0.49)	(0.47)
Real earnings in 1974 (thousand)	NA	NA	14.02	19.43	2.10	2.11
			(9.57)	(13.41)	(4.89)	(5.69)
Proportion unemployed in 1974	NA	NA	0.12	0.09	0.71	0.75
			(0.32)	(0.28)	(0.46)	(0.43)
Observations	297	425	15,922	2,490	185	260

Note: Standard deviations are shown in parentheses. The experimental data were drawn from the National Supported Work program. Tables 3, 5, and 6 in LaLonde (1986) use the data summarized in columns 1–4, while Dehejia and Wahba (1999a) primarily rely on the data in columns 3–6.

below), a subset of the original LaLonde male sample selected for containing information on 1974 earnings. Our reanalysis focuses on this dataset because it is widely used in the post-LaLonde methodological literature and includes two pretreatment outcomes—earnings in 1974 and 1975—which enable adjustment for longer earnings histories and placebo analyses. Results for the original LaLonde male samples and reconstructed female samples are reported in the Supplemental Appendix.

Econometric Approaches in LaLonde (1986)

To estimate the causal effect of the National Supported Work program on 1978 earnings using both experimental and nonexperimental data, LaLonde employed a variety of models that fall into two broad categories: regression methods, in which earnings serve as the outcome (referred to as the "earnings equation"), and selection models that also include a "participation equation," where the outcome is program participation. In LaLonde's paper, Tables 4 and 5 present the regression

results, while Table 6 reports the selection model results. These tables are reproduced in the Supplemental Appendix.

LaLonde reported training effect estimates for both female and male participants using seven estimators across six comparison groups: two large groups described in Table 1 and the aforementioned four subgroups defined through additional selection criteria to improve overlap with the National Supported Work demonstration sample. All regression models were linear and implicitly assumed constant treatment effects. Of the seven models, two are simple regressions estimated with or without controls for age, education, and race (but notably excluding 1975 earnings). Two models use a difference-in-differences estimator, replacing the outcome with the change in earnings between 1975 and 1978, estimated with and without controlling for age. Two more use a quasi-difference-in-differences specification, including 1975 earnings on the right-hand side to account for transitory shocks—known as the "Ashenfelter dip" (Ashenfelter 1978)—again estimated with and without controlling for age. The final specification includes all available pretreatment covariates, including 1975 earnings, 1975 unemployment status, and marital status.

In an additional analysis in the spirit of the modern causal literature, LaLonde also conducted placebo tests using 1975 earnings as the outcome. Because the training occurred after 1975, the true treatment effect on this placebo outcome is zero. Indeed, the placebo estimates from the linear regression approach on the experimental sample were close to zero.

LaLonde emphasized several key findings from the linear regression results. First, using the experimental data, all seven estimators produced similar estimates: around $851 for female participants and $886 for male participants. Second, when using nonexperimental comparison groups, the estimates diverged sharply from these benchmarks, often yielding large, negative values in both female and male samples, with modest standard errors. Third, the estimates from nonexperimental data varied widely across specifications, and goodness-of-fit tests offered little guidance for selecting models that aligned with the experimental results. Taken together, these findings led LaLonde to conclude that the regression adjustment methods commonly used at the time were not credible when applied to nonexperimental data.

In addition to the regression models, which rely on exogeneity of the treatment indicator, LaLonde also presented results from selection models that allow for potential endogeneity. These models use the two-step estimator proposed by Heckman (1978), which permits correlation between the error terms in the earnings and participation equations. Identification in this framework relies either on exclusion restrictions—variables that appear in the participation equation but not in the earnings equation—or on functional form and distributional assumptions (linearity and joint normality of the error terms). For both female and male samples, LaLonde used three comparison groups (the experimental controls, CPS-SSA-1, and PSID-1) and estimated four specifications for females and three for males. Each specification used a different set of excluded variables. A priori, no single specification was clearly more defensible based on economic or econometric reasoning.

LaLonde found that the selection model estimates using the experimental data remained close to $851 for females and $886 for males. However, those using nonexperimental data again varied substantially and deviated from the experimental benchmarks. He concluded that, although the two-step procedure brought the estimates somewhat closer to the benchmarks, it still yielded a "considerable range of imprecise estimates" (LaLonde 1986, p. 617).

The LaLonde-Dehejia-Wahba Data

The data reconstructed by Dehejia and Wahba (1999a) focused exclusively on male participants, noting that "estimates for this group were the most sensitive to functional-form specification" (p. 1054). They constructed a subsample from LaLonde's original data consisting of individuals with available information on 1974 earnings and unemployment status. They argued that this subsample remained a valid experimental sample because it was constructed solely using pretreatment information—such as month of assignment and employment history—thereby preserving the orthogonality of treatment assignment with respect to observed and unobserved characteristics. Notably, this subsample includes only 62 percent of the original treatment group in LaLonde's study. For nonexperimental controls, they used subsets of the same datasets LaLonde employed, restricted to units with 1974 earnings and unemployment data. This collection, now widely known as the LaLonde-Dehejia-Wahba (LDW) data, has become a standard benchmark in the causal inference literature. In particular, most methodological studies focus on the version combining the experimental treated units with CPS-SSA-1 controls. For brevity, we refer to the LaLonde-Dehejia-Wahba experimental sample as LDW-Experimental, and the samples combining these treated units with CPS-SSA-1 and PSID-1 controls as LDW-CPS and LDW-PSID, respectively.

As noted earlier, columns 5 and 6 of Table 1 report summary statistics for the LDW-Experimental sample. Compared to LaLonde's original male sample, participants in this subsample had higher unemployment rates and lower average earnings in 1975. The inclusion of 1974 earnings—available only in the LaLonde-Dehejia-Wahba data—further suggests that many participants faced long-term unemployment. These differences may help explain why the estimated training effect in this sample ($1,794) is more than twice that of LaLonde's original male sample ($886).

Subsequent Literature

The publication of LaLonde (1986) sparked a debate in the applied econometrics literature. For example, Heckman and Hotz (1989) responded to LaLonde's critique of nonexperimental evaluation methods by advocating the use of specification tests to rule out particularly poor estimators. However, this approach did not offer a clear way to distinguish among the many estimators that fit the data reasonably well but rely on different identification assumptions. As a result, it gained limited traction in subsequent work.

In contrast, Dehejia and Wahba (1999a) proposed alternative, more flexible estimators to address LaLonde's challenge. Their proposals included approaches

based on propensity score stratification and matching. Their estimates closely matched the experimental benchmark, leading them to conclude (p. 1062), in sharp contrast to LaLonde: "[T]he estimates of the training effect for LaLonde's . . . dataset are close to the benchmark experimental estimates and are robust to the specification of the comparison group and to the functional form used to estimate the propensity score. Our methods succeed for a transparent reason: They use only the subset of the comparison group that is comparable to the treatment group, and discard the complement."

The contrast between the conclusions of Dehejia and Wahba (1999a) and LaLonde (1986) led to a wave of methodological research aimed at probing and reconciling these findings. We turn to these developments in the next section.

Methodological Improvements since LaLonde (1986)

Over the past four decades, the applied econometrics literature has made significant progress—most notably in clarifying estimands (what is being estimated) and the assumptions required for identification (how to use data to recover them). Once these assumptions are clearly stated, researchers can assess how plausible they are. To facilitate discussion of methodological advances since LaLonde (1986), we begin with the "potential outcomes framework," then introduce the main causal estimand and a closely related statistical estimand. We then outline the key assumptions required for nonexperimental data to identify the causal estimand.

Causal and Statistical Estimands

In our discussion, we begin with the potential outcome framework, originally developed by Jerzy Spława-Neyman (1990) in the context of randomized experiments and extended to nonexperimental settings by Donald Rubin (Rubin 1974, 2006; Imbens and Rubin 2015). For each individual i, two potential outcomes are postulated: $Y_i(0)$ and $Y_i(1)$. In the LaLonde setting, they represent the individual's 1978 earnings had the individual not participated in the program and had the person participated, respectively. The causal effect for individual i is the difference $Y_i(1) - Y_i(0)$. Let W_i denote the binary treatment indicator, equal to 1 if the individual participated in the program and 0 otherwise. The observed outcome is linked to the potential outcomes by the identity $Y_i = (1 - W_i)Y_i(0) + W_iY_i(1)$. We also observe pretreatment characteristics or covariates X_i. In the LaLonde study, these include age, years of schooling, high school dropout status, marital status, and indicators for Black and Hispanic backgrounds.[2] Following Dehejia and Wahba (1999a), we may augment this vector to include 1974 and 1975 earnings and indicators for zero earnings (unemployment) in each year.

---

[2] The modern methodological literature has paid particular attention to settings where the vector of pretreatment variables is high-dimensional.

Our primary goal is to estimate the average treatment effect on the treated (ATT), which represents the individual-level causal effect averaged over for those who actually received the treatment.[3]

$$ATT = E[Y_i(1) - Y_i(0) \,|\, W_i = 1] = E[Y_i \,|\, W_i = 1] - E[Y_i(0) \,|\, W_i = 1].$$

In LaLonde's context, it means the average treatment effect of the job training program on individuals who actually participated. Because $Y_i(1) = Y_i$ for treated individuals (individuals with $W_i = 1$), the second equality follows. The first term on the right-hand side, $E[Y_i \,|\, W_i = 1]$, can be directly estimated from the observed data. The second term, $E[Y_i(0) \,|\, W_i = 1]$, cannot, because it involves counterfactual outcomes—in this case, what participants' earnings would have been had they not participated in the program.

Most analyses of the LaLonde data that explicitly allow for heterogeneous treatment effects focus on the ATT, as it makes little sense to estimate or even contemplate the effect of the program for nonparticipants with stable jobs and high earnings. In other contexts, researchers may be interested in the average treatment effect (ATE), which captures the average effect across the entire population, while the ATT focuses specifically on the treated group. LaLonde (1986) did not explicitly distinguish between different estimands, as his analysis did not consider treatment effect heterogeneity.

Because we cannot observe the counterfactual outcomes, we cannot directly estimate the ATT. This is the core of what Holland (1986) termed the "fundamental problem of causal inference." To make progress, we can instead estimate the covariate-adjusted difference in average outcomes between treated and control groups:

$$\text{(Covariate-adjusted difference)} \quad E[Y_i \,|\, W_i = 1] - E[E[Y_i \,|\, W_i = 0, X_i] \,|\, W_i = 1].$$

We refer to this as a *statistical estimand*, as opposed to a *causal estimand*, because it can be estimated from observed data when overlap holds, an assumption to which we return below. In LaLonde's context, the statistical estimand represents the difference between the average 1978 earnings of individuals who participated in the program and the average 1978 earnings of nonexperimental individuals with similar observed characteristics.

Unlike a statistical estimand, ATT is a causal estimand because it compares potential outcomes for the same individuals—specifically, those who participated in the program. These two estimands are not necessarily equal. The counterfactual mean for the treated individuals, $E[Y_i(0) \,|\, W_i = 1]$, may differ from $E[E[Y_i \,|\, W_i = 0, X_i] \,|\, W_i = 1]$, a weighted average of observed outcomes for individuals in the nonexperimental control group with similar characteristics. For

---

[3] For simplicity, here we do not distinguish between a finite study population and a super-population.

the statistical estimand to recover ATT, a key assumption—unconfoundedness—must hold. That is, if unconfoundedness holds, we can approximate the average untreated outcome for the treated individuals using outcomes from control individuals of similar observed characteristics.

A substantial body of subsequent research has focused on improving methods for estimating the covariate-adjusted difference, particularly in high-dimensional settings. Many recent approaches draw on machine learning techniques to estimate the relationships between covariates, treatment, and outcomes in a flexible manner. Formal results often rely on additional regularity conditions, such as the smoothness of conditional means and propensity scores. For formal treatments, see the references in Imbens and Wooldridge (2009) and Abadie and Cattaneo (2018). The methodological advances in this literature apply well beyond the LaLonde setting.

Our focus here is on the plausibility of the key assumptions. In particular, if unconfoundedness does not hold, we may still be able to estimate the covariate-adjusted difference robustly and precisely, but it cannot be interpreted as a causal effect and may have little substantive relevance.

Unconfoundedness

The unconfoundedness assumption was first explicitly introduced by Rubin (1978) as part of the "ignorable treatment assignment" concept (see also Rosenbaum and Rubin 1983a). If unconfoundedness holds, then treatment assignment can be viewed as effectively random once differences in observed covariates between the treatment and control groups are adjusted for. This assumption plays a central role in identifying causal effects from nonexperimental data. In the context of estimating the average treatment effect on the treated (ATT), unconfoundedness is formally stated as:

$$(\text{Unconfoundedness}) \quad W_i \perp Y_i(0) \,|\, X_i,$$

which means that treatment status is conditionally independent of the control potential outcome given the covariates. In LaLonde's context, unconfoundedness means that once we account for observable characteristics (including age, education, race, marital status, and prior earnings), whether someone was in the experimental treatment group or nonexperimental comparison group provides no additional information about what their 1978 earnings would have been had they not participated in the program.

Unconfoundedness is also referred to as conditional independence (Lechner 1999, 2002), or informally as exogeneity (Imbens 2004), or as selection on observables (Barnow, Cain, and Goldberger 1980). However, its definition departs from traditional econometric definitions of exogeneity, which are expressed in terms of residuals and specific functional forms. By contrast, the formal statement of unconfoundedness avoids functional form assumptions, focusing instead on the treatment assignment mechanism. It allows researchers to separate the essence of

the identification assumptions from functional form considerations, emphasizing the role of design rather than the full specification of the data-generating process.

A key result from Rosenbaum and Rubin (1983a) shows that unconfoundedness implies that conditioning on the scalar propensity score is sufficient to remove bias from covariate imbalance. This dimensionality reduction—from the full vector X_i to a single index—has made the propensity score a cornerstone of many modern estimators.

When the parametric model for the conditional expectation of the outcome—such as the earnings equation used in LaLonde (1986)—is correctly specified, unconfoundedness implies a zero conditional mean for the error term. Thus, the results reported in Tables 4 and 5 of the original LaLonde (1986) paper can be viewed as relying on a combination of unconfoundedness and correct functional form assumptions. At the time, however, the nonparametric framing of the identifying assumptions and the emphasis on assignment mechanisms were not yet standard in applied work.

In practice, unconfoundedness is a strong assumption, and its plausibility depends heavily on the context. When the treatment assignment mechanism is poorly understood, this assumption may not be credible. Nonetheless, often researchers can assess its plausibility through supplementary analyses. Tools such as placebo tests and sensitivity analyses can help probe the credibility and robustness of causal claims that rely on unconfoundedness. For a general discussion, see Rosenbaum and Rubin (1983a) and Imbens (2004).

In the next section, we illustrate the use of placebo tests to evaluate unconfoundedness. In the LaLonde setting, the ten covariates are clearly pretreatment variables and should be included in any adjustment strategy. In other settings, however, whether to adjust for a given covariate should is less obvious. Rosenbaum (1984) cautions against adjusting for post-treatment variables, and Cinelli, Forney, and Pearl (2022) offer guidance on selecting from among valid pretreatment covariates for causal inference.

Overlap and Balance

To identify the average treatment effect on the treated (ATT)—and to ensure that the statistical estimand is properly defined—we require an overlap assumption, which states that the propensity score is strictly less than one:

$$\text{(Overlap)} \quad \Pr(W_i = 1 \mid X_i) < 1.$$

In LaLonde's context, this assumption means that for any individual i assigned to the treatment group with a covariate profile $X_i = x_0$, there must also be individuals in the nonexperimental comparison group with the same profile; otherwise, $\Pr(W_i = 1 \mid X_i = x_0) = 1$, violating the overlap assumption. Overlap ensures that the weighted average $E[E[Y_i \mid W_i = 0, X_i] \mid W_i = 1]$, the second term in the statistical estimand, is well-defined. When overlap is violated, it can be restored by trimming

the treatment group—that is, by removing treated units whose covariate profiles are not represented in the control group.[4]

Overlap implies that treated and control units must share common support in their covariate distributions. Without overlap—for instance, when some covariate values appear in the treatment group but not in the control group—it becomes difficult to make credible comparisons, because estimates rely on extrapolation. Overlap is conceptually distinct from balance, which refers to the similarity in covariate distributions across groups. In a randomized control trial, overlap holds by design, and balance is achieved in expectation. In that case, balance can be further improved through stratification or post-stratification. In nonexperimental settings, ensuring overlap is a key condition for credible estimation of causal effects.

Overlap is especially important when researchers do not wish to impose strong functional form assumptions on the conditional means of potential outcomes or on the structure of treatment effect heterogeneity. When the number of covariates is small, overlap can be assessed by examining marginal or joint covariate distributions across treatment groups. But this becomes impractical in high-dimensional settings. In such cases, it is more effective to inspect the distribution of estimated propensity scores across treated and control groups. Lack of overlap in covariate distributions implies, and is implied by, a lack of overlap in propensity score distributions.

LaLonde did not explicitly discuss overlap, nor did he assess it beyond reporting covariate means by treatment status. Both the regression and selection models he used rely on correct functional form assumptions, which permit interpolation or extrapolation of treatment effects even when treated and control units differ substantially in their covariates—thus bypassing the need for overlap. Still, LaLonde clearly recognized the potential problem: to improve comparability, he trimmed the comparison groups based on "characteristics [that] are consistent with some of the eligibility criteria used to admit applicants into the NSW program" (LaLonde 1986, p. 611). However, by modern standards, his trimming procedures—such as removing all men working in March 1976 in one subset (CPS-SSA-2) or further excluding unemployed respondents with 1975 incomes above the poverty line in another (CPS-SSA-3)—are ad hoc and certainly do not guarantee overlap on all relevant covariates.

Over the past four decades, researchers have developed more systematic approaches for improving overlap, often using the propensity score. These methods vary depending on whether the goal is simply to ensure overlap or to further improve covariate balance. Improving overlap or balance typically requires dropping some units from the sample. Although this reduces the sample size, by ensuring better balance it may improve precision for estimates of the average treatment effect. In addition, the gain in robustness and reduction in bias often outweigh any loss in

---

[4] If the target parameter is the average treatment effect (ATE), a stronger version of the overlap assumption is needed: the propensity score must lie strictly between zero and one for all units.

precision. In practice, any increase in variance, even from substantial trimming, is typically modest.[5]

Several specific trimming strategies have been proposed. For instance, focusing on overlap for estimation of the ATT, Dehejia and Wahba (1999a) dropped control units with estimated propensity scores below the minimum observed in the treated group. Crump et al. (2009) proposed a more aggressive approach, selecting subsamples that minimize the variance of average treatment effect (ATE) estimates and recommending trimming units with propensity scores outside the $[0.1, 0.9]$ interval. Crump et al. (2006) and Li, Morgan, and Zaslavsky (2018) proposed improving balance through propensity score weighting, introducing "overlap weights" proportional to the product of the propensity score and one minus the propensity score. Another effective strategy—especially when targeting the ATT—is to match each treated unit to a control unit with a similar estimated propensity score. This approach not only ensures overlap but also tends to improve balance across the covariate distributions.

In practice, overlap—like unconfoundedness—is essential for credible estimation. In settings with poor overlap, such as the LaLonde nonexperimental samples, trimming the sample to ensure overlap is often more important than the choice of estimation method.

Estimation Given Unconfoundedness and Overlap

All estimators in LaLonde (1986) are linear in the covariates. Since then, a wide range of more flexible methods have been proposed to estimate average causal effects under unconfoundedness and overlap. These approaches can be broadly categorized into three groups: (1) outcome modeling, including linear regressions; (2) methods that directly adjust for covariate imbalance, including those based on propensity scores; and (3) doubly robust methods.

First, outcome modeling remains the most widely used approach among applied researchers. It typically involves regressing the outcome on the treatment indicator and covariates (usually, the level terms)—what LaLonde called the earnings equation. This method assumes linearity in covariates and constant treatment effects. A modest relaxation of this approach involves estimating two separate linear regressions for the treated and control groups, sometimes referred to as the Oaxaca-Blinder estimator (Kline 2011). More flexible alternatives include semiparametric and nonparametric methods to model the conditional means of the potential outcomes (for example, Heckman, Ichimura, and Todd 1997; Athey, Tibshirani, and Wager 2019).

---

[5] For example, suppose we have a sample with N_{tr} treated units and N_{co} control units. Under homoskedasticity and random assignment, the variance of the difference-in-means estimator is $\sigma^2(1/N_{tr} + 1/N_{co})$. In the LDW-CPS sample, starting with $N_{tr} = 185$ and $N_{co} = 15,922$, dropping 15,737 control units—a 99 percent reduction—raises the standard error by only about 30 percent in the "best-case scenario," which assumes no bias from including the additional controls.

The second group of methods focuses on directly adjusting for covariate imbalance between the treatment and control groups. This includes blocking on covariates (that is, grouping units with similar characteristics and comparing outcomes within groups, and then aggregating), covariate matching (for example, Abadie and Imbens 2006, 2008, 2011, 2016; Diamond and Sekhon 2013; Rubin 2006; Imbens 2015), and weighting methods to achieve covariate balance (for example, Hirano, Imbens, and Ridder 2003; Hainmueller 2012; Zubizarreta et al. 2023; Zubizarreta 2015).

Covariate matching is a nonparametric method that avoids imposing modeling assumptions. However, it suffers from the curse of dimensionality when many covariates are present, making it prone to large biases (Abadie and Imbens 2006). In such settings, adjusting for the estimated propensity score is often more practical. This can be implemented through blocking and matching (for example, Dehejia and Wahba 1999a; Abadie and Imbens 2011) or through "inverse propensity weighting" (IPW). IPW reweights observations by the inverse of their estimated propensity score, creating a pseudo-population in which treatment assignment is uncorrelated to observed covariates. Hirano, Imbens, and Ridder (2003) show that a variant (the Hájek variant) of the IPW estimator can achieve the semiparametric efficiency bound even when the propensity score is estimated nonparametrically. In the LaLonde (1986) setting, the IPW estimator reweights the nonexperimental control group based on estimated propensity scores so that its covariate distribution closely approximates that of the experimental treated group.

Because covariate imbalance is the sole source of bias under unconfoundedness, scholars have developed methods to improve balance either by refining propensity score estimation or by bypassing it entirely. For instance, Imai and Ratkovic (2014) propose estimating a covariate-balancing propensity score using the generalized method of moments, whereas Hainmueller (2012) introduces "entropy balancing" to achieve balance in specified covariate moments directly. Entropy balancing can be viewed as an IPW estimator that implicitly relies on a correctly specified propensity score model, where the link function is logistic and the log-odds are linear in covariates (Zhao and Percival 2017).

However, neither outcome modeling nor matching or weighting methods on their own are currently most favored in the methodological literature. Instead, hybrid approaches combine outcome modeling, such as regression, with techniques that address covariate imbalance, such as propensity score weighting, to leverage the strengths of both. Examples include regression within propensity score blocks (Rosenbaum and Rubin 1983b; Imbens 2015), matching followed by regression adjustment (Rubin 1973; Abadie and Imbens 2011), and methods that integrate weighting with regression (for example, Robins, Rotnitzky, and Zhao 1994; Robins and Rotnitzky 1995). These approaches are motivated by the fact that even if balancing or propensity score methods are consistent and efficient in large samples, combining them with outcome modeling can reduce small-sample bias and improve precision. For instance, in high-dimensional settings, the bias from a matching estimator due to remaining covariate imbalance may dominate

the variance, and regression adjustment that accounts for this imbalance can help reduce the bias (Abadie and Imbens 2011).

Robins and Rotnitzky (2001) introduced the term "double robustness," a key concept for hybrid methods. They show that if either the propensity score or the outcome model is correctly specified, the augmented inverse propensity weighting (AIPW) estimator—which combines propensity score weighting and regression—is consistent. AIPW can be viewed as an outcome model augmented by a correction term: an IPW estimator applied to the residuals from the outcome model, rather than the raw outcomes. The double robustness property arises because, if the outcome model is correctly specified, the correction term has mean zero even if the propensity score model is misspecified; if the outcome model is misspecified, a correctly specified IPW component can debias the regression model. Beyond robustness, AIPW is also desirable for its efficiency: when both models are correctly specified, it achieves the semiparametric efficiency bound (Bang and Robins 2005).

More recently, machine learning methods have become increasingly popular in applied causal inference (Van der Laan and Rose 2011; Wager and Athey 2017; Chernozhukov et al. 2017; Athey, Imbens, and Wager 2018; Athey, Tibshirani, and Wager 2019); see Athey and Imbens (2017, 2019) for reviews. These methods are particularly useful for estimating "nuisance parameters"—such as propensity scores or conditional outcome means—that are not of direct interest but are essential for identifying causal effects. In the LaLonde setting, the goal is to estimate the average treatment effect on the treated (ATT) on post-training wages, which requires modeling the propensity score and/or the conditional means of potential outcomes. Many modern estimators satisfy the "Neyman orthogonality" condition (Chernozhukov et al. 2017), which reduces the effect of small estimation errors in the nuisance parameters on estimates of the target parameter. In binary treatment settings like LaLonde's, these estimators resemble the AIPW estimator introduced earlier, but with the nuisance parameters estimated via flexible machine learning methods instead of parametric models. Chernozhukov et al. (2017, 2018) emphasize that this property ensures valid inference even when using machine learning algorithms that converge more slowly than is required for estimators based on only estimating the conditional outcome distributions or the propensity score.

Alternative Estimands and Heterogeneous Treatment Effects

Much of the methodological and applied research has focused on estimating average causal effects, such as the average treatment effect on the treated (ATT). However, other quantities may also be of interest. For example, researchers often seek to understand treatment effect heterogeneity among treated units, which can shed light on mechanisms, improve evaluations of effectiveness, and guide personalized policy design. For example, the MDRC (1980) team reported that the National Supported Work program had a large and positive effect on female participants, a significant impact on ex-addict male participants, a small and highly variable impact on ex-criminal male participants, and almost no effect on youth participants.

Econometrically, exploring treatment effect heterogeneity given observed characteristics involves estimating the "conditional average treatment effects on the treated" (CATT). Machine learning methods have been proposed to estimate CATT either nonparametrically or using low-dimensional representations, such as causal forests, while still permitting valid inference or error bounds (for example, Athey and Imbens 2016; Wager and Athey 2018; Athey, Tibshirani, and Wager 2019).

Another important, though less frequently used, estimand is the quantile treatment effects—defined as the difference between quantiles of the treated and untreated potential outcome distributions, either for the population or for the treated group. Under unconfoundedness and overlap, the full marginal distributions of potential outcomes are identified, enabling identification of quantile treatment effects. Firpo (2007) proposes a semiparametrically efficient inverse propensity weighting estimator for these quantities.

Validation through Placebo Analyses

While researchers can assess overlap using observed data, the unconfoundedness assumption is not directly testable. To evaluate the credibility of treatment effect estimates, the literature has developed two main strategies: placebo analyses and sensitivity analyses. Here, we focus on the former and relegate discussion of the latter to the Supplemental Appendix. Placebo analyses offer an indirect way to probe the plausibility of unconfoundedness. These analyses typically involve estimating a model similar to the main specification, but replacing the outcome variable with a pseudo-outcome—usually a variable known to be unaffected by the treatment. A common approach is to test for a treatment effect on a pretreatment variable, such as a lagged outcome, which should not be influenced by the treatment but may still correlate with unobserved confounders. Another variant tests the effect of a pseudo-treatment—often a proxy for the treatment—on the actual outcome. This strategy is often implemented using multiple control groups. For further discussion, see Rosenbaum (1987), Imbens and Rubin (2015), and Imbens (2015).

Although the term "placebo test" was not yet in use in the mid-1980s, LaLonde conducted such an analysis. He regressed 1975 earnings, a pretreatment variable, on the treatment indicator and covariates (reported in columns 2 and 3 of his Tables 4 and 5). Using nonexperimental data, he found that many of the estimated effects were large, negative, and statistically significant, suggesting a violation of unconfoundedness. One limitation of the LaLonde data is that they contain only a single pretreatment outcome. In contrast, the LaLonde-Dehejia-Wahba subset of the LaLonde data allows for placebo analyses that condition on an earlier pretreatment variable, 1974 earnings, when testing for associations between treatment and 1975 earnings. In general, when multiple pretreatment periods are available, placebo tests can be both statistically more powerful and substantively more credible.

Formally, a placebo test assesses a conditional independence restriction, which implies that the placebo outcomes are unrelated to treatment assignment once we account for the remaining observed covariates. A limitation of LaLonde's original test is that it checks only one implication of the full conditional independence

assumption: whether the average outcomes, after adjusting for observed covariates, are the same across groups. Imbens (2015) discusses other testable implications of this assumption.

Reanalyzing the LaLonde Data

To demonstrate the methodological advances since LaLonde (1986) in practice, we revisit the LaLonde data, including the LaLonde-Dehejia-Wahba data, the original LaLonde male samples, and the Lalonde-Calónico-Smith female samples. Our primary focus is on the LaLonde-Dehejia-Wahba data, while results for the other two datasets are reported in the Supplemental Appendix. For all three datasets, overlap is a central concern. The average treatment effect on the treated (ATT) and conditional average treatment effects on the treated (CATT) are our primary causal estimands.

Propensity Scores and Overlap

We focus on the LaLonde-Dehejia-Wahba (LDW) data because they include earnings and employment information from 1974. Our analysis uses three LDW datasets: (1) LDW-Experimental, consisting of 185 treated and 280 control individuals from the experimental sample; (2) LDW-CPS, which includes the same treated individuals and 15,922 controls from CPS-SSA-1; and (3) LDW-PSID, comprising the same treated individuals and 2,490 controls from PSID-1. As noted earlier, LaLonde constructed smaller subsamples to improve comparability, but we instead rely on more modern, data-driven methods to assess and address overlap issues.

First, we estimate propensity scores using generalized random forest (Athey et al. 2019), a machine learning method that flexibly models conditional probabilities. Unlike simple logistic regression, generalized random forest can capture complex nonlinearities and higher-order interactions in the covariates, potentially yielding more accurate estimates of the propensity scores.

Panels A–C of Figure 1 use these estimated propensity scores to assess overlap between treated and control units in each sample. Each panel plots histograms of the log-odds of the estimated propensity scores, defined as $\log(\hat{e}/(1-\hat{e}))$, where $\hat{e}$ is the estimated propensity score. We use the log-odds scale because it more effectively distinguishes differences at the tails of the distribution. For reference, a log-odds of -3 corresponds roughly to a probability of 0.05, and by symmetry, a log-odds of 3 corresponds to a probability of about 0.95.

In panel A, the LDW-Experimental sample shows near-perfect overlap: the treated and control groups have closely aligned propensity score distributions. In contrast, panel B and C reveal severe overlap problems in the nonexperimental samples, with many treated units having propensity scores that fall outside the support of the control group, and large segments of the control group exhibiting extremely low log-odds. Similar patterns are observed in the original LaLonde male samples (as shown in the Supplemental Appendix).

Figure 1
Assessing Overlap in the Lalonde-Dehejia-Wahba (LDW) Data
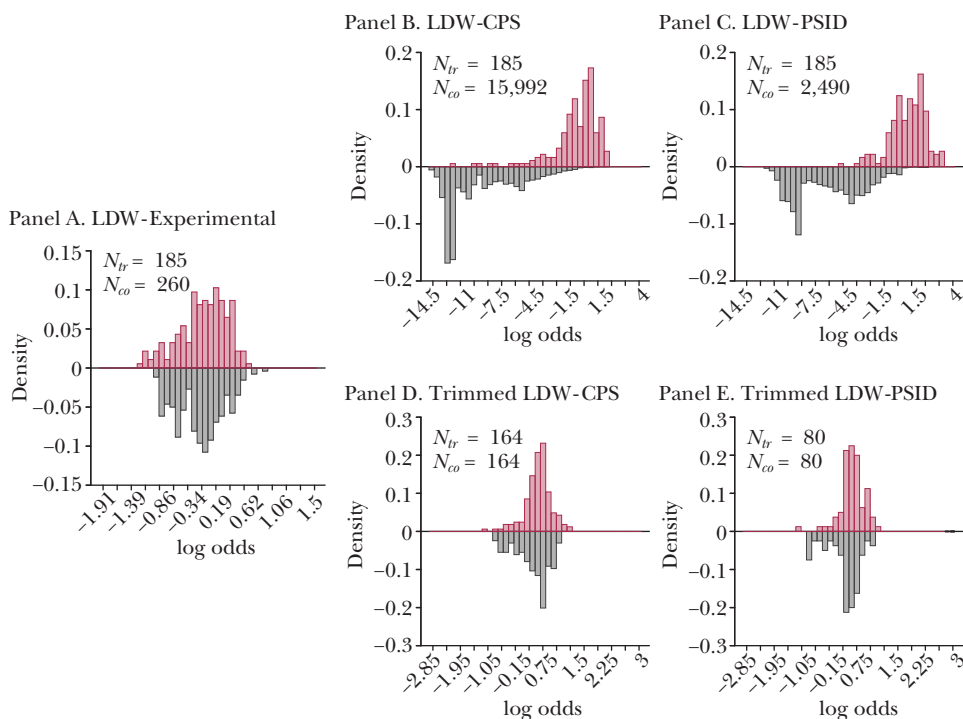
Source: Imbens and Xu (2025).
Note: Histograms show the log odds ratios for units in the treatment group (top, in maroon) and the control group (bottom, in gray). For unit i, the log odds ratio is defined as $\log(\hat{e}/(1-\hat{e}))$, where $\hat{e}$ is the estimated propensity score from the generalized random forest. Each figure represents a different sample. N_{tr} and N_{co} represent the numbers of treated and control units, respectively. A: LDW-Experimental. B: LDW-CPS. C: LDW-PSID. D: Trimmed LDW-CPS. E: Trimmed LDW-PSID. For panels D and E, the propensity scores are re-estimated after trimming.

To address these issues, we construct trimmed versions of the LDW-CPS and LDW-PSID samples. Following the earlier discussion, trimming can improve robustness with only modest loss of precision. We begin by augmenting each nonexperimental sample with the experimental controls and estimating each unit's probability of being in the experimental data using the generalized random forest technique. We then trim based on preset thresholds, which may exclude some treated units (Crump et al. 2009). Next, we re-estimate the propensity scores in the trimmed sample and perform 1:1 matching to refine the control group. This yields two sets of trimmed samples: one with experimental treated units and matched nonexperimental controls, and a second with treated and control units from the experiment, which serves as a benchmark. This two-step trimming and matching procedure improves overlap while preserving a comparison to an experimental

benchmark.[6] As shown in Figure 1, panels D and E, overlap improves substantially in both trimmed samples, albeit at the cost of smaller sample sizes.

Estimating the Average Treatment Effect on the Treated (ATT)

We now estimate the ATT using both the original LaLonde-Dehejia-Wahba nonexperimental samples and the newly constructed trimmed samples. We apply a range of estimators, some likely familiar to most readers and others perhaps more novel. Our primary goal is to compare these methods; again, full computational details are provided in the Supplemental Appendix.

Figure 2 presents the results. The top row reports experimental benchmarks using both the untrimmed and the trimmed LaLonde-Dehejia-Wahba samples. The remaining rows combine the experimental treated units with a nonexperimental control group and apply the following methods: difference-in-means, simple regression; regression with interactions; generalized random forest (GRF) for outcome modeling; "nearest neighbor matching" with bias correction (matching each treated unit with five control units based on covariates); inverse propensity weighting with GRF-estimated propensity scores; covariate balancing propensity score; entropy balancing; double/debiased machine learning using elastic net (DML-ElasticNet); and augmented inverse propensity weighting via GRF (AIPW-GRF). All estimators use the same ten covariates used in prior analyses.[7]

The left panel of Figure 2 shows ATT estimates and 95 percent confidence intervals from the LDW-CPS sample; the right panel shows results from the LDW-PSID sample. For each method, estimates from the full sample are shown in black, and those from the trimmed data are in red. Solid circles mark point estimates; lines represent 95 percent confidence intervals.

Using the LDW-CPS sample (left panel), all estimators yield positive ATT estimates, though they vary in magnitude. Nearest neighbor matching aligns most closely with the experimental benchmark of \$1,794; covariate balancing propensity score, entropy balancing, and AIPW-GRF also produce estimates near this benchmark. Despite numerical differences, the estimates are not statistically distinguishable from one another. In the trimmed LDW-CPS sample, the estimates show less variation across methods, although confidence intervals widen slightly. All estimates in the trimmed sample center around the experimental benchmark of \$1,911. This stability of estimates after trimming is consistent with the simulation results in Athey et al. (2024).

Using the LDW-PSID sample (right panel), estimates from the full data show more dispersion, ranging from \$4 to \$2,420. AIPW-GRF yields an estimate closest to the experimental benchmark. In the trimmed LDW-PSID sample, the experimental

---

[6] Details of the trimming procedure are provided in the Supplemental Appendix along with other computational details.

[7] The difference-in-means estimates are shown in the Supplemental Appendix but omitted here due to their extreme values in the LDW-CPS and LDW-PSID samples (\$−8,497 and \$−15,204, respectively). In the trimmed samples, they are closer to the other estimates (\$1,483 and \$−1,505).

Figure 2

ATT Estimates Given Unconfoundedness: LaLonde-Dehejia-Wahba (LDW) Samples
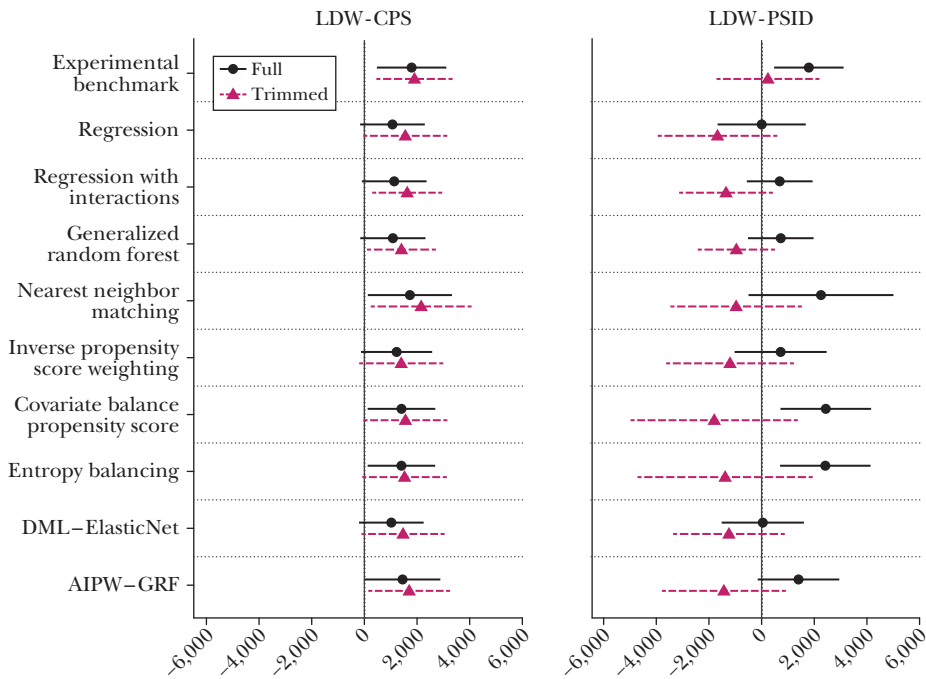
Source: Imbens and Xu (2025).
Note: The panels above show the average treatment effect on the treated (ATT) estimates and their 95 percent confidence intervals using four different samples: LDW-CPS and Trimmed LDW-CPS (left panel), and LDW-PSID and Trimmed LDW-PSID (right panel). Estimates based on corresponding experimental samples are presented at the top. Ten estimators are employed, including difference-in-means, linear regression, linear regression with interactions, generalized random forest for outcome modeling, 1:5 nearest neighbor matching with bias correction, inverse propensity score weighting with GRF-estimated propensity scores, covariate-balance propensity score, entropy balancing, double/debiased machine learning with elastic net (DML-ElasticNet), implemented using DoubleML, and augmented inverse propensity score weighting with GRF for both outcome modeling and propensity score estimation (AIPW-GRF), implemented using generalized random forest techniques.

benchmark is $306, which is not statistically different from zero at the 5 percent level. Although the estimates from the trimmed sample are all negative and more similar to each other, they seem qualitatively different from the experimental benchmark. However, due to large standard errors, these differences are not statistically significant.

Overall, the results suggest that improving overlap based on observed covariates reduces model dependence and estimate variability, thus yielding more robust estimates of the statistical estimand. However, the fact that many methods produce ATT estimates close to the experimental benchmark using LDW-CPS may have given

researchers a false sense of confidence that modern estimators can recover causal effects—even in settings where the unconfoundedness assumption is likely violated.

Treatment Effect Heterogeneity

We explore treatment effect heterogeneity by estimating the conditional average treatment effects on the treated (CATT), comparing estimates from the experimental and nonexperimental samples in the LaLonde-Dehejia-Wahba data. For simplicity, we focus on the LDW-CPS data—both full and trimmed—for the nonexperimental analyses, using the corresponding experimental samples for benchmarking. In the trimmed sample, 21 treated units lacking comparable controls are removed, and the nonexperimental control group is trimmed using one-to-one matching on the estimated propensity score with the remaining 164 treated units. Details on the trimming procedure are discussed earlier in the paper and in the Supplemental Appendix. We estimate CATT using a causal forest estimator (Athey, Tibshirani, and Wager 2017), with technical details presented in the Supplemental Appendix.

Figure 3 plots the estimated CATT at the covariate values of each treated unit, with experimental data on the *x*-axis and nonexperimental data on the *y*-axis, using both the full sample and the trimmed sample. Each dot represents a pair of CATT estimates for an individual with a given covariate profile. The red cross marks the pair of ATT estimates obtained using the AIPW-GRF estimator.

Figure 3 reveals several interesting patterns. First, there is substantial heterogeneity in the estimated treatment effects: CATT estimates based on the experimental data range from \$−236 to \$3,817 in the full sample and from \$−218 to \$4,324 in the trimmed sample. Second, in the full sample, although the AIPW-GRF estimator yields ATT estimates that closely match the experimental benchmark, the CATT estimates from the experimental and nonexperimental data diverge sharply from the 45-degree line. In particular, the nonexperimental CATT estimates span a much wider range (from \$−5,667 to \$7,102) far exceeding that of the experimental estimates, with over one-fourth of treated units obtaining negative CATT estimates. Third, improving overlap substantially enhances the robustness of the CATT estimates: in the trimmed sample, CATT estimates from the experimental and nonexperimental data roughly align along the 45-degree line, and the range of the nonexperimental estimates (from \$−2,877 to \$5,758) is much narrower than in the untrimmed sample.

Applying the same procedure to the LDW-PSID data yields CATT estimates ranging from \$−8,422 to \$4,870 in the full sample and from \$−5,088 to \$1,571 in the trimmed sample—showing greater deviation from the experimental benchmark than in the LDW-CPS case. Analysis of the LaLonde male sample and the reconstructed female data further demonstrates that recovering the CATT is substantially more difficult than estimating the ATT. The combination of the experimental subsample constructed by Dehejia and Wahba (1999a) and the CPS nonexperimental comparison group appears to be an outlier in its ability to recover experimental benchmarks. These additional results are reported in the Supplemental Appendix.
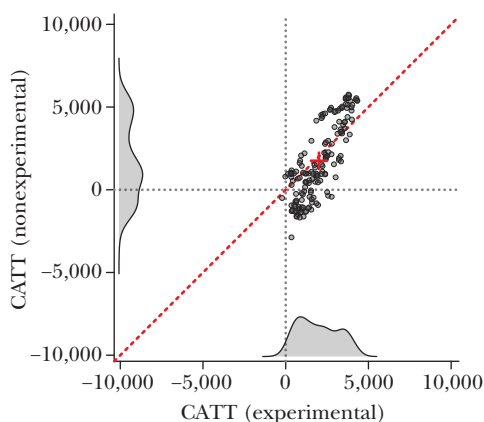
Figure 3
CATT Estimates: Experimental versus Nonexperimental
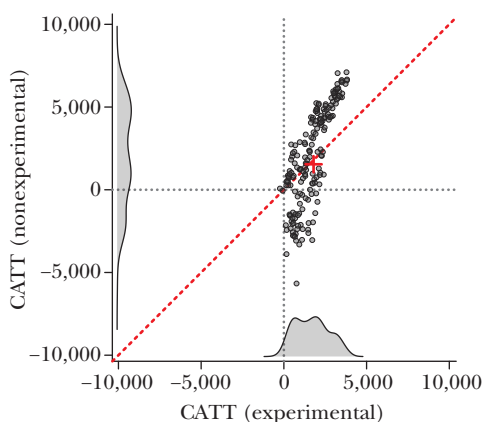
Source: Imbens and Xu (2025).
Note: In the above figures, the scatterplots compare conditional average treatment effects on the treated estimates (CATT) from the experimental data (*x*-axis) and the nonexperimental data (*y*-axis). Panel A compares CATT estimates from LDW-Experimental versus LDW-CPS, while panel B compares those from their trimmed versions, which exclude individuals with extreme estimated propensity scores. In each figure, a dot represents a CATT estimate based on the covariate values of a treated unit; the red cross marks the pair of ATT estimates from the two samples. The marginal distributions of the CATT estimates are shown as density plots along the axes—the bottom plot for the experimental data and the left-side plot for the nonexperimental data. The red dashed line indicates the 45-degree line. Uncertainty estimates are not shown in this figure.

Validation through Placebo Analyses

While modern nonexperimental methods may be effective in estimating the statistical estimand—the covariate-adjusted difference in average outcomes between treated and control groups—this does not imply that the estimate approximates the causal estimand, such as the ATT. Identifying the ATT requires the unconfoundedness assumption, which is fundamentally untestable. However, we can assess its plausibility through placebo analyses.

In the LaLonde setting, we use 1975 earnings as a placebo outcome and exclude both 1975 earnings and employment status from the set of conditioning variables. By construction, 1975 earnings could not have been affected by the treatment, which occurred afterward. We also construct two new trimmed samples, omitting these variables during the trimming process. We then estimate the ATT for the placebo outcome, adjusting for the remaining covariates using a range of estimators.

If the ATT estimates for the placebo outcome are close to zero, this lends support to the unconfoundedness assumption, as it suggests that even without conditioning on 1975 earnings or employment status, treatment assignment is likely independent of the untreated potential outcome $Y_i(0)$—that is, 1978 earnings had

Figure 4
Placebo Tests: 1975 Earnings as the Outcome
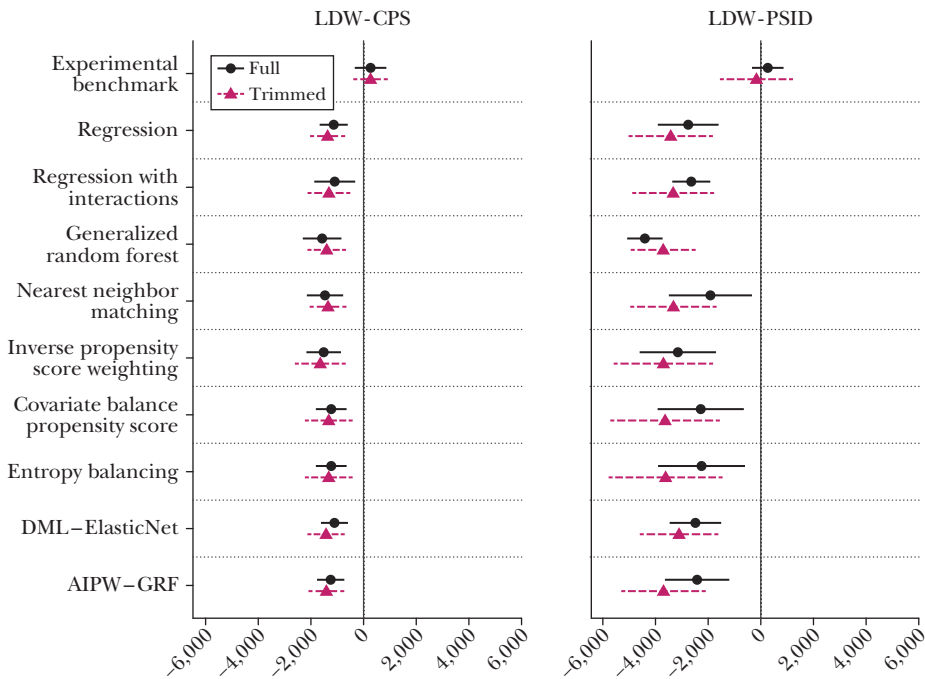
Note: The figures above show the placebo estimates and their 95 percent confidence intervals using four samples: LDW-CPS and Trimmed LDW-CPS (left panel), and LDW-PSID and Trimmed LDW-PSID (right panel). For each case, results based on the corresponding experimental samples are shown at the top as benchmarks. All estimates are produced using the same ten estimators introduced earlier.

the individual not participated in the program. Including these variables in the main analysis would then make unconfoundedness even more credible. However, if the placebo ATT estimates differ significantly from zero, this may indicate either that 1975 earnings and employment status are key confounders, or that unmeasured factors such as perseverance influence both treatment assignment and outcomes. In that case, the placebo analysis fails to bolster unconfoundedness in the main analysis.

Figure 4 presents the results. As expected, the experimental benchmarks are close to zero and statistically insignificant. In contrast, all estimators using nonexperimental data yield large, negative estimates. While trimming improves the stability of these estimates, they remain statistically different from zero.[8] Moreover, as shown

---

[8] The difference-in-means estimates are not shown here; they are $-12,118$, $-17,531$, $-14,056$, and $-4,670$ across the four panels, respectively. Details are available in the Supplemental Appendix.

in the Supplemental Appendix, CATT estimates for a similar placebo test based on experimental data cluster around zero, while their nonexperimental counterparts are consistently negative and sizable, indicating substantial bias. These patterns further illustrate the point we aim to emphasize: With sufficient overlap, modern methods can robustly estimate the statistical estimands, but not necessarily the causal estimands if unconfoundedness is violated.

Alternative Samples

For comparison, we also revisit the original male samples used in LaLonde (1986). These datasets lack information on 1974 earnings and employment status. We find that with sufficient overlap, most modern estimators yield estimates within relatively narrow ranges when using CPS-SSA-1 or PSID-1 as control groups. However, these estimates—most of which are negative—do not align with the experimental ATT benchmarks. Similar patterns are reported by Smith and Todd (2001, 2005). Using these nonexperimental data, modern methods also fail to recover the experimental CATT benchmarks, as noted earlier.

Using the LaLonde-Calónico-Smith female sample with PSID data as nonexperimental controls (Calónico and Smith 2017), we find that many modern methods produce estimates close to the experimental benchmarks, though standard errors are often quite large. As noted by Calónico and Smith (2017), selection appears to be less severe for female participants than for male participants in the National Supported Work program. However, overlap remains a significant challenge. Moreover, a placebo test using the number of children in 1975—a variable not included in LaLonde's original analysis—does not support the unconfoundedness assumption. CATT estimates also do not align with the experimental benchmarks.

As an example of a dataset that readily passes a placebo test, we include in the Supplemental Appendix a reanalysis of data from Imbens, Rubin, and Sacerdote (2001), who conducted an original survey to study the impact of lottery prize size in Massachusetts during the mid-1980s on the economic behavior of lottery players. The primary outcome is post-winning labor earnings, with data available for earnings in the six preceding years. While lottery outcomes may seem random, there are systematic differences in pretreatment variables, likely related to the number of tickets purchased or survey response rates. However, in this case, placebo tests provide strong evidence supporting the unconfoundedness assumption, bolstering the credibility of the causal estimates. The availability of six pretreatment earnings measures proves particularly valuable: they likely capture both selection and outcome-relevant factors and serve as strong placebo outcomes due to their comparability to the post-treatment outcome. We provide the data (Imbens, Rubin, and Sacerdote 2001) in the replication materials.

Summary

Our reexamination of the LaLonde data shows that when overlap is ensured, the choice of estimation method becomes less consequential, as most methods yield similar results. However, these estimates may not represent the causal

quantity of interest if the unconfoundedness assumption is violated. Supplementary analyses, such as placebo tests, can help assess how plausible this assumption is. Although in both the LDW-CPS and LaLonde-Calónico-Smith datasets, many modern methods appear to recover the experimental benchmark for the ATT, placebo tests fail to support unconfoundedness. In typical research settings where experimental benchmarks are unavailable, we cannot know whether such estimates credibly recover the causal estimand. Indeed, with the original LaLonde male sample or the LDW-PSID dataset, modern estimators fail to recover the ATT.

The answer to whether modern nonexperimental methods can credibly estimate causal effects, 40 years after LaLonde's critique, is therefore nuanced. Researchers can now use data-driven approaches to improve overlap and apply modern methods to reliably estimate the statistical estimand. However, credible causal inference still depends on the validity of the unconfoundedness assumption. If supplementary analyses—such as placebo tests—support this assumption, modern methods can yield highly credible estimates. Otherwise, they cannot.

Lessons Learned

What lessons has the methodological literature since LaLonde (1986) taught us? And what specific analyses should researchers conduct today in similar nonexperimental studies? We return to the five lessons mentioned in the introduction and offer practical recommendations.

First, any analysis of causal effects using nonexperimental data should begin with a careful investigation of the treatment assignment mechanism. A clear understanding of the "design" is essential for evaluating the plausibility of the unconfoundedness assumption. The case for relying on unconfoundedness is strongest when researchers believe that selection into treatment is driven by factors that are well understood, observed, and measured. When that condition holds, flexibly adjusting for observed pretreatment covariates can reduce reliance on strong modeling assumptions. In cases where important confounders are unobserved, researchers may turn to panel data methods that account for time-invariant unobserved confounding (for recent overviews, see Xu 2023; Arkhangelsky and Imbens 2024).

Second, the literature has underscored the importance of assessing and improving overlap in covariate distributions. The comparison groups LaLonde used differed substantially from the experimental sample, prompting him to discard some units based on age, employment status, and earnings. Since then, more systematic and data-driven approaches have been developed to diagnose and address lack of overlap, including propensity score-based trimming and weighting strategies. The loss of efficiency is often modest and a worthwhile cost.

Third, relatedly, the propensity score has become central both to diagnosing overlap and to estimating treatment effects. Researchers now routinely estimate the propensity score using flexible methods and evaluate overlap by comparing the

distribution of propensity scores across treated and control groups. When necessary, samples can be trimmed to improve comparability. While the concept of the propensity score was introduced just prior to LaLonde's work in Rosenbaum and Rubin (1983a), it has since become foundational. Doubly robust estimators, particularly those that combine outcome modeling with inverse propensity score weighting, yield consistent estimates when either the outcome model or the propensity score model is correctly specified. The integration of machine learning techniques into causal inference has further strengthened these methods by reducing dependence on ad hoc specification choices. We expect these methods to see broader adoption among economists and social scientists.

Fourth, attention has shifted from solely estimating average treatment effects to understanding treatment effect heterogeneity. Policymakers increasingly ask not only "Does it work?" but "For whom does it work?" and "Where does it cause harm?" The literature has developed tools to estimate conditional average treatment effects and quantile treatment effects. These estimands help unpack heterogeneous impacts and can inform personalized policy decisions. Large datasets and algorithmic advances have made it easier to estimate these effects flexibly and at scale (for example, Wager and Athey 2018). We encourage the estimation and visualization of these additional quantities of interest.

Finally, the credibility of causal estimates increasingly hinges on validation exercises—especially placebo tests. While LaLonde included some placebo analyses using lagged earnings, his main focus was on comparing nonexperimental estimates to experimental benchmarks. In contrast, modern practice places greater emphasis on formal diagnostic checks. Placebo tests, such as estimating effects on outcomes known to be unaffected by the treatment, offer informative checks on key identifying assumptions like unconfoundedness and should be more routinely used in empirical analyses.

To help researchers apply these lessons in practice, we provide a detailed online tutorial—including R code and the data used in this paper and the Supplemental Appendix—available at https://yiqingxu.org/tutorials/lalonde/. The tutorial replicates our analyses and can be easily adapted to other datasets.

References

Abadie, Alberto, and Matias D. Cattaneo. 2018. "Econometric Methods for Program Evaluation." *Annual Review of Economics* 10: 465–503.

Abadie, Alberto, and Guido W. Imbens. 2006. "Large Sample Properties of Matching Estimators for Average Treatment Effects." *Econometrica* 74 (1): 235–67.

Abadie, Alberto, and Guido W. Imbens. 2008. "On the Failure of the Bootstrap for Matching Estimators." *Econometrica* 76 (6): 1537–57.

Abadie, Alberto, and Guido W. Imbens. 2011. "Bias-Corrected Matching Estimators for Average Treatment Effects." *Journal of Business and Economic Statistics* 29 (1): 1–11.

Abadie, Alberto, and Guido W. Imbens. 2016. "Matching on the Estimated Propensity Score." *Econometrica* 84 (2): 781–807.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2010. "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con Out of Econometrics." *Journal of Economic Perspectives* 24 (2): 3–30.

Arkhangelsky, Dmitry, and Guido Imbens. 2024. "Causal Models for Longitudinal and Panel Data: A Survey." *Econometrics Journal* 27 (3): C1–61.

Ashenfelter, Orley. 1978. "Estimating the Effect of Training Programs on Earnings." *Review of Economics and Statistics* 60 (1): 47–57.

Ashenfelter, Orley, and David Card. 1985. "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs." *Review of Economics and Statistics* 67 (4): 648–60.

Athey, Susan, and Guido Imbens. 2016. "Recursive Partitioning for Heterogeneous Causal Effects." *Proceedings of the National Academy of Sciences* 113 (27): 7353–60.

Athey, Susan, and Guido W. Imbens. 2017. "The State of Applied Econometrics: Causality and Policy Evaluation." *Journal of Economic Perspectives* 31 (2): 3–32.

Athey, Susan, and Guido W. Imbens. 2019. "Machine Learning Methods That Economists Should Know About." *Annual Review of Economics* 11: 685–725.

Athey, Susan, Guido W. Imbens, Jonas Metzger, and Evan Munro. 2024. "Using Wasserstein Generative Adversarial Networks for the Design of Monte Carlo Simulations." *Journal of Econometrics* 240 (2): 105076.

Athey, Susan, Guido W. Imbens, and Stefan Wager. 2018. "Approximate Residual Balancing: Debiased Inference of Average Treatment Effects in High Dimensions." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 80 (4): 597–623.

Athey, Susan, Julie Tibshirani, Stefan Wager. 2019. "Generalized Random Forests." *Annals of Statistics* 47 (2):1148–78.

Bang, Heejung, and James M. Robins. 2005. "Doubly Robust Estimation in Missing Data and Causal Inference Models." *Biometrics* 61 (4): 962–73.

Barnow, Burt S., Glen G. Cain, and Arthur S. Goldberger. 1980. "Issues in the Analysis of Selectivity Bias." University of Wisconsin-Madison Institute for Research on Poverty Discussion Paper 600-80.

Calónico, Sebastian, and Jeffrey Smith. 2017. "The Women of the National Supported Work Demonstration." *Journal of Labor Economics* 35 (S1): S65–97.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, and Whitney Newey. 2017. "Double/Debiased/Neyman Machine Learning of Treatment Effects." *American Economic Review* 107 (5): 261–65.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. 2018. "Double/Debiased Machine Learning for Treatment and Structural Parameters." *Econometrics Journal* 21 (1): C1–68.

Chernozhukov, Victor, Christian Hansen, Nathan Kallus, Martin Spindler, and Vasilis Syrgkanis. 2024. "Applied Causal Inference Powered by ML and AI." Preprint, arXiv. https://doi.org/10.48550/arXiv.2403.02467.

Cinelli, Carlos, Andrew Forney, and Judea Pearl. 2022. "A Crash Course in Good and Bad Controls." *Sociological Methods and Research* 53 (3): 1071–1104.

Crump, Richard K., V. Joseph Hotz, Guido W. Imbens, and Oscar A. Mitnik. 2006. "Moving the Goalposts: Addressing Limited Overlap in the Estimation of Average Treatment Effects by Changing the

Estimand." NBER Working Paper 0330.

Crump, Richard K., V. Joseph Hotz, Guido W. Imbens, and Oscar A. Mitnik. 2009. "Dealing with Limited Overlap in Estimation of Average Treatment Effects." *Biometrika* 96 (1): 187–99.

Cunningham, Scott. 2021. *Causal Inference: The Mixtape.* Yale University Press.

Currie, Janet, Henrik Kleven, and Esmée Zwiers. 2020. "Technology and Big Data Are Changing Economics: Mining Text to Track Methods." *AEA Papers and Proceedings* 110: 42–48.

Dehejia, Rajeev H., and Sadek Wahba. 1999a. "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs." *Journal of the American Statistical Association* 94 (448): 1053–62.

Dehejia, Rajeev H., and Sadek Wahba. 1999b. *Reconstructed National Supported Work Demonstration Data.* https://users.nber.org/~rdehejia/data/.nswdata2.html.

Dehejia, Rajeev H., and Sadek Wahba. 2002. "Propensity Score-Matching Methods for Nonexperimental Causal Studies." *Review of Economics and Statistics* 84 (1): 151–61.

Diamond, Alexis, and Jasjeet S. Sekhon. 2013. "Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies." *Review of Economics and Statistics* 95 (3): 932–45.

Ding, Peng. 2024. *A First Course in Causal Inference.* CRC Press.

Firpo, Sergio. 2007. "Efficient Semiparametric Estimation of Quantile Treatment Effects." *Econometrica* 75 (1): 259–76.

Fraker, Thomas, and Rebecca Maynard. 1987. "The Adequacy of Comparison Group Designs for Evaluations of Employment-Related Programs." *Journal of Human Resources* 22 (2): 194–227.

Hainmueller, Jens. 2012. "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies." *Political Analysis* 20 (1): 25–46.

Heckman, James J. 1978. "Dummy Endogenous Variables in a Simultaneous Equation System." *Econometrica* 46 (4): 931–59.

Heckman, James J., and V. Joseph Hotz. 1989. "Choosing among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training." *Journal of the American Statistical Association* 84 (408): 862–74.

Heckman, James J., V. Joseph Hotz, and Marcelo Dabos. 1987. "Do We Need Experimental Data to Evaluate the Impact of Manpower Training on Earnings?" *Evaluation Review* 11 (4): 395–427.

Heckman, James J., Hidehiko Ichimura, and Petra E Todd. 1997. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme." *Review of Economic Studies* 64 (4): 605–54.

Hirano, Keisuke, Guido W. Imbens, and Geert Ridder. 2003. "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score." *Econometrica* 71 (4): 1161–89.

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81 (396): 945–60.

Huber, Martin. 2023. *Causal Analysis: Impact Evaluation and Causal Machine Learning with Applications in R.* MIT Press.

Huntington-Klein, Nick. 2021. *The Effect: An Introduction to Research Design and Causality.* CRC Press.

Imai, Kosuke, and Marc Ratkovic. 2014. "Covariate Balancing Propensity Score." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 76 (1): 243–63.

Imbens, Guido W. 2004. "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review." *Review of Economics and Statistics* 86 (1): 4–29.

Imbens, Guido W. 2015. "Matching Methods in Practice: Three Examples." *Journal of Human Resources* 50 (2): 373–419.

Imbens, Guido W., and Donald B. Rubin. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences.* Cambridge University Press.

Imbens, Guido W., Donald B. Rubin, and Bruce I. Sacerdote. 2001. "Estimating the Effect of Unearned Income on Labor Earnings, Savings, and Consumption: Evidence from a Survey of Lottery Players." *American Economic Review* 91 (4): 778–94.

Imbens, Guido W., and Jeffrey M. Wooldridge. 2009. "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature* 47 (1): 5–86.

Imbens, Guido W., and Yiqing Xu. 2025. *Data and Code for: "Comparing Experimental and Nonexperimental Methods: What Lessons Have We Learned Four Decades after LaLonde (1986)?"* Nashville, TN: American Economic Association; distributed by Inter-university Consortium for Political and Social Research, Ann Arbor, MI. https://doi.org/10.3886/E232201V1.

Kline, Patrick. 2011. "Oaxaca-Blinder as a Reweighting Estimator." *American Economic Review* 101 (3): 532–37.

LaLonde, Robert J. 1986. "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." *American Economic Review* 76 (4): 604–20.

Lechner, Michael. 1999. "Earnings and Employment Effects of Continuous Gff-the-Job Training in East Germany after Unification." *Journal of Business and Economic Statistics* 17 (1): 74–90.

Lechner, Michael. 2002. "Program Heterogeneity and Propensity Score Matching: An Application to the Evaluation of Active Labor Market Policies." *Review of Economics and Statistics* 84 (2): 205–20.

Li, Fan, Kari Lock Morgan, and Alan M. Zaslavsky. 2018. "Balancing Covariates via Propensity Score Weighting." *Journal of the American Statistical Association* 113 (521): 390–400.

MDRC (Manpower Development Research Corporation). 1980. *Summary and Findings of the National Subsidized Work Demonstration.* Ballinger Publishing Company.

Robins, James M., and Andrea Rotnitzky. 1995. "Semiparametric Efficiency in Multivariate Regression Models with Missing Data." *Journal of the American Statistical Association* 90 (429): 122–29.

Robins, James M., and Andrea Rotnitzky. 2001. Comments on "Inference for Semiparametric Models: Some Questions and an Answer." *Statistica Sinica* 11 (4): 920–36.

Robins, James M., Andrea Rotnitzky, and Lue Ping Zhao. 1994. "Estimation of Regression Coefficients When Some Regressors Are Not Always Observed." *Journal of the American Statistical Association* 89 (427): 846–66.

Rosenbaum, Paul R. 1984. "The Consequences of Adjustment for a Concomitant Variable That Has Been Affected by the Treatment." *Journal of the Royal Statistical Society Series A: Statistics in Society* 147 (5): 656–66.

Rosenbaum, Paul R. 1987. "The Role of a Second Control Group in an Observational Study." *Statistical Science* 2 (3): 292–306.

Rosenbaum, Paul R., and Donald B. Rubin. 1983a. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70 (1): 41–55.

Rosenbaum, Paul R., and Donald B. Rubin. 1983b. "Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome." *Journal of the Royal Statistical Society Series B: Methodological* 45 (2): 212–18.

Rubin, Donald B. 1973. "The Use of Matched Sampling and Regression Adjustment to Remove Bias in Observational Studies." *Biometrics* 29 (1): 185–203.

Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66 (5): 688–701.

Rubin, Donald B. 1978. "Bayesian Inference for Causal Effects: The Role of Randomization." *Annals of Statistics* 6 (1): 34–58.

Rubin, Donald B. 2006. *Matched Sampling for Causal Effects.* Cambridge University Press.

Smith, Jeffrey A., and Petra E. Todd. 2001. "Reconciling Conflicting Evidence on the Performance of Propensity-Score Matching Methods." *American Economic Review* 91 (2): 112–18.

Smith, Jeffrey A., and Petra E. Todd. 2005. "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?" *Journal of Econometrics* 125 (1–2): 305–53.

Spława-Neyman, Jerzy. 1990. "On the Application of Probability Theory to Agricultural Experiments. Essay On Principles. Section 9." *Statistical Science* 5 (4): 465–72.

Van der Laan, Mark J., and Sherri Rose. 2011. *Targeted Learning: Causal Inference for Observational and Experimental Data.* Springer.

Wager, Stefan. 2024. "Causal Inference: A Statistical Learning Approach." Unpublished.

Wager, Stefan, and Susan Athey. 2018. "Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests." *Journal of the American Statistical Association* 113 (523): 1228–42.

Xu, Yiqing. 2023. "Causal Inference with Time-Series Cross-Sectional Data: A Reflection." In *Oxford Handbook of Engaged Methodological Pluralism*, edited by Janet M. Box-Steffensmeier, Dino P. Christenson, and Valeria Sinclair-Chapman. Oxford University Press.

Zhao, Qingyuan, and Daniel Percival. 2017. "Entropy Balancing Is Doubly Robust." *Journal of Causal Inference* 5 (1): 20160010.

Zubizarreta, José R. 2015. "Stable Weights That Balance Covariates for Estimation with Incomplete Outcome Data." *Journal of the American Statistical Association* 110 (511): 910–22.

Zubizarreta, José R., Elizabeth A. Stuart, Dylan S. Small, and Paul R. Rosenbaum. 2023. *Handbook of Matching and Weighting Adjustments for Causal Inference.* CRC Press.

Why Regulate Junk Fees?

Neale Mahoney

StubHub is an online marketplace for buying and selling tickets. For years, event ticket sellers like StubHub, Ticketmaster, and others initially displayed a base price and then added service and other fees at check-out—often after consumers had selected seats and entered credit card information. The fees can add a significant amount to the base price. A US Government Accountability Office (2018) report found that these fees averaged 27 and 31 percent of the ticket's total price across the primary and secondary markets, respectively. A shopping analysis by the Council of Economic Advisers (2023) documented fees of $34 to $37 for tickets with a $87 base price.

In response to consumer complaints about these additional fees, in January 2014, StubHub switched to an "all-in, upfront" pricing model, in which consumers were shown a single price, inclusive of fees, at the start of the shopping process. The goal was to improve its brand and shift industry pricing practices. In August 2015, StubHub carried out a field experiment to better understand the dynamics of the pricing change. Specifically, half of US customers were randomly selected to see ticket prices with all fees listed up front, while the others saw a lower price, with fees added at the back end. The results of this experiment, published several

■ *Neale Mahoney is TJ Wijaya Professor of Economics and the Trione Director of the Stanford Institute for Economic Policy Research (SIEPR), both at Stanford University, Stanford, California. He is also a Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. He worked in the White House National Economic Council in 2022 and 2023 and was closely involved in the development of regulations on junk fees. His email address is nmahoney@stanford.edu.*

years later in Blake et al. (2021), showed that consumers who received the lower up-front price and back-end additional fees exhibited greater demand on both the extensive margin (more purchases) and the intensive margin (higher-priced seats). Indeed, demand was higher even among experienced consumers who had visited the website at least ten times and might have expected the back-end fees.

However, the industry did not follow StubHub's switch to all-in, up-front pricing, and StubHub lost significant market share to rivals who continued to offer lower up-front prices and back-end additional fees. In fall 2015, StubHub concluded it had no choice and reverted to a model of showing the lower base price and adding back-end fees at check-out. Since then, company representatives have advocated for a regulatory or legislative solution to what they see as a collective action problem that harms the reputation of their industry (for example, Dooley 2023).

Of course, the business practice in which sellers list a price and then add mandatory fees at check-out extends well beyond online ticket-sellers. Cable television providers may require additional payments for "broadcast television" or "regional sports," which were not advertised in the promotional materials and you only learn about when you get your monthly bill (FCC 2024b). Hotels may add a mandatory "resort fee" that you first notice when reviewing your folio at check-out. Restaurants may add a "kitchen appreciation fee" that you only learn about when you are paying the tab.

These and other examples have become known as "junk fees," defined by Deese, Mahoney, and Wu (2022) as "fees designed either to confuse or deceive consumers or to take advantage of lock-in or other forms of situational market power." For economists, a more useful working definition may be fees designed to exploit consumers' behavioral biases.

In this essay, I will begin by providing concrete examples of junk fees, which include mandatory back-end fees (discussed above) and shrouded add-on fees (described below). I will then turn to what economic theory and evidence teach us about their consequences for market functioning. Junk fees can raise prices by making comparison-shopping more difficult, distort purchase behavior by inducing consumers to make purchases that they would have avoided if they knew the full cost up front, and distort the direction of innovation by encouraging firms to invest in creating new or larger junk fees rather than lowering costs or improving product value.

In the last section, I will discuss policies taken to address junk fees, including a series of rules issued by the Federal Trade Commission (FTC) and the Consumer Financial Protection Bureau (CFPB). (A disclosure is in order here: I worked at the White House National Economic Council in 2022 and 2023 and was closely involved in the development and implementation of this policy agenda.)

Many of these policy changes are recent, and some have been challenged or overturned by courts. In response, policy action has moved to the states: To date, California, Colorado, Massachusetts, Minnesota, Rhode Island, and Virginia have passed sweeping bans on junk fees. Several other states have passed or are

considering various types of junk fee legislation. From the standpoint of economic research, the dynamic policy environment is a boon. It provides useful variation for empirical research and the opportunity to provide timely evidence on how these policies affect consumers and markets while policy is still evolving.

Categories of Junk Fees

Markets work best when consumers can readily observe the full cost of a potential purchase. Pricing practices that make it hard to anticipate the full cost can degrade market functioning. In this section, I will discuss two types of problematic fees: (1) back-end fees that are mandatory but not included in the base price and (2) fees for additional products or services that are shrouded or surprising in magnitude.

Mandatory Back-End Fees

Perhaps the most straightforward example of a junk fee is an add-on fee or surcharge that is mandatory, but *not* included in the base or up-front price, like the back-end service charges in the StubHub example.

Hotel "resort fees" or "facilities fees" that are not included in the base price and may only become salient to consumers at check-out are another example. More than a decade ago, the FTC (2012a) warned hotel operators that their online reservations sites might be violating the law by failing to disclose these fees properly, prompting some changes in practice (Sullivan 2017). A more recent analysis by French (2024) found that among hotels that charged resort fees, the average fee was $38.82 per night.

This category also includes mandatory "broadcast television" or "regional sports" fees charged by cable television providers that were not advertised in the promotional materials and you may only learn about when you get your monthly bill (FCC 2024b).

The dining industry has experienced rapid growth in this type of mandatory back-end fee. Restaurants are increasingly adding "kitchen appreciation fees" and other mandatory surcharges to the tab. Food delivery apps, such as DoorDash and Uber Eats, add a series of mandatory fees for services, delivery, and taxes at check-out that are hard to predict and raise the total price by as much as 30 percent (Council of Economic Advisers 2023).

Nonmandatory but Shrouded Add-On Fees

A second example is "shrouded" add-on fees for additional products or services. This fee is not mandatory—one does not have to buy the add on—but is not clearly and conspicuously disclosed during the initial purchase of the base good. One high-profile example is airline "family seating fees," a term used for the add-on fees that families must pay to guarantee sitting together when buying certain types of plane tickets (Kelleher 2023).

Another example is the surprise out-of-network charges patients face when they receive care at a hospital that is inside their insurance network, but are treated in that facility by an out-of-network physician (Cooper, Scott Morton, and Shekita 2020).[1]

Of course, many add-on fees can be readily justified by the costs of customization. There is nothing wrong with a restaurant charging extra for mushrooms on your pizza or a hotel charging you extra for a room with an ocean view. A related explanation is differences in total costs. For instance, restaurants may charge diners a hefty amount for an after-dinner coffee, not because of the cost of the coffee per se, but because the diners who order coffee may linger longer at their tables after dinner, with a corresponding opportunity cost to the restaurant.

Other add-on fees may be designed for price discrimination. Familiar examples include airlines charging high fees for business-class upgrades, automakers charging large markups on premium features like heated seats, and restaurants charging significant markups on expensive wine. Even in competitive markets, when there are fixed costs, firms need to charge markups above marginal cost to break even. Whether price discrimination is desirable as a normative question needs to be determined on a case-by-case basis.

A telltale feature of add-on fees justified by customization, total costs, or price discrimination is that they do not need to be shrouded to function. Pizzerias do not hide the surcharge for adding mushrooms to your pizza. Flyers would not be surprised to be charged a hefty markup for a business-class upgrade.

The type of fees we are concerned about here are those that only "work" if they are shrouded at the point when customers make their initial purchase decision. Sometimes, these fees remain shrouded until the consumer receives the bill. Out-of-network charges at in-network hospitals or bounced check fees paid by the depositor are good examples. However, even if the add-on price becomes salient when the consumer is offered the add-on, the pricing practice may still be problematic because, at that point in the transaction process, the firm has substantial situational pricing power that allows it to charge a markup well above the competitive rate.

Unfair Pricing Practices

Another motivation for junk fee policies, which is admittedly less comfortable for economists, is the desire to ban "unfair" market practices. Let me provide two examples and then attempt to distill what I think policymakers are getting at.

First, consider the bank practice of processing checking account transactions from largest to smallest within a day to maximize overdraft fees, a practice followed by nearly half of the 44 largest US banks as recently as 2016 (Pew 2016; Carrns

---

[1] These out-of-network fees are technically nonmandatory, as the patient could ask about the physicians network status and refuse treatment, although it seems reasonable to question how much discretion a patient has, especially in an acute health event.

2014). In particular, suppose a consumer has $1,000 in a checking account at the beginning of the day and buys a $3 cup of coffee, $87 sneakers, and has $1,200 automatically deducted for rent. If the bank posts transactions chronologically, the customer is charged a single $35 overdraft fee for the rent payment. However, with high-to-low processing, the account is first charged for $1,200, followed by $87, and lastly $3, resulting in three $35 overdraft fees totaling $105 (Di Maggio, Ma, and Williams 2020).[2]

Second, consider the late-fee model used by Blockbuster Video, back in the days when customers went to those stores to rent a movie in the physical format of a CD or a VHS tape. As readers of a certain age may recall, Blockbuster would charge consumers $2.99 for a five-day video rental, which works out as 60 cents per day. Return the video late, however, and you would incur a $2.99 per day late fee. Of course, late returns incur an opportunity cost, and it was not unreasonable for Blockbuster to charge late fees. But it is hard to imagine that the daily opportunity cost would increase from 60 cents to $2.99 between days 5 and 6.[3]

Bank fees that are often triggered by mistakes are similar to the Blockbuster example. Examples include check account overdraft fees that are assessed even when their accounts showed a sufficient balance at the time of purchase (Consumer Financial Protection Bureau 2022d, 2024a) and bounced check fees assessed on the checks' depositor who cannot control whether the check-writer has a sufficient balance (Consumer Financial Protection Bureau 2022c).

What I think strikes people as unfair about these pricing practices is that they seem designed to trick consumers into making mistakes, and when they do, the fees seem disproportionate to the underlying offense. For the person buying a $3 coffee to know they were incurring a $35 over-limit fee, they would have to remember that their automatic rent payment is deducted the same day and know about high-to-low processing. The Blockbuster pricing model is a textbook example of what some policymakers refer to as "tripwire pricing," a pricing model that seems designed to extract higher payments from those who make a mistake or have the misfortune of a late return.

This general sentiment that the punishment should fit the crime—or in this case, that the fee should fit the fault—is known as "proportionality" in law and philosophy. In US law, regulation targeting junk fees is often brought under prohibitions on unfair and deceptive acts and practices (UDAAP). Both the FTC (1980) and the CFPB (2012) have released publications outlining the legal criteria for what constitutes an unfair pricing scheme under this authority.

---

[2]The Consumer Financial Protection Bureau (CFPB) and the Federal Deposit Insurance Commission (FDIC) have issued guidance cautioning banks against this behavior (Consumer Finanial Protection Bureau 2015; Federal Deposit Insurance Corporation 2023).

[3]When Blockbuster was sued for these practices, data covering a period during which they had partially reformed their practices show that the firm was earning $800 million per year or 16 percent of all revenue from late fees; I do not know of any estimates of what share of Blockbuster revenue came from late fees before that partial reform to its pricing scheme.

Proliferation of Junk Fees across Markets

Junk fees are not new. Hotel resort fees have existed since at least 1997 (Sullivan 2017). A 2003 article documented instances of cell phone operators and rental car companies advertising "low" prices for their offerings and then charging customers additional costs buried in the fine print during check-out (Ayres and Nalebuff 2003). Since at least 2006, ticket agencies have charged mandatory back-end fees for music and theater performances (David 2007; McVeigh 2008).

While data on time trends are scarce, there is a sense that these pricing practices are becoming more prevalent. In their literature review, Greenleaf et al. (2016) write about the "proliferating use" of these pricing practices across a "wide range of markets." Restaurants' use of mandatory surcharges, such as "kitchen appreciation" fees, increased 400 percent between 2018 and 2024, albeit from a low base (as reported by McCorvey 2022; see also Square 2024).

There are several plausible explanations for the growth of these pricing practices. One is the diffusion of revenue-maximizing practices across firms and industries. As one example, the consulting firm IdeaWorks (2025) offers an Ancillary Revenue Master Class that teaches companies about how to apply these techniques in their business. Another is the shift towards e-commerce. The online check-out process, after consumers have entered credit card information and when there are time and psychic costs of abandoning the purchase, may provide a ripe opportunity for the insertion of add-on fees.

Online shopping also enables firms to conduct A/B testing, offering some randomly selected consumers an all-in, up-front price or a lower listed price with back-end fees listed later. Such testing may have allowed online firms to discover the revenue boost from these practices, of which they were not previously aware. While A/B testing can be used by a firm to measure the short-term revenue gain from these pricing practices, it is not as well-suited to quantifying any long-run costs to future sales or the firm's brand. As such, some firms may be overusing these pricing practices relative to what would maximize long-run profits.

Finally, the post-pandemic inflation surge may have accelerated the diffusion of these pricing practices. In press statements and earnings calls, companies pointed to inflation to justify decisions to add these fees.[4] The more rapid diffusion of junk fees during an inflationary period is consistent with a fixed cost of adopting these practices that is only worthwhile when inflation creates a strong incentive for nominal price increases.

Back-End Sales Taxes

A discussion of surcharges would not be complete without a discussion of sales taxes, which are a mandatory add-on fee that is sometimes not included

---

[4]Companies across industries, such as Harley-Davidson (motorcycles), Peloton (stationary bikes and treadmills), Disney World (amusement park), have cited rising costs as a reason for the addition of surcharges (Harley-Davidson 2021; Thomas 2022; Gasparro and Rubin 2022). Some restaurants explained that they added fees to account for rising ingredient costs (Hay 2023; Smith 2023).

in the posted price. Even for a public economist, the rules on tax disclosure are bewildering. Consider a trip to the gas station. The posted price for gasoline includes all taxes. Buy a cup of coffee in the convenience store, however, and the sales tax will not be included in the posted price. Grab a six-pack of beer, and the posted price will include the excise tax but *not* the sales tax, which will be added at the register. The public economics literature includes several influential papers documenting consumers' nonresponsiveness to taxes that are hidden at check-out and the implications of this nonsalience for tax design (Chetty, Looney, and Kroft 2009; Finkelstein 2009).

Stepping outside of this research, my view is back-end taxes represent a short-sighted response to political economy pressures that is often unwise in the long run. As Chetty, Looney, and Kroft (2009) shows, the less salient a tax, the less the burden or incidence will fall on producers. Given this, it is not surprising that a policymaker who is looking to overcome industry opposition to tax may allow the tax to be included at the back end. For these same reasons, I believe policymakers should think twice before implementing back-end taxes. Politicians like policies to counter junk fees because it shows they are fighting for consumers. The same logic indicates that back-end taxes, which can leave consumers feeling nickel-and-dimed at check-out, are self-defeating and should be avoided.

Consequences of Junk Fees

Junk fees can have deleterious consequences for market functioning. By obscuring the full price, junk fees increase the cost of product comparison and thereby raise equilibrium prices in standard search models. Conditional on equilibrium prices, they can also distort purchase behavior, inducing consumers to make purchases that they would not have made if they knew the full cost upfront. Moreover, the ability to engage in these practices can distort the direction of innovation, encouraging firms to invest in creating new or larger junk fees rather than lowering the cost of improving the value of their offerings.

Partitioned Pricing and Drip Pricing

In the academic literature, the terms "partitioned pricing" and "drip pricing" are often used to refer to junk fee pricing practices. "Partitioned pricing" was coined by Morwitz, Greenleaf, and Johnson (1998) to describe a practice of splitting the total price into two or more mandatory pieces, such as a base price and a required surcharge. In their seminal paper, the authors hypothesized that partitioned pricing would lead consumers to underestimate the total price of the product because of behavioral biases: specifically, either because they anchored on the base price (and downweighed the surcharge) or because they ignored the surcharge completely. Consistent with their theory, they found in laboratory experiments that partitioned pricing raised willingness to pay and likelihood of purchase relative to all-in pricing.

Subsequent research, reviewed by Greenleaf et al. (2016), has replicated these findings across contexts and research methodologies. For example, researchers have shown that the distortions can be larger when surcharges are hidden in the fine print or displayed as a percentage of the base price rather than as a dollar amount, and that the distortion is decreasing in the size of the surcharge (for example, Kim 2006; Sheng, Bao, and Pan 2007; Xia and Monroe 2004).

"Drip pricing" is used to describe the pricing practice where mandatory surcharges "drip" in as the consumer proceeds through the purchase process. The mechanism underlying drip pricing has been examined in experimental settings. In a series of laboratory experiments using subjects from Amazon MTurk and business students, Santana, Dallas, and Morwitz (2020) showed that consumers who are exposed to the full dripped price at the end of the shopping process are unlikely to comparison shop. This is due to both higher perceived search costs and the expectation that competitors will employ similar pricing practices. They also showed that consumers engage in self-justification, exhibiting more criticism of others' decisions not to comparison-shop than their own.

Search Costs and Equilibrium Outcomes

While the research literature has mainly studied partitioned pricing and drip pricing practices at the firm level, standard search cost models indicate these types of pricing practices raise equilibrium prices in the market (Stigler 1961; Ellison and Wolitzky 2012). Drip prices increase the cost of searching across products in the marketplace—for example, comparing the total cost for food delivery across delivery apps would entail entering the full order on each app and clicking through to the check-out page; comparing prices for event tickets would require selecting seats and entering credit card information on each website. Across most search models, these search frictions will cause high prices in equilibrium. Partitioned pricing, by making it harder to discover the full price of products in the marketplace, would similarly raise equilibrium prices.

Moreover, these types of pricing practices may also redistribute profit across firms. As the StubHub example makes clear, firms that decide not to use these types of back-end fees will seem more expensive and lose market share and profits to their rivals (Federal Trade Commission 2024a).

Innovation Effects

The prospect of earning revenue from junk fees can distort the direction of innovation. Heidhues, Kőszegi, and Murooka (2016) introduce a model in which firms can either invest in value-increasing innovations or in exploitative innovations such as hidden back-end fees. They show that if higher hidden fees are not offset by a lower base price, firms may have larger incentives to engage in exploitative than value increasing innovation. The formal model captures a broader intuition. If regulations allow firms to engage in these pricing practices, then firms may devote more resources to coming up with new or larger junk fees (or lobbying for approval to charge these fees), rather than to improving the quality or lowering the cost of their offerings.

Options for Policymakers

In this section, I consider the need for and efficacy of policy options to address junk fees. Will a competitive market protect against the distortionary effects of junk fees? Can requirements for transparency overcome these effects? Can explicit regulations, like those that require "all-in" pricing, have consumer benefits that exceed unintended consequences?

Can the Market Fix the Market?

In principle, market forces might constrain or even eliminate junk fee pricing practices. Competition might force companies to dissipate any excess profits in the form of a lower base price. In competitive markets, companies might have an incentive to advertise that they have all-in, up-front pricing and win the market share of consumers who value this feature of the shopping experience. While these forces certainly operate in many settings, a growing literature shows that competitive forces are not generally sufficient to eliminate the harm from junk fees.

The "Chicago School" argument that competition erodes profits from junk fees is described in Ellison (2005): Suppose that products have a base price and an add-on price that consumers only learn after purchasing the product. If there is strong competition on the base price, then even though firms can earn an excess markup on the add-on, the firms will dissipate their profits via a lower base price. In a sense, the base price operates like a "loss leader" that companies offset with the marked-up add-ons.

Over the last 20 years, a literature has identified significant limitations to this argument. Ellison (2005) specifies a Hotelling-type model in which high-type consumers have stronger preferences for the add-on and are less likely to switch firms—for instance, because the add-on is a luxury good that appeals to high-income customers or because it is a late or overdraft fee that is disproportionately incurred by unsophisticated consumers who are less likely to bargain shop. Ellison shows that, in this scenario, obscuring add-on prices raises total prices and firm profits. Firms do not dissipate the higher add-on prices with a lower base price because the marginal consumers attracted by a lower base price are less likely to buy the add-on and therefore less attractive to the firm. Regulation that caps add-on fees can increase consumer welfare, raising the surplus of high-type consumers via lower add-on prices and the surplus of lower-type consumers by providing them a lower-cost opportunity to buy an add-on that they value.

Given that customers are often annoyed by add-on fees, why do firms not advertise that, unlike their competitors, they do not charge add-ons, winning over customers who want a more transparent experience?

In an influential paper, Gabaix and Laibson (2006) specify a model of add-on pricing in which there are sophisticated consumers who anticipate the high-priced add-on and take costly preemptive action to substitute away from it and naïve customers who do not. They show that this model exhibits a "curse of debiasing." In a market with add-on fees, naïve customers cross-subsidize sophisticated consumers

via their purchase of the high-priced add-on. However, a transparent firm cannot poach naive consumers by informing them about add-on fees. Once the naïve consumers are informed, they prefer to remain at the price-shrouding firm, take preemptive action, and get cross-subsidized by the remaining naifs. Relative to a setting without shroud prices, the equilibrium has lower total welfare and a transfer of surplus from naïve to sophisticated customers.

Is Greater Transparency the Answer?

What about requiring transparency of add-on fees through disclosure requirements? US law generally requires accurate disclosure of prices and other terms in consumer product markets. For instance, the Truth In Lending Act (TILA) is intended to aid comparison shopping in credit markets by establishing uniform disclosure requirements, such as the Annual Percentage Rate (APR) (Consumer Financial Protection Bureau 2015), among other requirements. Deceptive practices, defined as "a material representation, omission or practice that is likely to mislead a consumer acting reasonably in the circumstances," are prohibited under Section 5 of the FTC Act on Unfair and Deceptive Acts and Practices (UDAAPs) (FTC 2021). Many other federal agencies and all 50 states have similar UDAAP laws on the books (Carter 2009; Bettino et al. 2024).

Due to these and other laws, many products are accompanied by voluminous disclosures. My Chase checking account includes a 26-page fee schedule, a two-page rate sheet, and a 29-page deposit account agreement with basic terms and conditions. The Fidelity 500 Index Fund, a standard fund designed to track the S&P 500, provides investors with a 32-page prospectus breaking down detailed fees and expenses, investment policies, and financial highlights, along with a somewhat more manageable six-page summary (Fidelity Investments 2024a, b). A standard mortgage closing process includes dozens of pages of disclosures. While requirements that firms provide accurate information are necessary, it seems unreasonable to expect busy people to read and comprehend the disclosures for the products they are considering to a sufficient extent.

Could disclosure be made more effective? While there are no resounding success stories, two areas seem promising. The first is to provide curated information to consumers at the point of purchase ("the information you need when you need it"). For instance, credit card providers are required to provide loan terms in a standardized format known as a "Schumer Box" (Young 2024), which is a table with two panels: the top panel shows interest rates on unpaid balances, cash advances, balance transfers, and the like, while the bottom panel shows annual fees, transaction fees, and penalty fees. Similarly, the Credit Card Accountability Responsibility and Disclosure Act of 2009 (CARD Act), requires that credit card statements prominently display the cumulative interest charges from paying the minimum versus an amount that would pay off the balance in three years. This disclosure caused a sharp increase in consumers paying the three-year amount, but not more than a modest decrease in overall interest payments (Agarwal et al. 2015). In another product area, a recent rule from the Federal Communications Commission (FCC)

requires that broadband providers offer information on costs and performance in an easy-to-read, standardized "nutrition labels" type format.

A second potentially promising area is so-called "smart disclosure" (Thaler 2015). Firms would be required to provide information on costs and terms in a standardized, machine-readable format, with the idea that third parties might develop tools that help consumers shop for and use products in a more optimal way. While there is suggestive evidence that search tools or other decision aids could be helpful (Häubl and Triffs 2000; Murray and Häubl 2008), it is unclear whether smart disclosure laws will cause the development of third-party services that would meaningfully enhance consumer surplus. As the literature on financial advisors warns, there is no guarantee that third parties will operate in consumers' best interest (in this journal, Egan, Matvos, and Seru 2024). Third parties are often profit maximizing, and even third parties that provide a valuable service to consumers can end up extracting more from the market than the surplus they generate (Edelman and Wright 2015).

Given the underwhelming track record of disclosure laws to date, it is worth considering bolder policies. Here is one way to think about it: Consumers' beliefs can be better aligned with product terms either by shifting consumers' beliefs towards product terms via disclosure rules or by shifting product terms towards consumers' beliefs via product regulation. When the evidence shows that it is hard to shift consumer beliefs (disclosure rules are ineffective), it is logical to consider whether it would be more effective to shift product terms closer to what consumers expect (product regulation). While such an approach will not yield the first best of fully informed consumers and completely unconstrained firms, it may yield higher welfare than the status quo.

Regulations to Require All-In, Upfront Pricing

Economic principles provide a strong argument in favor of all-in, up-front pricing requirements for mandatory back-end fees. First, holding equilibrium prices fixed, the ability to observe the full price upfront should reduce the time and effort consumers spend on the shopping process. In the context of online ticket sales, for example, if a consumer spends time selecting a seat and entering credit card information only to abort the transaction when the add-on fees are revealed, that time and effort are wasted.

Second, such a requirement should reduce purchase distortions. Staying on the event ticket example, if a consumer selects a more expensive seat upfront because they were not anticipating a $40 service charge but does not reverse their decision when the cost is revealed because of the time or psychic cost of restarting the purchase process (Santana, Dallas, and Morwitz 2020), the purchase decision is distorted.

Third, such a requirement should reduce market-wide prices in equilibrium. If full prices for tickets are more easily observed, it will be easier to comparison shop, which will increase competition and reduce prices in the market.

Motivated by these arguments, several federal agencies recently took steps to require all-in upfront pricing in industries they regulate. In March 2024, the Federal Communications Commission (2024a) finalized a rule requiring all-in pricing of

cable and satellite TV, arguing that providers had obfuscated the full price and made it difficult to comparison shop by failing to include required "broadcast television," "regional sports," and other similar fees in the advertised price in promotional materials. Under the rule, providers are required to list an aggregate all-in price as a prominent single line item on bills and promotional materials. Any breakdown of fees, such as in the case of bundled video programming and phone services or discounts, can still be provided, but only as a supplement to the prominently displayed all-in price. Other amounts beyond the service-related costs, such as taxes, equipment fees, and franchise fees, are excluded from the all-in rule. The rule went into effect in December 2024 (or March 2025 for smaller operators).

In December 2023, the FTC (2023) finalized the CARS (Combating Auto Retail Scams) rule, which required auto dealers to disclose the actual price any consumer can pay to get the car in all advertising or promotional materials (excluding only required government charges). The rule further requires dealers to get consumers' explicit consent for any added charges and prohibits charges for add-ons deemed not to provide a clear benefit to consumers, such as duplicative warranty programs, service contracts for oil changes on an electric vehicle, or software services on vehicles that do not support the software. However, the rule was struck down on procedural grounds by the Fifth Circuit Court of Appeals in January 2025.[5]

In December 2024, the FTC finalized a rule that requires all-in, up-front pricing for firms in the live-event ticketing and short-term lodging industries. In particular, the rule requires sellers clearly and conspicuously disclose the total price inclusive of mandatory fees whenever they offer, display, or advertise any price, and it requires the total price be more prominently displayed than other pricing information. While the proposed rule was sweeping in the industries covered, the final rule was narrowed to cover event tickets, such as StubHub and Ticketmaster, and short-term lodging, such as hotels and Airbnb.[6] The rule went into effect in May 2025.

In conjunction with federal rulemaking, a number of states passed their own junk fees laws, often anticipating or expanding the scope of federal rules. As of September 2025, California, Colorado, Minnesota, Massachusetts, Rhode Island, and Virginia have passed sweeping junk fee laws that extend beyond event tickets and short-term lodging covered federally. Other states, including Illinois, New York, and North Carolina, have introduced bills requiring all-in, up-front pricing for certain businesses.

---

[5] The case is *National Automobile Dealers Association; Texas Automobile Dealers Association vs. Federal Trade Commission*, US Court of Appeals for the Fifth Circuit, Case No. 24-60013, opinion filed January 27, 2025, https://www.ca5.uscourts.gov/opinions/pub/24/24-60013-CV0.pdf.
[6] These rules build on several agreements concerning junk fees that the Biden administration reached with industry players using the bully pulpit. In June 2023, event ticket vendors, including Ticketmaster–Live Nation and SeatGeek, agreed to switch to upfront pricing (Renshaw 2023). The Ticketmaster agreement only applied to venues that Live Nation controls and excludes sales at sports arenas where pricing is determined in part by the stadium owners. Bully pulpit pressure and heightened media scrutiny also prompted short-term rental companies Airbnb and VRBO and apartment rental platforms Zillow and Appartments.com to move towards all-in upfront pricing (Airbnb 2022; Gangitano 2023; Pohle 2022).

There are also efforts to implement at the state level rules that have been blocked by federal courts. For instance, in September 2025, the California Legislature passed a version of the CARS rule (mentioned above) that had been blocked on procedural grounds at the federal level.

The efforts have not been without political opposition. In California, the restaurant industry pushed back strongly against the all-in pricing requirement, arguing that their prices had always been transparent and that the sticker shock from the higher up-front prices would reduce demand and cause job losses in the industry (California Restaurant Association 2024; Saria 2024; Synergy Restaurant Consultants 2024). Putting aside the internal incoherence of arguing that all-in pricing would depress demand in the presence of already transparent pricing, the lobbying effort was successful: Two days before the bill was to go into effect, restaurants received a carve-out from the law (Fantozzi 2024).

An Emerging Research Agenda

The flurry of policy activity around the subject of junk fees, including the staggered implementation of policies at the state level, provides motivation and useful variation for economic research. Scholars can use the empirical variation to estimate the effect of these policies on consumers, firms, and markets, while also building and refining models that help us better understand and interpret these effects.

Positive Analysis of Junk Fee Regulation

A recent but growing literature is undertaking a positive (that is, nonnormative) analysis of the intended and unintended consequences of fee regulation. Conceptually, one natural question to ask is whether regulation of a hidden fee—say, rules limiting bank overdraft fees or add-on fees at hotels or restaurants—will be offset by a higher base price.

To help guide analysis, Agarwal et al. (2014) provide a model where they parameterize the degree of competition in the market and the salience of the hidden fees. In this model, when a fee is fully hidden (that is, the salience parameter is zero), it functions like a negative cost in that it raises profits without affecting demand. Accordingly, a regulation that clamps down such a fee will be passed through to prices in the same manner as a positive shock to costs.

Consistent with the theory, the empirical literature finds that when fees are nonsalient and competition is limited, the pass-through of fee regulation to the base price need not be one-for-one. For instance, Agarwal et al. (2015) studied restrictions on credit card over-limit and late fees in the 2009 CARD Act. Using a research design that compares consumer credit cards, which were targeted by the law, and small business credit cards, which were excluded, they find that the restrictions significantly reduced fee revenue with no offsetting increase in interest rates or reduction in access to credit.

In a sweeping study, Sarin (2019) analyzed three separate consumer protection policies: the CARD Act restrictions of over-limit and late fees discussed above, an opt-in requirement for bank overdraft fees, and a provision of the Wall Street Reform and Consumer Protection Act (the "Dodd-Frank Act") called the Durbin amendment, which imposes a cap on debit card interchange fees (the "swipe fees" that retailers charge to process debit card payments). Like the CARD Act fee restrictions, Sarin finds that banks were unable to offset lost revenue from reduced overdraft fees. In contrast, banks responded to the lower revenue from the swipe fee cap by eliminating free checking accounts and raising maintenance fees for accounts where there was already a charge. Consistent with the theory above, Sarin argues that these heterogeneous results can be jointly rationalized by considering the difference in salience and competition across these markets.

Directions for Research

While economists are building a strong evidentiary foundation, there is still much to be learned about the intended and unintended consequences of policies to address junk fees. I would suggest three broad areas in which research could be especially productive.

First, researchers can extend and deepen our understanding of consumer decision-making in the face of junk fee pricing practices, building on the seminal work of Morwitz et al. (1998) and others. For these questions, researchers may particularly benefit from the use of laboratory or field experiments to precisely control the decision-making environments and uncover underlying behavior mechanisms.

Second, researchers can estimate consumer and firm responses to government policies, make use of natural experiments provided by these policies, along with increasingly available data on prices and purchases.

Third, researchers should seek to quantify the market-level equilibrium effects of these policies. To do so, researchers may choose to specify and estimate structural models of consumer demand and firms' pricing, using the natural experiments from policy implementation to estimate or validate the models.

In promoting junk fee policies, policymakers often reference a desire to protect certain groups of consumers (US Department of Housing and Urban Development 2023; Consumer Financial Protection Bureau 2024b; Newsom 2024). For example, to motivate regulation of bank fees, policymakers cited research showing credit and debit card fees are higher in zip codes with lower incomes and a high percentage of Black residents (Consumer Financial Protection Bureau 2022b). Researchers can help inform policymakers by documenting any heterogeneity in the fee burden and in the benefits of these policies across different population groups.

Conclusion

Markets work best when consumers can readily observe the full price of products available. Pricing practices that make it hard to anticipate the full cost can

degrade market functioning, raising equilibrium prices through their effect on search costs, inducing consumers to make purchases they would not have made if they knew the full cost up front, and potentially distorting the direction of innovation. Junk fee regulation, at its core, seeks to enhance the functioning of markets. While some economists may resist regulatory intervention in markets, well-designed junk fee policies can enhance market functioning and should have broad support.

References

Agarwal, Sumit, Souphala Chomsisengphet, Neale Mahoney, and Johannes Stroebel. 2014. "A Simple Framework for Estimating Consumer Benefits from Regulating Hidden Fees." *Journal of Legal Studies* 43 (S2): S239—52.

Agarwal, Sumit, Souphala Chomsisengphet, Neale Mahoney, and Johannes Stroebel. 2015. "Regulating Consumer Financial Products: Evidence from Credit Cards." *Quarterly Journal of Economics* 130 (1): 111–64.

Airbnb. 2022. "Airbnb Is Introducing Total Price Display and Updating Guest Checkout."Airbnb Newsroom, November 7. https://news.airbnb.com/airbnb-is-introducing-total-price-display-and-updating-guest-checkout/.

American Economic Liberties Project. n.d. "End Junk Fees." https://www.endjunkfees.com/.

Ayres, Ian, and Barry Nalebuff. 2003. "In Praise of Honest Pricing." *MIT Sloan Management Review* 45 (1): 24–28.

Bettino, Diane A., Martin J. Bishop, Greyson K. Van Dyke, Thomas C. Hardy, and Rebecca R. Hanson. 2024. "State UDAP Laws." Bloomberg Law. https://www.bloomberglaw.com/external/document/XEH90GM0000000/banking-overview-state-udap-laws.

Blake, Tom, Sarah Moshary, Kane Sweeney, and Steve Tadelis. 2021. "Price Salience and Product Choice." *Marketing Science* 40 (4): 619–36.

California Restaurant Association. 2024. "The California Restaurant Association Responds to SB 478 Guidance and FAQ Released by California Attorney General's Office." California Restaurant Association Newsroom, May 8. https://www.calrest.org/news/california-restaurant-association-responds-sb-478-guidance-and-faq-released-california-attorney.

Carrns, Ann. 2014. "Customers Can Lose When Banks Shuffle Payments." *New York Times*, April 11.

Carter, Carolyn L. 2009. "A 50-State Report on Unfair and Deceptive Acts and Practices Statutes." Unpublished.

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. "Salience and Taxation: Theory and Evidence." *American Economic Review* 99 (4): 1145–77.

Consumer Financial Protection Bureau. 2012. *Unfair, Deceptive, or Abusive Acts or Practices (UDAAPs) Examination Procedures.* CFPB.

Consumer Financial Protection Bureau. 2015. *Laws and Regulations: Truth in Lending Act.* CFPB.

Consumer Financial Protection Bureau. 2022a. *Credit Card Late Fees.* CFPB.

Consumer Financial Protection Bureau. 2022b. "CFPB Finds Credit Card Companies Charged $12 Billion in Late Fee Penalties." CFPB, March 29. https://www.consumerfinance.gov/about-us/newsroom/cfpb-finds-credit-card-companies-charged-12-billion-in-late-fee-penalties-in-2020/.

Consumer Financial Protection Bureau. 2022c. "CFPB Issues Guidance to Help Banks Avoid Charging Illegal Junk Fees on Deposit Accounts." CFPB, October 26. https://www.consumerfinance.gov/about-us/newsroom/cfpb-issues-guidance-to-help-banks-avoid-charging-illegal-junk-fees-on-deposit-accounts/.

Consumer Financial Protection Bureau. 2022d. "CFPB Orders Regions Bank to Pay $191 Million for Illegal Surprise Overdraft Fees." CFPB, September 28. https://www.consumerfinance.gov/about-us/newsroom/cfpb-orders-regions-bank-pay-191-million-for-illegal-surprise-overdraft-fees/.

Consumer Financial Protection Bureau. 2024a. "CFPB Orders Navy Federal Credit Union to Pay More than $95 Million for Illegal Surprise Overdraft Fees." CFPB, November 7. https://www.consumerfinance.gov/about-us/newsroom/cfpb-orders-navy-federal-credit-union-to-pay-more-than-95-million-for-illegal-surprise-overdraft-fees/.

Consumer Financial Protection Bureau. 2024b. "CFPB Report Highlights Junk Fees Charged by School Lunch Payment Platforms." CFPB, July 25. https://www.consumerfinance.gov/about-us/newsroom/cfpb-report-highlights-junk-fees-charged-by-school-lunch-payment-platforms/.

Consumer Financial Protection Bureau. 2024c. "Closing Disclosure Explainer." https://www.consumerfinance.gov/owning-a-home/closing-disclosure/.

Consumer Reports. 2022. "New York Governor Signs Bill to Protect Event Buyers from Hidden Fees and Other Unfair Practices into Law." Press release, June 30. https://advocacy.consumerreports.org/press_release/new-york-governor-signs-bill-to-protect-event-ticket-buyers-from-hidden-fees-and-other-unfair-practices-into-law/.

Cooper, Zack, Fiona Scott Morton, and Nathan Shekita. 2020. "Surprise! Out-of-Network Billing for Emergency Care in the United States." *Journal of Political Economy* 128 (9): 3626–77.

Council of Economic Advisers. 2023. "How Junk Fees Distort Competition." The White House. https://bidenwhitehouse.archives.gov/cea/written-materials/2023/03/21/how-junk-fees-distort-competition/.

David, Cara Joy. 2007. "How Much Does It Cost to Buy a $110 Theater Ticket?" *New York Times*, February 10. https://www.nytimes.com/2007/02/10/theater/10fees.html.

Deese, Brian, Neale Mahoney, and Tim Wu. 2022. "The President's Initiative on Junk Fees and Related Pricing Practices." Press release, White House, October 26. https://www.presidency.ucsb.edu/documents/white-house-press-release-the-presidents-initiative-junk-fees-and-related-pricing.

Di Maggio, Marco, Angela T. Ma, and Emily Williams. 2020. "In the Red: Overdrafts, Payday Lending and the Underbanked." NBER Working Paper 28242.

Dooley, Laura. 2023. "StubHub Remarks for the White House Convening on the Economic Case for 'Junk' Fee Policies." Speech, March 21, 2023. https://bidenwhitehouse.archives.gov/wp-content/uploads/2023/03/StubHub-Laura-Dooley-remarks.pdf.

Edelman, Benjamin, and Julian Wright. 2015. "Price Coherence and Excessive Intermediation." *Quarterly Journal of Economics* 130 (3): 1283–328.

Egan, Mark, Gregor Matvos, and Amit Seru. 2024. "The Problem of Good Conduct among Financial Advisors." *Journal of Economic Perspectives* 38 (4): 193–210.

Ellison, Glenn. 2005. "A Model of Add-On Pricing." *Quarterly Journal of Economics* 120 (2): 585–637.

Ellison, Glenn, and Alexander Wolitzky. 2012. "A Search Cost Model of Obfuscation." *Rand Journal of Economics* 43 (3): 417–41.

Fantozzi, Joanna. 2024. "Restaurant Surcharges are Officially an Exception to the California Junk Fee Law." *Nation's Restaurant News*, July 1. https://www.nrn.com/news/restaurant-surcharges-are-officially-exception-california-junk-fee-law.

FCC (Federal Communications Commission). 2024a. "Broadband Consumer Labels." https://www.fcc.gov/broadbandlabels.

FCC (Federal Communications Commission). 2024b. "FCC Adopts 'All-In' Cable and Satellite Video Pricing." https://www.fcc.gov/document/fcc-adopts-all-cable-and-satellite-video-pricing.

Federal Deposit Insurance Corporation. 2023. *Supervisory Guidance on Charging Overdraft Fees for Authorize Positive, Settle Negative Transactions.*

Federal Trade Commission. 1980. "FTC Policy Statement on Unfairness." FTC, December 17. https://www.ftc.gov/legal-library/browse/ftc-policy-statement-unfairness.

Federal Trade Commission. 2012a. "FTC Warns Hotel Operators that Price Quotes that Exclude 'Resort Fees' and Other Mandatory Surcharges May be Deceptive." Press release, November 28. https://www.ftc.gov/news-events/news/press-releases/2012/11/ftc-warns-hotel-operators-price-quotes-exclude-resort-fees-other-mandatory-surcharges-may-be.

Federal Trade Commission. 2012b. "The Economics of Drip Pricing." May 21. https://www.ftc.gov/news-events/events/2012/05/economics-drip-pricing.

Federal Trade Commission. 2021. "A Brief Overview of the Federal Trade Commission's Investigative, Law Enforcement, and Rulemaking Authority." FTC, May. https://www.ftc.gov/about-ftc/mission/enforcement-authority.

Federal Trade Commission. 2023. "FTC Announces CARS Rule to Fight Scams in Vehicle Shopping." Press release, December 12. https://www.ftc.gov/news-events/news/press-releases/2023/12/ftc-announces-cars-rule-fight-scams-vehicle-shopping.

Federal Trade Commission. 2024a. "Federal Trade Commission Announces Bipartisan Rule Banning Junk Ticket and Hotel Fees." Press release, December 17. https://www.ftc.gov/news-events/news/press-releases/2024/12/federal-trade-commission-announces-bipartisan-rule-banning-junk-ticket-hotel-fees.

Federal Trade Commission. 2024b. *FTC CARS Rule: Combating Auto Retail Scams—A Dealer Guide.* FTC.

Fidelity Investments. 2024a. "Fidelity 500 Index Fund/FXAIX." Prospectus, April 29. files.agfinancial.org/Retirement/Prospectus/Fidelity-500.pdf.

Fidelity Investments. 2024b. "Fidelity 500 Index Fund/FXAIX. Summary Prospectus." April 29. https://fundresearch.fidelity.com/prospectus/sec.

Finkelstein, Amy. 2009. "E-ztax: Tax Salience and Tax Rates." *Quarterly Journal of Economics* 124 (3): 969–1010.

French, Sally. 2024. "How to Avoid Hotel Resort Fees (and Which Brands are the Worst)." NerdWallet, September 25. https://www.nerdwallet.com/article/travel/hotel-resort-fees.

Gabaix, Xavier, and David Laibson. 2006. "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets." *Quarterly Journal of Economics* 121 (2): 505–40.

Gangitano, Alex. 2023. "Zillow, Apartments.com Pledge to Show All-In Pricing for Prospective Renters: White House." *The Hill*, July 19. https://thehill.com/homenews/administration/4104009-zillow-apartments-com-pledge-to-show-all-in-pricing-for-prospective-renters-white-house.

Gasparro, Annie, and Gabriel T. Rubin. 2022. "The Hidden Ways Companies Raise Prices." *Wall Street Journal*, February 12. https://www.wsj.com/articles/companies-hidden-inflation-consumer-price-index-11644549254.

Greenleaf, Eric A., Eric J. Johnson, Vicki G. Morwitz, and Edith Shalev. 2016. "The Price Does Not Include Additional Taxes, Fees, and Surcharges: A Review of Research on Partitioned Pricing." *Journal of Consumer Psychology* 26 (1): 105–24.

Harley-Davidson, Inc. 2021. "CEO Jochen Zeitz on Q2 2021 Results—Earnings Call Transcript." July 21. https://seekingalpha.com/article/4440427-harley-davidson-inc-hog-ceo-jochen-zeitz-on-q2-2021-results-earnings-call-transcript.

Häubl, Gerald, and Valerie Trifts. 2000. "Consumer Decision Making in Online Shopping Environments." *Marketing Science* 19 (1): 4–21.

Hay, Kiara. 2023. "Some Restaurants Adding 'Inflation Fees' to Bills Amid Rising Food Costs." *WXYZ Detroit*, March 10. https://www.wxyz.com/news/some-restaurants-adding-inflation-fees-to-bills-amid-rising-food-costs.

Heidhues, Paul, Botond Kőszegi, and Takeshi Murooka. 2016. "Exploitative Innovation." *American Economic Journal: Microeconomics* 8 (1): 1–23.

IdeaWorks. 2025. "Ancillary Revenue Assessment." https://ideaworkscompany.com/services/ancillary-assessment/.

Kelleher, Suzanne Rowan. 2023. "These Airlines Charge Families Extra to Sit Together." *Forbes*, March 7. https://www.forbes.com/sites/suzannerowankelleher/2023/03/07/airlines-charge-families-extra-sit-together/.

Kim, Hyeong Min. 2006. "The Effect of Salience on Mental Accounting: How Integration versus Segregation of Payment Influences Purchase Decisions." *Journal of Behavioral Decision Making* 19 (4): 381–91.

McCorvey, J. J. 2022. "Restaurants Add New Fees to Your Check to Counter Inflation." *Wall Street Journal*, June 2. https://www.wsj.com/articles/waiter-theres-a-fee-in-my-soup-11654139870.

McVeigh, Karen. 2008. "Hidden Charges Add 30% to Cost of Seat for Music and Theatre Fans." *The Guardian*, May 29. https://www.theguardian.com/stage/2008/may/29/culture.theatre.

Morwitz, Vicki, Eric Greenleaf, and Eric Johnson. 1998. "Divide and Prosper: Consumers' Reactions to Partitioned Prices." *Journal of Marketing Research* 35 (4): 453–63.

Murray, Kyle B., and Gerald Häubl. 2008. "Interactive Consumer Decision Aids." In *Handbook of Marketing*

Decision Models, edited by Berend Wierenga, 55–77. Springer.

Newsom, Gavin. 2024. "Governor Newsom Signs Consumer Protection Bills Targeting Medical Debt, Overdraft Fees, and Unfair Subscription Practices." Press release, September 24. https://www.gov.ca.gov/2024/09/24/governor-newsom-signs-consumer-protection-bills-targeting-medical-debt-overdraft-fees-and-unfair-subscription-practices/.

Pew. 2016. *Consumers Need Protection From Excessive Overdraft Costs.* The Pew Charitable Trusts.

Pohle, Allison. 2022. "Why Airbnb Cleaning Fees Cost So Much Now." *Wall Street Journal*, April 12. https://www.wsj.com/articles/why-airbnb-cleaning-fees-cost-so-much-now-11649717791.

Renshaw, Jarrett. 2023. "Ticketmaster, Others Agree to Upfront, All-In Prices as Part of Biden War on Junk Fees." *Reuters*, June 15. https://www.reuters.com/world/us/ticketmaster-live-nation-agree-all-in-prices-part-biden-war-junk-fees-2023-06-15/.

Santana, Shelle, Steven K. Dallas, and Vicki G. Morwitz. 2020. "Consumer Reactions to Drip Pricing." *Marketing Science* 39 (1): 188–210.

Saria, Lauren. 2024. "A New Law That Goes into Effect July 1 Could 'Upend' California's Restaurant Industry." *Eater San Francisco*, May 3. https://sf.eater.com/2024/5/3/24148146/california-junk-fee-ban-restaurants-bars.

Sarin, Natasha. 2019. "Making Consumer Finance Work." *Columbia Law Review* 119 (6): 1519–96.

Sheng, Shibin, Yeqing Bao, and Yue Pan. 2007. "Partitioning or Bundling? Perceived Fairness of Surcharge Makes a Difference." *Psychology and Marketing* 24 (12): 1025–41.

Smith, Stacey Vanek. 2023. "What My $30 Hamburger Reveals about Fees and How Companies Use Them to Jack Up Prices." *NPR*, July 27. https://www.npr.org/2023/07/27/1189665392/fees-inflation-white-house-service-hidden.

Square. 2024. "Summer Quarterly Restaurant Report: Service Fees Continue to Grow, Wages Stay Above Inflation." August 13. https://squareup.com/us/en/press/summer-restaurant-report-2024.

Stigler, George J. 1961. "The Economics of Information." *Journal of Political Economy* 69 (3): 213–25.

State of California Department of Justice, Office of the Attorney General. 2025. "SB 478—Hidden Fees." https://oag.ca.gov/hiddenfees.

Sullivan, Mary W. 2017. "Economic Analysis of Hotel Resort Fees." Bureau of Economics, Federal Trade Commission. https://www.ftc.gov/system/files/documents/reports/economic-analysis-hotel-resort-fees/p115503_hotel_resort_fees_economic_issues_paper.pdf.

Synergy Restaurant Consultants. 2024. "SB 478: New Ban on Adding Service Charges at California Restaurants." May 14. https://www.synergyconsultants.com/blog-posts/sb-478-new-ban-on-adding-service-charges-at-california-restaurants.

Thaler, Richard. 2015. *Misbehaving: The Making of Behavioral Economics.* W.W. Norton.

Thomas, Lauren. 2022. "Peloton Is about to Tack on Hundreds of Dollars in Fees to Its Bike and Treadmill, Citing Inflation." *CNBC*, January 16. https://www.cnbc.com/2022/01/17/peloton-is-about-to-tack-on-hundreds-of-dollars-in-fees-to-its-bike-and-treadmill-citing-inflation.html.

US Government Accountability Office. 2018. *Event Ticket Sales: Market Characteristics and Consumer Protection Issues.* GAO-18-347.

US Department of Housing and Urban Development. 2023. "Biden-Harris Administration and Secretary Fudge Take on Junk Fees in Rental Housing." Press release, July 19. https://archives.hud.gov/news/2023/pr23-146.cfm.

Xia, Lan, and Kent B. Monroe. 2004. "Price Partitioning on the Internet." *Journal of Interactive Marketing* 18 (4): 63–73.

Young, Kenley. 2024. "What Is a Schumer Box and How Do You Read It?" NerdWallet, December 19. https://www.nerdwallet.com/article/credit-cards/how-to-read-schumer-box.

Retrospectives: W. E. B. Du Bois, Harvard Economics, and Marginalist Wage Theory

Daniel Kuehn

This feature addresses the history of economic terms and ideas. The hope is to deepen the workaday dialogue of economists, while perhaps also casting new light on ongoing questions. If you have suggestions for future topics or authors, please contact Joseph Persky, University of Illinois at Chicago (jpersky@uic.edu).

W. E. B. Du Bois (1868–1963) was a public intellectual, civil rights leader, and scholar of race and Black life. His work is closely associated with sociology and history, but after receiving his first bachelor's degree at Fisk University in 1888 (he would earn a second at Harvard), Du Bois studied economics at both Harvard University (1888–1892) and the University of Berlin (1892–1893). When he was forced to return from Berlin due to the expiration and nonrenewal of his Slater Fund scholarship, Du Bois completed his PhD at Harvard in history rather than economics (for details on this educational transition, see Lewis 2009). But by the end of the 1890s, he was a professor of history and economics at Atlanta University (1897–1910).

In this essay, I show that as a graduate student studying economics with Frank Taussig at Harvard University in the early 1890s, Du Bois was a well-read and innovative contributor to the contemporary debates on wage theory. In his unpublished 158-page 1891 manuscript, *A Constructive Critique of Wage Theory*, Du Bois adapts Carl Menger's value theory to develop a marginal productivity theory

■ *Daniel Kuehn is a Principal Research Associate, Urban Institute, and Adjunct Professor, George Washington University, both in Washington, DC. His email address is dkuehn@ urban.org.*

of wages that anticipates both the concept of the marginal revenue product of labor and the labor-leisure tradeoff that would be formalized by Wicksteed (1910). The first persuasive account of how marginalism relates to the determination of factor prices is usually credited to John Bates Clark's *The Distribution of Wealth* (1899). Clark's marginal productivity theory was more formally developed than Du Bois's, and it was not dependent on the awkward Mengerian ranked values employed by Du Bois. Nevertheless, by rescuing Du Bois's early work on wage theory from obscurity, and detailing how it built on other theories that also anticipated Clark (1899), we can cultivate a deeper appreciation for the breadth of his scholarship and his place in the history of economics.

In the late 1880s and early 1890s, when Du Bois studied economics at Harvard, the prominent doctrine for the determination of wages remained classical wage theory. Classical wage theory often carried an implication of what Ferdinand Lassalle (1863) derisively called the Iron Law of Wages—that is, wages would be pushed toward subsistence in the long run. One reason wages were supposedly pushed towards subsistence was that they were constrained by the "wages fund." According to this wages fund theory, workers were paid out of the capital available each period after deducting the cost of raw materials and profits, so that average wage levels were determined by the size of the capital stock and the size of the workforce. Assuming Malthusian forces governed the size of the workforce, the wages fund theory went hand in hand with the Iron Law of Wages, and implied that political or labor advocacy for higher wages was self-defeating. When Du Bois studied with Taussig, Taussig was working on the book *Wages and Capital*, which sought to extend wages fund theory, address its shortcomings, and temper many of its harsher policy implications.

Of course, Du Bois did not continue to work on marginalist economics. During his *Wanderjahre* in Berlin in 1892 and 1893, the renowned economists Gustav Schmoller and Adolph Wagner powerfully influenced Du Bois's methodology and his understanding of the social role of social science. Du Bois himself emphasizes the significance of the German Historical School of Economics for his work in his many autobiographical writings.

Instead of pursuing a marginalist theory of wages after Berlin, Du Bois became one of the founders of the field of Black Labor Studies in the late nineteenth and early twentieth centuries (Grossman 1974; Wilson 2006; Rutherford 2024). Black Labor Studies was an empirical effort to understand the condition of Black workers in the United States, conducted almost exclusively by Black social scientists. Wilson (2006, p. 17) describes *The Philadelphia Negro*, published by Du Bois in 1899, as setting "a fifty-year methodological standard" for Black Labor Studies, which he would continue to promote through the American Negro Academy and Atlanta University. Professionally—and to a certain extent intellectually—Black Labor Studies operated parallel to rather than in conjunction with academic economics, due to discrimination against Black scholars and the segregation of social spaces (like universities and conferences) where social science is conducted. This empirical work on the economic and social life of Black Americans, informed by the German Historical

School of Economics, was Du Bois's ultimate legacy in economics. But it was not the sort of economics that Du Bois was practicing as a graduate student at Harvard.

Harvard University, Frank Taussig, and Classical Wage Theory

Understanding Du Bois's economic studies at Harvard generally and his work on wage theory in particular must begin with understanding his relationship with Frank Taussig (1859–1940), his economics professor as an undergraduate and graduate student. In the literature on Du Bois, Taussig is often presented as a towering figure of orthodox economics, but this is a misleading image for understanding his role in Du Bois's career. When Du Bois enrolled at Harvard, Taussig's star was still early in its rise. At the time, Taussig was only 28 years old and had only been an assistant professor for two years. He had not yet started his 41-year tenure as editor of the *Quarterly Journal of Economics*, nor had he embarked on any of the major public service roles for which he would later become famous.

In 1888, the year Du Bois enrolled as an undergraduate, Taussig published his *The Tariff History of the United States*, which was the culmination of several years of work on protectionism and international trade. Taussig then turned his attention to wage theory, which became the centerpiece of his classroom instruction and his own studies for the next several years,[1] culminating in the 1896 publication of *Wages and Capital. Wages and Capital* was a historical and theoretical investigation that attempted to rescue the wages fund theory from the rigidities of classical economics. Taussig (1896) rejected the close relationship between the wages fund theory and the Iron Law of Wages. He saw "the main cause of this unsatisfactory treatment" of the wages fund by Smith, Malthus, and Ricardo as being their undue emphasis on "the standard of living" and "the principle of population" as governing wages, rather than market forces (Taussig 1896, p. 323). Obsession with the Iron Law of Wages was, in Taussig's view, the reason no important advances on the wages fund theory had been made until the second half of the nineteenth century.

To get out of this theoretical rut, Taussig (1896) expands the classical concept of capital to include goods at the intermediate stage of production, appealing to observation of the "machinery of distribution" (that is, the way people are actually paid) and the theoretical advances of Austrian economists, principally the work of Eugen von Böhm-Bawerk.[2] The Austrian economists pointed out that there is a time structure of production, so that workers' productive activity would only generate revenue in the future. Workers had to be paid out of "the stock of wealth already produced" (Taussig 1896, p. 315), which could be considered a type of wages fund. The Austrian concept of the time structure of production, with labor applied to a

---

[1] See Irwin Collier's reproduction of Taussig's final exams for the years he taught Du Bois, at https://www.irwincollier.com/harvard-history-of-economic-theory-final-exam-questions-taussig-1891-94/.

[2] Schumpeter, Cole, and Mason (1941) recall Taussig telling Schumpeter that Böhm-Bawerk was Ricardo's only rival as an economic theorist.

series of unfinished intermediate goods to produce a flow of income, reinforced Taussig's concept of a wages fund that was more "elastic" than a fund restricted by the final product of labor. Taussig felt that Böhm-Bawerk's explanation of wage determination was not as well developed as his theory of interest, but that "on these topics economic theory will gain by following the main trend of the exposition which has finally resulted from the labors of the Austrian school" (Taussig 1896, p. 318).

In his economics classes, Taussig taught Du Bois the classical and contemporary wage theorists. This instruction, along with ongoing debates among economists about factor price determination in the context of the marginal revolution, led Du Bois to write and submit a 158-page manuscript entitled *A Constructive Critique of Wage Theory* as an entry for the 1891 Toppan Prize, a Harvard essay contest. The *Constructive Critique* shows a different side of Du Bois's economic thought, pre-dating his time in Berlin or his more empirically focused work in *The Philadelphia Negro* and the reports he produced at Atlanta University.

A Constructive Critique of Wage Theory (1891)

I am certainly not the first to discuss Du Bois's *Constructive Critique*. However, the usual pattern is for biographers and historians to only offer passing references to the essay, and only through the lens of the Du Bois they know best: Du Bois the Marxist, Du Bois the sociologist, or Du Bois the social reformer.[3] All of these facets of Du Bois were already germinating at Fisk and Harvard, and might therefore be found in the *Constructive Critique*, if we look hard enough or extrapolate. But here, I advocate for reading the *Constructive Critique* through the lens of the W. E. B. Du Bois who wrote it. In 1891, Du Bois was a young, intelligent, curious, aspiring economist who (like his teacher Taussig) was dissatisfied with classical wages fund theory and inspired by what he read in the early issues of the *Quarterly Journal of Economics* in his late nights in Harvard's Political Economy Library.[4] We have to ask

---

[3] For example, Buschendorf (2017) notes that Du Bois discusses Marx in the *Constructive Critique*, which he argues provides evidence of an early familiarity with Marxist economics. While this is true as far as it goes, the *Constructive Critique* only spends two pages on Marx and characterizes Marx as "the great expounder and elaborator" (Du Bois 1891, p. 55) of Johann Karl Rodbertus, rather than a wholly original thinker. Later, Du Bois (1891, p. 118) even more icily notes that it is "now generally conceded by the Socialists themselves" that Marx and Rodbertus "carried the Ricardian principle of Cost of Production too far." Other mentions of the *Constructive Critique* read Du Bois's social and policy goals into the text. Lewis (2009, p. 88) notes that Du Bois (1891) "groped for a formulation of the wages problem that was less Darwinian than the prevailing one, more benign and community oriented," but then dismisses the *Constructive Critique* as not making any contribution to economic theory (a point I hope I will have decisively refuted in this essay). Oliver (2014) makes the vague suggestion that Du Bois (1891) "develops the thesis that society had an obligation to regulate profits as the only mechanism for distributing wealth." None of these brief references provide any discussion of the details of the actual content of the marginalist wage theory Du Bois developed in the *Constructive Critique*, perhaps because none of these authors are invested in understanding Du Bois as an aspiring economist.

[4] See Du Bois's "In the Library," for a short description of one such late night that Du Bois spent in the Harvard Political Economy library. The essay is available through in the W. E. B. Du Bois papers held

the question, what motivated Du Bois to write the *Constructive Critique*? He wrote it because he was trying to resolve a puzzle that was engulfing economics in the late nineteenth and early twentieth centuries: how were wages determined?

Du Bois (1891) begins his *Constructive Critique* with an eclectic comparison of wages in various industries and occupations in Brussels, New York City, Chicago, Ontario, and Massachusetts. He observes that these wages vary widely, motivating "the questions we are about to investigate," namely, "is there any discernable scientific law regulating the amount of return to personal exertion" (Du Bois 1891, p. 7). The first two-thirds of Du Bois's *Constructive Critique* comprises an "exposition of theories" of 19 economists on the determination of wages. John Elliott Cairnes, Alfred Marshall, Francis Amasa Walker, and William Stanley Jevons receive the most detailed individual treatment, although the largest section of the literature review is dedicated to the Austrian School of Economics, principally covering Carl Menger and Eugen Böhm-Bawerk.

Part of the reason why Du Bois's discussion of the Austrian school is so much longer is that it develops Austrian marginalism from first principles. Du Bois does this because his own wage theory would lean most heavily on Carl Menger's value theory, in which Menger argued that the value of a good was not determined by the characteristics of the good itself, but by the ability of the good to satisfy human needs and wants, which in turn can be ranked in a hierarchy of wants. In his discussion, Du Bois introduces the difference between utility and value, the distinctively Mengerian ranking of wants, and the conclusion that value is determined by "Final" or marginal utility. Du Bois then moves on to the implications of Austrian marginalism for factor prices, writing that factor prices are "fixed in the last resort, by the finished product's Final Utility" (Du Bois 1891, p. 98). This formulation inverts classical cost-of-production theories of value that determined the "value" of the final product by adding up the costs of the factors of production. Du Bois further recognizes that this theory was relevant to factor prices, including wages. This extension of marginal analysis into the determination of factor prices had only recently been pioneered by John Bates Clark (1888, 1889, 1890, 1891), von Wieser (1889), and Böhm-Bawerk (1890).

Du Bois (1891) quotes Böhm-Bawerk (1890) at length to explain wage determination, and then summarizes in his own words that "the level of Wages at any given time depends on the quantity of present goods offered for future goods, as compared with the number of laborers offering future goods for present, i.e., offering to be hired" (Du Bois 1891, p. 102). The logic is that workers produce future goods, which they offer to employers in exchange for present goods (that is, their wages). This is closely related to Taussig's (1896) reliance on Böhm-Bawerk for his elastic version of the wages fund theory a few years later. Despite his familiarity with the Austrian economists, Du Bois was uncharacteristically intellectually

---

by the University of Massachusetts-Amherst library at https://credo.library.umass.edu/view/collection/mums312. "In the Library" was written for Du Bois's English 12 class and dated April 30, 1891, which was the day before he finished the *Constructive Critique*, May 1, 1891. Assuming Du Bois wrote like many of us do—up to the last minute available—he seems to have been in the library that night putting the finishing touches on the manuscript when he dashed off his assignment for English 12.

insecure about his summary of Austrian value theory, writing in a footnote that "the line of reasoning here is rather obscure and in attempting to put it briefly I fear I have made it worse" (Du Bois 1891, p. 99).

Typically, twentieth-century economists associate marginalist theories of wage determination with John Bates Clark. It may come as a surprise, then, that Du Bois's discussion of Clark is relatively brief at three-and-a-half pages, or about the same amount that he wrote on Emil Sax (an Austrian economist better known for his contributions to public finance than wage theory). Du Bois does not even include Clark with Jevons and Menger in the marginalist branch of his genealogy of economic thought (Du Bois 1891, p. 13). Instead, Du Bois places him in the branch of theorists who elaborated on Ricardian rent theory, which matches how Clark describes himself in his 1891 article, "Distribution as Determined by a Law of Rent."

Du Bois (1891) stayed current with controversies in wage theory. He reviewed the objective cost theories of Silas McVane, a Harvard professor, and the subjectivist criticism that McVane provoked from Böhm-Bawerk (1890). The debate between McVane and Böhm-Bawerk lasted from 1890 to 1894 and eventually drew in Friedrich Wieser and David Green (for an excellent recent review of this debate, see Barbieri and Filho 2024). In his discussion of all of these economists, Du Bois was highly dependent on recent volumes of the *Quarterly Journal of Economics* for source material—a clear stamp of the institutional significance of Harvard on Du Bois's work. In his review of wage theories, 13 of the 29 references were *Quarterly Journal of Economics* articles. Du Bois was especially reliant on the journal for contemporary economists and for the Austrians. Although he read German, six of the eight references in his review of Austrian economics were *Quarterly Journal of Economics* articles in English. The only book of the Austrians that he used, Böhm-Bawerk's *Capital and Interest*, was William Smart's English translation.

Du Bois (1891) emerged from his review of wage theory a committed pluralist, ready to acknowledge the contributions of all the authors he read. "In spite of all seeming diversity," Du Bois (1891, p. 114) reflects, "these theories of Wages when brought together have a great deal in common . . . It would seem indeed that the chief difference was a difference of emphasis." Du Bois's own emphasis in the last section of the *Constructive Critique* would be on developing a theory that extended Austrian marginalism by applying it symmetrically to labor supply and labor demand.

The *Constructive Critique's* "Statement of a Theory" (Du Bois 1891, pp. 130–56) begins with the observation that human beings have a variety of wants, but that the reality of the world dictated that those wants would be in conflict. Effort would have to be sacrificed to acquire what was desired, either through production or exchange. Du Bois then argues that every person has a hierarchy of wants, ordered by the amount of sacrifice they would make to acquire what they desire. He called this a "balancing of gratifications" (Du Bois 1891, p. 132).[5] Following Menger,

---

[5] Du Bois articulated this theory of value the prior December in an assignment for his English 12 course at Harvard titled simply "Value." W. E. B. Du Bois Papers (MS 312). Special Collections and University Archives, University of Massachusetts Amherst Libraries.

Diagrammatic Representation of Capitalist, Laborer, and Consumer Rankings of Wants and Sacrifices from Du Bois (1891)
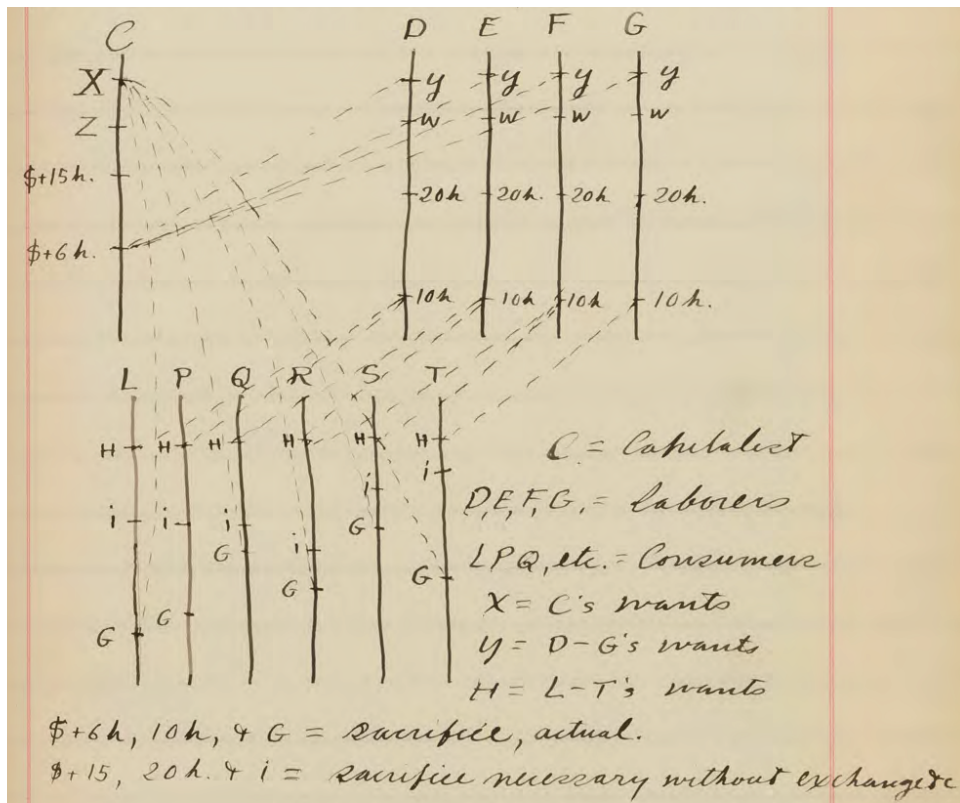
Source: Du Bois (1891, p. 144). Permission to republish granted by the Harvard University Archives.

Du Bois represents these wants through ordinal rankings, which he illustrates with a series of diagrams that built up to a full diagram of the circuit of exchange reproduced here in Figure 1.

The diagram presents the ranking of wants for a capitalist (*C*), four laborers (*D*, *E*, *F*, and *G*), and six consumers (*L*, *P*, *Q*, *R*, *S*, and *T*). The rankings include both the sacrifice that an individual makes as a result of an action (say, the laborers giving up a certain number of hours of their time for labor), as well as other wants, such as "bread, meat, clothes, Sunday excursions, etc. etc." (Du Bois 1891, p. 141). The same is true for the capitalist, who sacrifices their own time and wage payments made to laborers, and who wants "horses, houses, carriages, social position, etc. etc." (Du Bois 1891, p. 140). Du Bois includes different sacrificed wants on the rankings, depending on whether division of labor and exchange is used or not. For example, the diagram shows that without exchange, the laborer would have to sacrifice 20 hours of labor to produce their wants, *y*, for themselves. Their wants, *y*,

are ranked higher than 20 hours of labor, so they would be willing to work in autarky to produce them. However, if under exchange the laborer sold ten hours of their labor to the capitalist for $6, they would be able to afford their wants, *y*, at a much lower sacrifice of time.

The ten hours of labor that the laborer sells to the capitalist does not produce their wants directly, nor does it produce the wants of the capitalist (*X*) directly. Instead, it supplies inputs for other consumer goods (*H*) demanded by the six consumers in Du Bois's diagram. These consumers have to sacrifice wants as well to get their consumer goods, which Du Bois identifies as a price of *G*, paid to the capitalist. When consumers pay for their goods with *G*, this completes Du Bois's circuit of exchange. The capitalist pays the workers and supplies his own wants out of the payments collected from the consumers.

This diagrammatic presentation left open the question of how the wage rate, $6, was ultimately determined. Du Bois (1891, p. 146) provides a solution to this problem based on subjective opportunity cost:

> Wages, then, insofar as they are advanced are determined 1st by the wants of the capitalist, 2nd by the capitalist's estimate of the ultimate capability of the product to satisfy those wants. I have already explained the working of the first; the second will in the long run approximate the actual value of the goods in exchange.

As Du Bois had discussed in prior sections of the *Constructive Critique*, the "value of the goods in exchange," his second determinant of wages, is equal to "the sacrifice of gratification which will be made for it" by the consumer (Du Bois 1891, p. 133). Taken together, this is a statement of wages as equal to the marginal revenue product of labor. Du Bois is careful to point out that the capitalist does not want the product of labor for himself. Instead, the capitalist estimates "the ultimate capability of the product to satisfy" his wants through sales to consumers. Du Bois identifies this need to think in terms of what would ultimately be called the marginal revenue product of labor as the "peculiarity of [the] value of labor" compared to other commodities. Typically, "the value of a commodity has reference only to the wants of buyer and seller," Du Bois (1891, p. 156) explains, but "the value of labor has reference to buyer, seller, and ultimate product."

Du Bois (1891, pp. 145–6) also appreciates that marginal revenue from the product of labor alone would not provide an equilibrium wage rate, so he turns to the problem of labor supply, again using subjective opportunity cost concept of what a laborer will sacrifice to earn a wage,

> If the laborers do not find it to their advantage to take the proffered wages, then no transaction will take place, unless the capitalist is willing to increase his sacrifice, or, maintaining present sacrifice to work for less return, *Z*, for instance (see diagram). The relative return of the laborers and employer will depend on the sacrifice they are willing to undergo to satisfy their wants. If

wages are not advanced, i.e., if the sacrifice of the laborer is greater, then, by definition, the value of his want is greater and will call for more sacrifice on the part of the employer to fulfil it; another equilibrium will thus be reached giving the laborer a larger share.

Du Bois (1891) acknowledges that his theory "has obvious resemblances to the Austro-Jevons Utility Theory," and he is therefore attentive to highlighting what he understands to be the differences between them. He alleges that the major weakness of Jevons and the Austrians was that they did not explain why a certain quantity was demanded and therefore a certain price was paid to satisfy specific wants. Du Bois (1891, p. 149) writes,

> According to [the "Austro-Jevons theory"], value depends on the indispens-ableness of a certain thing to satisfy a certain want, i.e., its Final Utility; the question naturally arises why is this thing indispensable? It is answered because of its scarcity, but what has mere scarcity got to do with wants? . . . If air exists by the world-full it is just as "indispensable" that I should have a breath, as it is if there is only a cubic foot in creation. Here we come upon the weakness of the whole theory which is so noticeable in the treatment of value and its con-nection with Cost of Production.

Du Bois's solution is that it is not scarcity *per se* that, combined with marginal utility, identified the value of a thing. Instead, the sacrifice of the opposite party in an exchange gives a thing value. This sacrifice, moreover, can be compared directly with other human wants in a way that scarcity cannot because sacrifices are conflicting wants. An hour of labor is the sacrifice of an hour of leisure, which can be ranked against other wants. In effect, the Du Bois criticism of Jevons and the Austrians and his claim to fill "a manifest gap" (Du Bois 1891, p. 151) in the marginalist theory of wages is the classic criticism of Alfred Marshall, that "we might as reasonably dispute whether it is the upper or the under blade of a pair of scissors that cuts a piece of paper, as whether value is governed by utility or cost of production."[6]

Du Bois's economic analyses at Harvard resulted in a universal, marginalist theory of wage determination, and were therefore antithetical to the particularist and historical views of Gustav Schmoller and the German economists with whom he would come to identify. But it is doubtful that he inherited these marginalist views from Frank Taussig, the professor who taught him wage theory. Taussig was not an enthusiastic exponent of marginalism in the 1890s. Perhaps even more importantly,

---

[6] Interestingly, Marshall's best-known discussion of the limits of the marginal productivity theory of wages in *Principles of Economics* was not centered on the Marshallian scissors criticism. Instead, Marshall (1920, p. 518) insists that marginal productivity theory "has by itself no real meaning; since in order to estimate net product, we have to take for granted all the expenses of production of the commodity on which he works." This is a version of the "adding-up" problem resolved by Wicksteed (1894) and discussed in detail by Stigler (1941). Du Bois (1891) does not take up the adding-up problem.

Taussig was simply working on a different problem than Du Bois. Taussig was interested in the relationship between average wages and capital (the core of the wages fund theory), not the determination of individual wages in a market.

Labor and Leisure

The application of Mengerian value rankings by Du Bois led him to provide one of the earliest acknowledgements that a labor-leisure trade-off determines individual labor supply in the marginalist framework. The early marginalists spent relatively little time on the labor supply problem, and Jevons—who gave it more attention than Walras or Menger—conceived of labor supply as a trade-off between labor income and the disutility of labor. Derobert (2001) calls this a "labor-income" model, in contrast with the now canonical "leisure-income" model of labor supply in which the benefit of labor income is balanced against the benefit of leisure. Wicksteed (1910) is traditionally credited with developing the formal leisure-income model of labor supply, although Spencer (2003) identifies a precursor to Wicksteed in David I. Green's (1894) article, "Pain-Cost and Opportunity-Cost."

Both Wicksteed and Green were influenced by Carl Menger's early articulation of the principle of opportunity cost. As we have seen, Du Bois also read Menger closely. Du Bois writes of the leisure-income trade-off that "if the largest sacrifice that *A* will make for a coat be ten days labor, then we know the want of a coat is one degree higher in his estimation than the enjoyment of ten days leisure. Thus with men there must arise in every day life a certain balancing of gratifications: this gives rise to the phenomenon of value" (Du Bois 1891, pp. 132–3). Du Bois's analysis of the labor supply decision as a trade-off between leisure and the consumption made possible by labor income (the coat) is, of course, not developed as formally as Wicksteed (1910) or even Green (1894), but it is an elaboration on Mengerian value theory that anticipates the trajectory of marginalist labor economics.

Du Bois and the University of Berlin

The *Constructive Critique of Wage Theory* was a work of marginalist economics—even Austrian marginalist economics—but Du Bois did not continue in that theoretical tradition for long. In 1892 he began two years of graduate study at the University of Berlin, following in the footsteps of Taussig and many other American social scientists. Taussig had studied at the University of Berlin in late 1879 and early 1880, and maintained a deep respect for the work and influence of Adolph Wagner (Schumpeter, Cole, and Mason 1941). Given Du Bois's wide reading in economics at Harvard, it is no surprise that he was already familiar with the work of Wagner and Schmoller in 1891. Du Bois (1891, p. 16) wrote in the

Constructive Critique about how the two German scholars approached the question of a scientific law in economics:

> The extreme revolt as best represented by Schmoller of Berlin denied, at least at present, the possibility of any law, since we had not done the preliminary work of collecting the facts from historical research and from present observations. The more moderate revolt as represented by Wagner, Cohn, etc. said that the laws of the classical school could be modified and extended by taking cognizance of a much larger number of facts than older Economists did.

Du Bois obviously took Wagner's route rather than Schmoller's in the *Constructive Critique* itself, but even after arriving in Berlin he continued to respect Wagner's compromise with Schmoller's "extreme" perspective. In notes on his time in Berlin, Du Bois (1968, p. 166) wrote that Wagner's "hobby is the discovery of the golden mean between the warring extremes of his science," giving British economics in particular "due credit for their great work and agree[ing] with them more fully than with the younger German radicals headed by Schmoller." Time in Berlin therefore only reaffirmed Du Bois's view of the two men that he had developed at Harvard.

Wagner's "golden mean" reminds us that while Berlin was a major center of work in the German Historical School of Economics, it was not parochial or insular intellectually. As Kolev and Dekker (2023) point out, the divergence between Austrian and German economists in the *Methodenstreit* of the 1870s had largely subsided so that by the 1890s—when Du Bois came to Berlin—there was a greater degree of "convergence" between the two communities of scholars. Streissler (1990) similarly reminds us that Menger had great (and reciprocated) respect for the Older German Historical School of Economics, despite his methodological dispute with Schmoller and the Younger German Historical School of Economics. Yet, despite seeing Schmoller as the "extreme" view in Berlin in the early 1890s, Du Bois would more fully adopt Schmoller's attitude than any other economist. As a participant in Schmoller's seminar, Du Bois acquired a new appreciation for the German Historical School's tools of rigorous historical inquiry as a means of elaborating on economic processes and perhaps ultimately deriving scientific laws. He would never return to the abstract theorizing of his *Constructive Critique*.

Du Bois and Taussig, Redux

Before concluding, it is worth returning to the relationship between Du Bois and Taussig. Taussig is sometimes used as a foil in the literature on Du Bois, and Du Bois himself contributes to this tendency. Near the end of his life, in the pages of *The Massachusetts Review*, Du Bois remembers Taussig for "his reactionary British economics of the Ricardo school" (Du Bois 1960, p. 440). Taussig's courses dedicated "long hours on the 'Wages-fund'" and to "supporting dying Ricardean economics" (Du Bois 1960, p. 448). Du Bois repeats the cutting line "his reactionary British

economics" in his posthumous 1968 *The Autobiography of W. E. B. Du Bois*. These two recollections of Taussig are commonly cited by Du Bois scholars.[7] But this is not Du Bois's only recollection of his Harvard economics professor.

In 1946, in the pages of *The Chicago Defender*, Du Bois (1946, p. 15) remembers Taussig's instruction in a somewhat kindlier tone, and with a more subtle description of his interest in Ricardo and the wages fund theory. Du Bois (1946, p. 15) introduces Taussig as "the great American economist" and, instead of referring to "his" reactionary British economics, describes Ricardian economics as having "a certain dialectical interest" for Taussig, perhaps in recognition of the fact that Taussig's true project was to resolve the older contradictions of the wages fund theory. By this point in his life, Du Bois was committed to Marxist analysis and thinking in terms of dialectics and internal contradictions came naturally to him. Du Bois (1946, p. 15) wrote of the classical wages fund theory that "it was small, predetermined and could not be increased because of the two first demands made upon income," which were payments for raw materials and profits. With this rigid fund available for paying wages, "the Ricardians of the late nineteenth century could prove without shadow of doubt the uselessness of labor unions and the futility of strikes." On this understanding the wages fund "was fixed and within its limits men must live and reproduce the species."

This account of the wages fund theory from Du Bois (1946)—while perhaps an accurate depiction of how critics of labor unions used the theory—was not Taussig's view. Taussig (1896, p. 84) argued that it was "clear that nothing in the nature of a predetermined and rigid wages fund can be found." Interest and rent were variable, like any price, and the profits of the capitalist were the most flexible payment of all. Taussig's "elastic" wages fund opened the possibility that a general strike could raise wages by reducing capitalists' profits (Taussig 1896, p. 101). Du Bois also saw the labor movement as destined to disprove the rigid, classical, version of the wages fund theory. "Did not [Walter] Reuther put his hand upon the present difficulty in the beginning of the automobile strike?" Du Bois (1946, p. 15) asks. "He demanded higher wages at the expense of lower profit," a prospect deemed unthinkable by "reactionary" Ricardians.

Over time, Du Bois's subtler 1946 appreciation of Taussig's "dialectical interest" in the wages fund theory was forgotten and Taussig himself stood in as representative of reactionary Ricardian economics in Du Bois's memory. This evolving memory of Taussig resonates with Lewis's argument that Du Bois repeatedly "recaptured and

---

[7] Some criticism of Taussig by Du Bois scholars is almost certainly unjustified. For example, Lewis (1993) suggests that Du Bois wrote his (undated) anti-silver essay "The Free Coinage Controversy" for Taussig's Political Economy 1 class. Lewis (1993, p. 109) calls it "his Taussig essay" and "a plutocrat's delight" that "plump[ed] for the gold standard." The difficulty with laying Du Bois's "plutocrat's delight" at Taussig's feet is that in the essay Du Bois discusses the crisis of 1893 and quotes monetary statistics from 1894, putting it firmly after his time in Berlin and years after his Political Economy 1 class. Moreover, Gustav Schmoller (whose banner Du Bois took up after Berlin) was well known as a staunch gold standard advocate. Of course, this complicates the standard narrative, where *Taussig* is supposed to be the plutocratic boogeyman in Du Bois's education and *Schmoller* is the broad-minded social reformer.

revised [his Harvard experience] in fine-tuned memoirs that expertly combined truth with license" (Lewis 1993, p. 79). Even if Du Bois's recollections of Taussig half a century or more after the fact end up providing an unfair rendition of Taussig's own ideas, they nevertheless illustrate how closely Du Bois associated his Harvard economics training with developments in wage theory. Later correspondence between Du Bois and Taussig in the Du Bois Papers is sparse, but unfailingly cordial.[8] In his later years, Du Bois repeatedly recalled and named only Taussig among all of his Harvard economics professors, not because of any personal conflict with him (or any enduring friendship either), but because of Taussig's association in his mind with wage theory.

Conclusion

Despite his stature as a scholar and his formal training in economics, Du Bois has not been claimed as enthusiastically by economists as he has been by sociologists (Rabaka 2010, 2023; Morris 2015). Fortunately, just as sociologists have been reevaluating Du Bois's profile in the history of their field, some economists have been returning to Du Bois as well.

Boston (1991) provides an excellent early review of Du Bois's contributions to economics, focusing on his studies of Black life in *The Philadelphia Negro* (1899), his work for the US Bureau of Labor, and the Atlanta Studies that Du Bois either authored or supervised. Prasch (2008, p. 315) similarly presents Du Bois's historically contextualized empirical studies of Black Americans' economic experiences as "in the best traditions of the historicist school." He attributes our failure to see Du Bois as an economist to racism in American academic institutions that excluded Du Bois, but also to "the gulf between the historical and neoclassical schools" (p. 311). Oliver (2014, p. 67) identifies Du Bois's grounding in "the Berlin Historical economists" and their conflict with "laissez-faire economics" as providing a template for his own clashes with prominent white economists in the American Economic Association (AEA). Darity (1994) provides details on one particular clash between Du Bois and the AEA Committee to Investigate the Condition of the Negro, orchestrated in 1900 by Walter Willcox of Cornell University. Du Bois has also been claimed as a precursor to modern advances in economics. Stewart (2023) argues that Du Bois's work on identity and class is a crucial precursor to modern stratification economics, a point that is also developed by Numa and Zahran (forthcoming). Darity (2025) disputes Numa and Zahran's (forthcoming) view that Du Bois anticipated stratification economics, although he concedes the narrower point that Du Bois influenced stratification economics through his analysis of psychic benefits and costs in his idea of the white worker's "psychological wage" (Darity 2022).

---

[8] This is consistent with a number of accounts reported in Cord (2024) of Taussig's good relations with students.

This essay has sought to broaden our estimation of Du Bois's economic insights and contributions, and to better understand a neglected stage in his intellectual development. In *A Constructive Critique of Wage Theory*, Du Bois developed a wage theory that was at the frontier of marginal analysis at the time, and that anticipated important developments in marginal productivity theory and the theory of labor supply. Racism probably would have prevented Du Bois from winning the Toppan Prize for his *Constructive Critique*, but if he had been recognized by winning the prize, it is interesting to consider potential alternative trajectories for Du Bois after 1891. Frank Taussig's winning essay for the Toppan Prize became his first book, *Protection to Young Industries as Applied in the United States*. That book launched his career as an analyst and commentator on the tariff. Perhaps Du Bois would have published his *Constructive Critique* as a book on marginalist wage theory. Such a book would have put Du Bois in direct dialogue with Francis Amasa Walker, John Bates Clark, and other wage theorists of the late nineteenth century.

Du Bois had all the pieces in place to be a major contributor to the development of American marginalism, but he went another route. That alternative route was fruitful as well, and it deserves to be more comprehensively reincorporated into economic science. But reading the *Constructive Critique* also highlights, once again, the brilliant and creative economic mind of W. E. B. Du Bois.

References

Barbieri, Fabio, and Marcelo Lourenço Filho. 2024. "From 'Tired Muscles' to 'Might-Have-Beens': A Debate on the Nature of Costs in the Late Nineteenth Century." *Journal of the History of Economic Thought* 46 (1): 27–47.

Böhm-Bawerk, Eugen V. 1890. *Capital and Interest: A Critical History of Economic Theory.* Translated by William Smart. Macmillan.

Boston, Thomas D. 1991. "W. E. B. Du Bois and the Historical School of Economics." *American Economic Review* 81 (2): 303–06.

Broderick, Francsis L. 1958. "The Academic Training of W. E. B. DuBois." *Journal of Negro Education* 27 (1): 10–16.

Buschendorf, Christa. 2017. "'Race' as a Function of Capitalism and Imperialism: W. E. B. Du Bois's Appropriation of Marxian Ideas in 'The Black Flame.'" *Amerikastudien/American Studies* 62 (4): 567–84.

Clark, John B. 1888. "Capital and Its Earnings." *Publications of the American Economic Association* 3 (2): 1–69.

Clark, John B. 1889. "The Possibility of a Scientific Law of Wages." *Publications of the American Economic Association* 4 (1): 39–69.

Clark, John B. 1890. "The Law of Wages and Interest." *Annals of the American Academy of Political and Social Science* 1 (1): 43–65.

Clark, John B. 1891. "Distribution as Determined by a Law of Rent." *Quarterly Journal of Economics* 5 (3): 289–318.

Clark, John Bates. 1899. *The Distribution of Wealth*. Macmillan.

Cord, Robert A. 2024. "Frank W. Taussig (1859–1940)." In *The Palgrave Companion to Harvard Economics*, edited by Robert A. Cord, 113–42. Palgrave Macmillan.

Darity, William, Jr. 1994. "Many Roads to Extinction Early AEA Economists and the Black Disappearance Hypothesis." *History of Economics Review* 21 (1): 47–64.

Darity, William, Jr. 2022. "Position and Possessions: Stratification Economics and Intergroup Inequality." *Journal of Economic Literature* 60 (2): 400–426.

Darity, William, Jr. 2025. "Race Improvement: W. E. B. Du Bois, Eugenics, and Stratification Economics." Unpublished.

Derobert, Laurent. 2001. "On the Genesis of the Canonical Labor Supply Model." *Journal of the History of Economic Thought* 23 (2): 197–215.

Du Bois, W. E. B. 1891. *A Constructive Critique of Wage Theory*. https://id.lib.harvard.edu/alma/990052954990203941/catalog.

Du Bois, W. E. B. 1899. *The Philadelphia Negro: A Social Study*. University of Pennsylvania Press.

Du Bois, W. E. B. 1946. "The Winds of Time: Outmoded Wage Theory." *The Chicago Defender*, March 16, 15.

Du Bois, W. E. B. 1960. "A Negro Student at Harvard at the End of the 19th Century." *The Massachusetts Review* 1 (3): 439–58.

Du Bois, W. E. B. 1968. *The Autobiography of W. E. B. Du Bois: A Soliloquy on Viewing My Life from the Last Decade of Its First Century*. International Publishers Co., Inc.

Francis, Dania V., Bradley L. Hardy, and Damon Jones. 2022. "Black Economists on Race and Policy: Contributions to Education, Poverty and Mobility, and Public Finance." *Journal of Economic Literature* 60 (2): 454–93.

Green, David I. 1894. "Pain-Cost and Opportunity-Cost." *Quarterly Journal of Economics* 8 (2): 218–29.

Grossman, Jonathan. 1974. "Black Studies in the Department of Labor, 1897–1907." *Monthly Labor Review* 97 (6): 17–27.

Kolev, Stefan, and Erwin Dekker. 2023. "Carl Menger's Smithian Contributions to German Political Economy." *Review of Austrian Economics* 36 (2): 247–69.

Lassalle, Ferdinand. 1863. *Offenes Antwortschreiben an das Zentralkomitee zur Berufung eines Allgemeinen Deutschen Arbeiterkongresses zu Leipzig*. Edited by Eduard Bernstein. Singer, 1919.

Lewis, David Levering. 1993. *W. E. B. Du Bois, 1868–1919: Biography of a Race*. Owl Books, Henry Holt and Company.

Lewis, David Levering. 2009. *W. E. B. Du Bois: A Biography 1868–1963*. Henry Holt and Company.

Marshall, Alfred. 1920. *Principles of Economics*. Macmillan.

Morris, Aldon D. 2015. *The Scholar Denied: W. E. B. Du Bois and the Birth of Modern Sociology*. University of California Press.

Numa, Guy, and Sammy Zahran. Forthcoming. "W.E.B. Du Bois and Economics: A Reappraisal." *Journal of Economic Literature*.

Oliver, Lawrence J. 2014. "W. E. B. Du Bois and the Dismal Science: Economic Theory and Social Justice." *American Studies* 53 (2): 49–70.

Prasch, Robert E. 2008. "W. E. B. Du Bois's Contributions to US Economics (1893–1910)." *Du Bois Review: Social Science Research on Race* 5 (2): 309–24.

Rabaka, Reiland. 2010. *Against Epistemic Apartheid: W. E. B. Du Bois and the Disciplinary Decadence of Sociology*. Lexington Books.

Rabaka, Reiland. 2023. "Embryonic Intersectionality: W. E. B. Du Bois and the Inauguration of Intersectional Sociology." *Journal of Classical Sociology* 23 (4): 536–60.

Rutherford, Malcolm. 2024. "Racism, Segregation, Acceptance: American Economics and Black Labor Studies, 1885–1965." *Review of the History of Economic Thought and Methodology* 1 (1).

Schumpeter, Joseph A., Arthur H. Cole, and Edward S. Mason. 1941. "Frank William Taussig." *Quarterly Journal of Economics* 55 (3): 337–63.

Spencer, David A. 2003. "'The Labor-Less Labor Supply Model' in the Era before Philip Wicksteed." *Journal of the History of Economic Thought* 25 (4): 505–13.

Stewart, James B. 2023. "Connecting the Threads: Stratification Economics and the Political Economy of W. E. B. Du Bois." *Review of Black Political Economy* 50 (3): 269–90.

Stigler, George J. 2017. *Production and Distribution Theories*. Routledge.

Streissler, Erich W. 1990. "The Influence of German Economics on the Work of Menger and Marshall." *History of Political Economy* 22 (S1): 31–68.

Taussig, Frank W. 1883. *Protection to Young Industries as Applied in the United States: A Study in Economic History*. Moses King.

Taussig, Frank W. 1888. *The Tariff History of the United States*. G. P. Putnam's Sons.

Taussig, Frank W. 1896. *Wages and Capital: An Examination of the Wages Fund Doctrine*. D. Appleton and Company.

von Weiser, Friedrich F. 1889. *Der natürliche Werth*. A. Hölder.

Wicksteed, Phillip H. 1894. *An Essay on the Co-ordination of the Laws of Distribution*. Macmillan.

Wicksteed, Phillip H. 1910. *The Common Sense of Political Economy*. Macmillan.

Wilson, Francille Rusan. 2006. *The Segregated Scholars: Black Social Scientists and the Creation of Black Labor Studies, 1890–1950*. University of Virginia Press.

Wood, Stuart. 1889. "The Theory of Wages." *Publications of the American Economic Association* 4 (1): 5–35.

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

N. Gregory Mankiw delivered the Martin Feldstein Lecture at the Summer Institute of the National Bureau of Economic Research on the subject, "The Fiscal Future" (July 10, 2025, https://www.nber.org/research/videos/2025-17th-annual-feldstein-lecture-n-gregory-mankiw-fiscal-future). "I have no doubt that this path of a rising debt-to-GDP ratio will stop at some point. The open questions are how and when it will stop. . . . There are only five ways to stop this upward trajectory. They are (1) extraordinary economic growth, (2) government default, (3) large-scale money creation, (4) substantial cuts in government spending, and (5) large tax increases. I would encourage you to try to assign probabilities to these possible outcomes. Individually, each of these outcomes seems highly unlikely. But the probabilities you assign must sum to at least one. I say 'at least' because more than one of these

■ *Timothy Taylor is Managing Editor, Journal of Economic Perspectives, based at Macalester College, Saint Paul, Minnesota.*

outcomes could occur." Mankiw views the final option as "the most likely outcome in the long run," in part because the other options seem "implausible or unacceptable." He says: "To close a fiscal gap of 4 percent of GDP with only increased revenue, the United States would need to raise overall tax revenue by about 14 percent. That is a huge tax hike, but it would bring us only about halfway toward the level of taxation that prevails in the United Kingdom. U.S. taxes would remain below the OECD average and well below the levels in France, Italy, and Sweden. From a strictly economic standpoint, that is entirely feasible."

Alan Benson and Kathryn Shaw review the research on "What Do Managers Do? An Economist's Perspective" (*Annual Review of Economics* 2025, 17: 635–664, https://www.annualreviews.org/content/journals/10.1146/annurev-economics-082222-073303). "Research on managers has also made clear that managers help explain enormous variation in the productivity of firms. Better managers could be better in many dimensions: They may be better at recruiting, retaining, training, motivating, allocating, or evaluating. Much of the work on hiring and evaluating has pinned down the importance of leveraging managers' observation of and experience with workers. Incidentally, this mirrors a characteristic premise of the employer learning literature: Namely, that firms do not really observe match quality or ability until the worker is on the job. The empirical literature on managers corroborates this assumption but also reminds us that this knowledge often resides with the manager and not the firm. Evidently, inducing managers to act in the best interests of the firm is not straightforward. Studies are replete with examples of how managers' decisions reflect some combination of the incentives provided to them by the firm and their own considerations. Removing subjective assessments and the exercise of discretion by managers may reduce bias but also comes at the cost of the firm's ability to make well-informed decisions."

Several OECD economists—Francesco Filippucci, Peter Gal, Katharina Laengle, Matthias Schief, and Filiz Unsal—discuss "Opportunities and Risks of Artificial Intelligence for Productivity" (*International Productivity Monitor*, Spring 2025, https://www.csls.ca/ipm/48/OECD_Final.pdf). "AI could contribute between 0.3 and 0.7 percentage points to annual aggregate TFP [total factor productivity] growth in the United States over the next decade. The predicted impacts across different scenarios are highest in the United States, followed by the United Kingdom, Germany, Canada, France and Italy, and lowest in Japan. These figures indicate that Generative AI will likely be an important source of aggregate productivity growth over the next 10 years but also clarify that the expected gains from the current generation of AI technologies may not be extraordinary. For comparison, the latest technology driven boom linked to information and communication technologies (ICT) has been estimated to have contributed up to 1–1.5 percentage points to annual TFP growth in the United States during the decade starting in the mid-1990s."

The World Bank has published *21st-Century Africa: Governance and Growth*, a collection of eight chapters, edited by Chorching Goh (May 2025, https://openknowledge.worldbank.org/entities/publication/03bdf5e3-daba-404a-a638-25c5511c2376).

From the "Main Message" section at the start of the report: "Over the past 25 years, Africa has achieved notable progress . . . Mortality rates have fallen, with life expectancy rising from 50 years in 1998 to 61 years in 2022. School attendance has improved, with primary school enrollment increasing from 80 percent in 1999 to 99 percent in 2022 and secondary school enrollment increasing from 26 percent to 45 percent over the same period. The early 2000s saw strong economic growth fueled by high commodity prices. China emerged as a trade and investment partner, and the continent experienced a massive inflow of foreign capital from 17.6 percent of gross domestic product (GDP) in 1998 to 38.1 percent in 2018. . . . Aid dependence has declined, tax revenues have increased, and the median poverty rate fell by about 10 percentage points to about 43 percent." On the other side: "Africa remains the world's biggest development challenge. . . . By 2030, 90 percent of the world's extremely poor population will live in Africa. . . . Sub-Saharan Africa's share of the global economy remains at 2 percent . . . Private investment remains low, with the informal economy accounting for 59 percent of total nonagricultural employment. . . . Only 51 percent of the African population has access to electricity, compared to the global average of 91 percent." Overall, "Africa's income level per capita would be 40% higher if it had grown at the global average since 1990." The volume includes chapters on all of these topics and more.

William F. Maloney, Xavier Cirera, and Maria Marta explore *Reclaiming the Lost Century of Growth: Building Learning Economies in Latin America and the Caribbean* (World Bank, 2025, https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099060625113019654). "In 1850, the average income in the Latin America and the Caribbean (LAC) region was on a par with that of Japan, Spain, and Sweden at about 40 percent of the United States's income, and substantially above that of Korea . . . Viewing the subsequent evolution, the central growth question centers not so much on the lost decade of the 1990s, when the region slipped from 30 percent to 20 percent of US income, but rather why, beginning at the turn of the 20th century, the Asian and Nordic comparators, as well as the colonial mother countries of Portugal and Spain, were able to catch up to about 60 percent of US levels. At the same time, LAC's relative position remained unchanged. . . . Economic convergence is driven by firms and farms adopting frontier technologies, enabling growth rates faster than those of the countries that are inventing these technologies. . . . [F]or much of the past two centuries, LAC has been relatively slow at adopting technologies ranging from steamships to computers. Simulations suggest these lags can explain 83 percent of this lack of convergence or divergence . . ."

Stephen Cecchetti and Kermit L. Schoenholtz describe the state of play in "Crypto, tokenisation, and the future of payments" (*CEPR Policy Insights* 146, August 2025, https://cepr.org/publications/policy-insight-146-crypto-tokenisation-and-future-payments). "When historians look back at the decades following Bitcoin's introduction, they will ask: 'Why has crypto not 'taken off' in the way its creators and early backers hoped?' We offer three tentative answers. First, despite the hype about the speed and efficiency of digital transactions, it

turns out that transfers of Bitcoin and Ether—the leading cryptoassets—remain slow and costly. On 14 August 2025, it took an average of more than 15 minutes to confirm a Bitcoin transaction. And that time varies widely: on several days in September 2024, it took more than 2,000 minutes! . . . Small retail payments are especially costly (say, 5% for a payment of $20) in part because even the limited number of retailers who are willing to accept Bitcoin in payment typically do not wish to hold it. Second, the competition from traditional finance is intense, helping to lower costs and speed up payments. . . . Third, while both governments and private groups are expanding their efforts to track illicit crypto payments, the reputational damage from criminal activity lingers. In addition, spectacular failures in the past—such as the collapse of the FTX exchange—encourage consumer doubts about the reliability of crypto custodians. Similarly, dire head-lines about crypto-related kidnapping and torture probably deter potential crypto users who do not trust custodians and instead would consider owning a digital wallet." They argue that, in the future, "tokenised deposits and tokenised money market funds" are likely to be more important for transactions than crypto or stablecoins.

Ina Hajdini, Adam Shapiro, A. Lee Smith, and Daniel Villar discuss "Inflation Since the Pandemic: Lessons and Challenges" (August 2025, *Federal Reserve Finance and Economics Discussion Series* 2025-070, https://www.federalreserve.gov/econres/feds/files/2025070pap.pdf). From the abstract: "This paper reviews the drivers of the post-pandemic U.S. inflation surge and subsequent decline, including the behavior and role of inflation expectations. The sharp rise in inflation reflected severe imbalances between supply and demand stemming from the shocks of the pandemic and the policy response. Measures of short-term inflation expecta-tions increased alongside realized inflation, especially those of households and firms, which may have contributed to inflation's persistence through price- and wage-setting behavior. However, measures of longer-term inflation expectations remained generally well anchored, which likely prevented a larger or more lasting increase in inflation. The stability of longer-term inflation expectations, together with easing supply and demand imbalances, allowed inflation to fall from its peak in mid-2022 without a large increase in unemployment."

Fernando Avalos, Sebastian Doerr and Gabor Pinter raise the issue of "The global drivers of private credit" (*BIS Quarterly Review*, March 2025, pp. 13–30, https://www.bis.org/publ/qtrpdf/r_qt2503b.htm). "Private credit funds have increased their assets under management (AUM) from about $0.2 billion in the early 2000s to over $2,500 billion today. . . . Most funds operate as closed-end structures that lock in capital for their life cycle, which typically ranges from five to eight years. They do not trade on exchanges and are not available to retail investors, which makes them illiquid and subject to lighter regulation. The life cycle of funds usually matches the average maturity of their loan portfolios." For additional background, the IMF devoted Chapter 2 of its semiannual *Global Financial Stability Report* to "The Rise and Risks of Private Credit" (April 2025, https://www.imf.org/en/Publications/GFSR/Issues/2024/04/16/global-financial-stability-report-april-2024).

Central Bank Accountability

Ben S. Bernanke suggests "Improving Fed Communications: A Proposal" at the Federal Reserve Second Thomas Laubach Research Conference (May 15–16, 2025, https://www.federalreserve.gov/conferences/second-thomas-laubach-research-conference.htm). "Effective communication—about what the Fed sees in the economy and how it plans to respond—helps households and businesses better understand the economic outlook, clarifies and explains the Fed's policy strategy, and builds trust and democratic accountability. . . . The centerpiece . . . would be forecasts of key economic and policy variables at varying horizons, drawn from a comprehensive macroeconomic forecast led and 'owned' by the Board staff (possibly with some input and commentary from policymakers . . .). Because the underlying forecast would be internally consistent and based on explicit economic assumptions, it would provide greater insight than the projections of individual FOMC participants into the factors affecting the outlook for the economy and policy. Critically, a fully articulated baseline forecast would also facilitate the public discussion of economic scenarios that differ from that baseline. Besides highlighting the inherent uncertainty of economic forecasts, the publication of selected alternative scenarios and their implications could facilitate a subtle but important shift in the Fed's communications strategy. Specifically, it would allow the FOMC to provide policy guidance that is more explicitly contingent on how the economy evolves, underscoring for the public that the future path of policy is not unconditional ('on a preset course') but depends sensitively on economic developments and risk management considerations."

New Zealand was the first country to require that its central bank focus on an inflation target. Oliver Sikes tells how it happened in "How one Kiwi tamed inflation" (*Works in Progress*, June 12, 2025, https://worksinprogress.co/issue/how-one-kiwi-tamed-inflation/). He describes the New Zealand economy and Reserve Bank in the 1980s: "The government controlled large portions of many industries, including banking, insurance, and utilities, and the agricultural sector was supported by generous subsidies, price guarantees, and low-interest loans. Imports of goods were also tightly controlled—Kiwis needed government approval to subscribe to an overseas magazine. . . . Unlike today's central banks, which mainly control inflation through adjusting interest rates, New Zealand used direct regulatory controls on financial institutions. The government forced banks to hold specific amounts of government debt and set limits on interest rates for savers. It used capital controls to restrict money flows in and out of the country, allowing it to retain a fixed exchange rate. . . . Inflation began to fall in 1982, but only after [Prime Minister] Muldoon imposed a complete freeze on prices and wages, which then coincided with an economic contraction. . . . Recent experiences overseas showed countries could successfully rein in high inflation using tight monetary policy. Paul Volcker's Federal Reserve dropped inflation from 13.5 percent in 1980 to 3.2 percent in 1983 by raising rates to nearly 20 percent." Minister of Finance Roger Douglas asked the Reserve Bank to set public inflation goals, which ended up with a goal of inflation

in the range of 0–2 percent. "Michael Reddell, head of the Reserve Bank's monetary policy unit, said it was settled on 'more by osmosis than by ministerial sign-off'. . . . David J. Archer, a former Assistant Governor, said inflation targets were eventually chosen 'as the least bad of the alternatives available.'"

Andrew Levin asks, "Is the Federal Reserve Overstaffed or Overworked? Insights from the Fed's Financial Statements" (*George Mason University Mercatus Center Policy Brief,* March 27, 2025, https://www.mercatus.org/research/policy-briefs/federal-reserve-overstaffed-or-overworked-insights-feds-financial-statements). From the abstract: "This policy brief compares the operating expenses of the Federal Reserve with other large federal agencies. Inflation-adjusted salaries of Fed employees increased 67 percent from 2007 to 2024 but remained practically unchanged at other federal agencies. The Fed is now completing a $2.5 billion building upgrade at its DC headquarters—about ten times the cost of office renovations at the Ronald Reagan Building just a few blocks away. The Fed's management of its securities portfolio has resulted in unprecedented operating losses of $220 billion since mid-2022, and the total cost to taxpayers will be about $1.5 trillion over coming years. Each of these outcomes reflects an opaque and hierarchical institution with no constraints on the costs or efficacy of its programs and operations. Congress urgently needs to take specific steps to strengthen the Fed's public accountability."

Interviews

Tim Sablik poses the questions in "Carmen Reinhart: On twin financial and currency crises, the future of the dollar, and sovereign debt" (*Econ Focus*: Federal Reserve Bank of Richmond, Third Quarter 2025, https://www.richmondfed.org/publications/research/econ_focus/2025/q3_interview). Here's Reinhart: "When we talk about the dollar's dominance, it's important to first remember that central banks and investors are not buying greenbacks, they're buying Treasuries. And it is the unmatched liquidity of the Treasury market that supports the role of the dollar. . . . When the euro came into being, for a while it looked like, while it might not replace the dollar, you could have a situation with dual reserve currencies. Before the global financial crisis, investors tended to view all European debt—whether it was French debt, German debt, Greek debt, or Irish debt—as close substitutes. Of course, the global financial crisis completely destroyed that perception. What it boils down to is that you have very fragmented debt markets in the eurozone that don't offer the liquidity of the U.S. Treasury market. The euro is a unified currency, but there is no unification of the underlying assets that support the currency. Others have argued that the Chinese renminbi could be a contender to replace the dollar. I've never really entertained that possibility because, as Rudi Dornbusch used to say, people only go to a party if they think they can leave whenever they want to. China has capital controls, which directly impacts the liquidity of their debt market. How could you have as a reserve currency an underlying asset that in a time of need you can't sell?"

Dean Karlan served as Chief Economist at USAID from November 2022 to February 2025. Santi Ruiz interviews him "How to Fix Foreign Aid: USAID's former Chief Economist reflects on DOGE" (*Statecraft*, July 31, 2025, https://www. statecraft.pub/p/how-to-fix-foreign-aid). Here's Karlan: "There had never been an Office of the Chief Economist before. In a sense, I was running a startup, within a 13,000-employee agency that had fairly baked-in, decentralized processes for doing things. . . . [T]he reality is, we were running a consulting unit within USAID, trying to advise others on how to use evidence more effectively in order to maximize impact for every dollar spent. We were able to make some institutional changes, focused on basically a two-pronged strategy. One, what are the institutional enablers—the rules and the processes for how things get done—that are changeable? And two, let's get our hands dirty working with the budget holders who say, 'I would love to use the evidence that's out there, please help guide us to be more effective with what we're doing.' There were a lot of willing and eager people within USAID."

Discussion Starters

Matthew McCaffrey, Joseph T. Salerno, and Carmen-Elena Dorobat make a case for "The History of Economic Thought as a Living Laboratory" (*Cambridge Journal of Economics*, March 2025, 235–53, https://academic.oup.com/cje/article/49/2/235/7978384). "We argue that the history of thought can be conceived as a living laboratory of economic theorising. It is living in that it is a vital and valuable part of economics rather than a dead branch of it. It is a laboratory in that it functions as a proving ground in which theories from many different times and contexts can be examined, compared, critiqued, combined and developed. In other words, history of thought can be conceived as a method of doing economics rather than an isolated or niche field within it." Along the way, they quote Joseph Schumpeter: "[O]ur minds are apt to derive new inspiration from the study of the history of science. Some do so more than others, but there are probably few that do not derive from it any benefit at all. A man's mind must be indeed sluggish if, standing back from the work of his time and beholding the wide mountain ranges of past thought, he does not experience a widening of his own horizon . . . But, besides inspiration every one of us may glean lessons from the history of his science that are useful, even though sometimes discouraging. We learn about both the futility and the fertility of controversies; about detours, wasted efforts, and blind alleys; about spells of arrested growth, about our dependence on chance, about how not to do things, about leeways to make up for. We learn to understand why we are as far as we actually are and also why we are not further. And we learn what succeeds and how and why."

Olivier Blanchard offers "Convergence? Thoughts about the Evolution of Mainstream Macroeconomics over the Last 40 Years" (Peterson Institute for International Economics, Working Paper 25-8, May 2025, https://www.piie.com/sites/default/files/2025-05/wp25-8.pdf). "Let me state my two main conclusions. First,

starting from sharply different views, there has been substantial convergence, both in terms of methodology and in terms of architecture. Second, this convergence has been mostly in the right direction, allowing future research to build on the existing conceptual structure. Put strongly, macroeconomics may have a claim to calling itself a mature science. . . . As macroeconomists, we should stop self-flagellating and not accept flagellation from others. . . . I see the minimalist [New Keynesian] model as the basic unit in an erector set. By itself, the basic unit is not extremely useful, but you can plug into it a whole set of extensions. You can extend it to introduce myopia . . . and reduce the role of expectations. You can replace rational expectations with other expectation formation mechanisms. You can extend it to include borrowing constraints, which lead to a more important role for current variables and more realistic consumption dynamics. You can extend it to more than one country. You can extend it to introduce various forms of heterogeneity and derive aggregate implications. In short, it provides a common and generally understood structure from which to start and organize research and discussion."

The American Economic Association

Symposia

Government Debt

Yili Chien, Wenxin Du, and Hanno Lustig,
"Japan's Debt Puzzle: Sovereign Wealth Fund from Borrowed Money"

Karen Dynan and Douglas Elmendorf,
"Putting US Fiscal Policy on a Sustainable Path"

Zsolt Darvas, Lennard Welslau, and Jeromin Zettelmeyer,
"Sovereign Debt and Fiscal Integration in the European Union"

Sebastian Horn, Carmen M. Reinhart, and Christoph Trebesch,
"China's Lending to Developing Countries: From Boom to Bust"

The East Asian Tigers

Reda Cherif and Fuad Hasanov,
"Industrial Policy, Asian Miracle Style"

Nancy Birdsall, "The World Bank's *East Asian Miracle*:
Too Much a Product of Its Time?"

Articles

Ben S. Bernanke,
"Credit, Debt-Deflation, and the Great Depression Revisited"

Guido W. Imbens and Yiqing Xu,
"Comparing Experimental and Nonexperimental Methods: What Lessons Have
We Learned Four Decades after LaLonde (1986)?"

Neale Mahoney, "Why Regulate Junk Fees?"

Features

Daniel Kuehn, "Retrospectives: W. E. B. Du Bois,
Harvard Economics, and Marginalist Wage Theory"

Timothy Taylor, "Recommendations for Further Reading"

AMERICAN ECONOMIC ASSOCIATION