

the CATO JOURNAL

VOLUME 41 NUMBER 1 • PUBLISHED BY THE CATO INSTITUTE • WINTER 2021

Editor
JAMES A. DORN

Managing Editor
AMANDA GRIFFITHS

Book Review Editor
TREVOR BURRUS

EDITORIAL BOARD

TERRY L. ANDERSON
Property and Environment
Research Center

CHARLES W. BAIRD
California State University,
East Bay

RANDY E. BARNETT
Georgetown University Law
Center

JOHN H. COCHRANE
University of Chicago

KEVIN DOWD
Durham University

RICHARD A. EPSTEIN
University of Chicago

JAGADEESH GOKHALE
University of Pennsylvania

JAMES GWARTNEY
Florida State University

STEVE H. HANKE
Johns Hopkins University

RANDALL S. KROSZNER
University of Chicago

STEPHEN MACEDO
Princeton University

ANTONIO MARTINO
University of Rome, Luiss

ALBERTO MINGARDI
Università IULM

JEFFREY A. MIRON
Harvard University

KEVIN M. MURPHY
University of Chicago

GERALD P. O'DRISCOLL
Cato Institute

SAM PELTZMAN
University of Chicago

JUDY SHELTON
Fredericksburg, Virginia

VERNON L. SMITH
Chapman University

JOHN B. TAYLOR
Stanford University

ROLAND VAUBEL
Mannheim University

RICHARD E. WAGNER
George Mason University

LAWRENCE H. WHITE
George Mason University

THE CATO JOURNAL (ISSN 0273-3072) is published in the winter, spring/summer, and fall by the Cato Institute, 1000 Massachusetts Ave., N.W., Washington, D.C. 20001-5403. Annual subscriptions are \$22 for individuals and \$50 for institutions. Single copies are \$8. Two-year and three-year subscriptions are \$38 and \$55, respectively, for individuals, and \$85 and \$125, respectively, for institutions. Foreign subscribers should add \$5 per year for regular delivery and \$10 per year for airmail delivery.

Correspondence regarding subscriptions, changes of address, acquiring back issues, advertising and marketing matters, and so forth should be addressed to the Publications Department. All other correspondence, including requests to quote or reproduce material, should be addressed to the editor.

Unsolicited manuscripts cannot be returned and will be acknowledged only if accompanied by a stamped, self-addressed envelope. There is no submission fee. Articles should focus on specific public policy issues, be scholarly, but also be intelligible to the interested lay reader. Accepted manuscripts must conform to the CATO JOURNAL's style requirements. A style sheet is available on request.

Authors may submit their papers electronically, preferably in MS Word, to the editor (jdorn@cato.org), along with an abstract of no more than 250 words.

Unsolicited book reviews cannot be returned and will be acknowledged only if they are accepted for publication. They should conform to the above requirements and should provide an in-depth analysis of major studies of public policy. Reviews should range from 800 to 1,000 words.

The views expressed by the authors of the articles are their own and are not attributable to the editor, the editorial board, or the Cato Institute.

THE CATO JOURNAL is listed in the *Journal of Economic Literature*; *Current Contents/Social and Behavioral Sciences*; *America: History and Life*; *Historical Abstracts*; *International Political Science Abstracts*; and ABI/INFORM database.

Printed in the United States of America.

Copyright © 2021 by the Cato Institute.
All rights reserved.

ARTICLES

George S. Tavlas

Modern Monetary Theory Meets Greece and Chicago 1

Kam Hon Chu

Financial Development in Hong Kong and China:
A Hayekian Perspective 25

Scott Lincicome

U.S. Trade Policy toward China: Learning the Right Lessons 45

Diego A. Díaz and Cristian Larroulet

Impact of Institutions in the Aftermath of Natural Disasters 65

Anna Bocharnikova

Economic Well-Being under Plan versus Market:
The Case of Estonia and Finland 81

Tanner Corley and Marcus M. Witcher

Barber Licensing in Arkansas: Public Health or Private Gain? 115

Isabella M. Pesavento

How Misaligned Incentives Hinder Foster Care Adoption 139

Alan Reynolds

The Economic Impact of Tax Changes, 1920–1939 159

BOOK REVIEWS

Why Culture Matters Most

David C. Rose

Reviewed by **Ryan Bourne**

195

Competition Overdose: How Free Market Mythology Transformed Us from Citizen Kings to Market Servants

Maurice E. Stucke and Ariel Ezrachi

Reviewed by **Thomas A. Hemphill**

198

The Rise of Carry: The Dangerous Consequences of Volatility Suppression and the New Financial Order of Decaying Growth and Recurring Crisis

Tim Lee, Jamie Lee, and Kevin Coldiron

Reviewed by **Allan M. Malz**

205

Classical Liberalism and the Industrial Working Class: The Economic Thought of Thomas Hodgskin

Alberto Mingardi

Reviewed by **Diego Zuluaga**

209

MODERN MONETARY THEORY MEETS GREECE AND CHICAGO

George S. Tavlas

The issue is not where to find the money. . . . The money is there.

—George Papandreou (2009)

Coming up with money is the easy part.
[T]he money will be there.

—Stephanie Kelton (2020)

During the fall of 2009, George Papandreou headed the ticket of the Panhellenic Socialist Movement, known by its acronym PASOK, against the then-governing conservative party, New Democracy, in the Greek national elections. Papandreou ran on a platform that featured highly expansive fiscal spending. During a press conference on September 13, 2009, he was asked where he would find the money to fund his party's spending proposals. His answer was that given in the above quotation, by which he meant that Greece had abundant fiscal space to increase government spending; he believed that tax

Cato Journal, Vol. 41, No. 1 (Winter 2021). Copyright © Cato Institute. All rights reserved. DOI:10.36009/CJ.41.1.1.

George S. Tavlas is Alternate to the Governor of the Bank of Greece on the European Central Bank's Governing Council and Distinguished Visiting Fellow, Hoover Institution, Stanford University. He is grateful to John Cochrane, Harris Dellas, Jim Dorn, Steve Hanke, Ed Nelson, Ronnie Phillips, George Selgin, and Michael Ulan for helpful comments. He thanks Elisavet Bosdelekidou and Maria Monopoli for excellent technical support.

revenues could be sharply raised through stricter enforcement of laws against tax evasion. On October 4, PASOK won a landslide electoral victory, garnering 43.9 percent of the popular vote, compared with 33.5 percent for the second-place, incumbent New Democracy party, with the result that Papandreou became Greece's prime minister. In the following months, a sovereign-debt crisis erupted in Greece that, within a year, engulfed much of the euro area through contagion. In November 2011, Papandreou resigned the premiership, becoming the first Greek prime minister in almost 50 years to be forced out of office by his own cabinet. An article in the *Financial Times*, reporting on his ouster, stated: "George Papandreou will be remembered by Greeks with more than a trace of bitterness as the man who smilingly declared 'the money's there'" (Hope 2011). In the next Greek elections, held in June 2012, PASOK won only 12.3 percent of the vote.

In her influential book *The Deficit Myth*, Stephanie Kelton advances the view that the money could have been there for Greece if the country had not been part of the euro area. In particular, Kelton provides a diagnosis of what went wrong in Greece and a roadmap to the economic promised land for countries that follow her advice. Greece, it turns out, is Kelton's poster child for the way not to enter the promised land. "We all know what happened in Greece," she writes (Kelton 2020: 81). Greece, according to Kelton, began its journey into turbulent economic waters when it abandoned its national currency, the drachma, in 2001 and adopted the euro. By doing so, the country gave up its monetary sovereignty. It no longer was able to print its own currency to pay its debts, with the result that—well, "we all know what happened." Or do we? The problem is that Kelton does *not* know what happened.

If Greece is Kelton's poster child for the way not to run an economy, Kelton has become the poster child of the group of economists who advocate Modern Monetary Theory, or MMT. Kelton's *The Deficit Myth* made the *New York Times* bestseller list (for hardcover nonfiction) and has been the object of numerous reviews—including in this journal.¹ MMT itself has been embraced by several high-profile politicians.

¹See Dowd (2020). Despain (2020) expressed the view that "Kelton's book achieves a revolution in political economy." As I discuss below, other reviewers have been critical.

This article is an assessment of what went wrong in Greece through the lens of Kelton’s exposition of MMT. To set the stage, I first describe the central characteristics of MMT. As I will document, Kelton asserts that MMT builds on the idea of functional finance, developed in the 1940s by Abba Lerner. Functional finance views the issue of a balanced budget as of secondary importance; the primary purpose of functional finance is to ensure noninflationary full employment. Next, I provide a narrative of what went wrong in Greece. As I show, the origins of the Greek crisis had nothing to do with the loss of monetary sovereignty. Greece had monetary sovereignty in the 1980s but nevertheless found itself ensnared in financial crises because it engaged in fiscal profligacy. The fiscal profligacy led to high inflation, something that, as we shall see, Kelton says will not happen to a country that has monetary sovereignty. After Greece entered the euro area in 2001, the country acted as though it had unlimited fiscal space, resulting in the outbreak of the financial crisis in 2009. The common denominator of the crises of the 1980s and the 2009 crisis was the absence of fiscal discipline.

I then show that MMT is, in fact, a combination of two ideas developed in the 1940s: Lerner’s idea of functional finance and a proposal developed at the University of Chicago by Henry Simons, Lloyd Mints, and Milton Friedman. Like Kelton, the Chicago economists believed that fiscal deficits should be entirely backed by money creation. The Chicago proposal, however, was formulated with the objectives of disciplining fiscal policy and limiting the amount of money that could be created. The proposal aimed to attain price stability at full employment. In contrast, MMT provides no fiscal discipline and no limit on money creation, despite Kelton’s claims to the contrary. Moreover, the Chicagoans were concerned that discretionary policies can be destabilizing in light of long and variable lags. Consequently, their policy framework was rules based. As I document, functional finance’s failure to take account of the destabilizing properties of discretionary policies provided the basis of a highly critical assessment of Lerner’s functional finance proposal by Friedman.

MMT: Core Characteristics

At the “heart of MMT,” writes Kelton (2020), is the “distinction between currency *users* and the currency *issuer*” (original italics, p. 18). A currency issuer is a country that has monetary sovereignty.

Four conditions, she contends, are needed to attain monetary sovereignty: (1) a country must issue its own currency (p. 19); (2) it must borrow in that currency (p. 145); (3) it must let its currency float against other currencies (p. 145); and (4) the currency in question must be inconvertible—that is, it is “also important that [countries] don’t promise to convert their currency into something of which they could run out (e.g., gold or some other country’s currency)” (pp. 18–19). Countries like the United States, Japan, the United Kingdom, Australia, Canada, and “many more,” Kelton asserts, are currency issuers. These countries “never [have] to worry about running out of money.” The United States, for example, “can always pay the bills, even the big ones” (p. 19) because it can always print enough dollars to pay those bills.² Kelton writes: “Congress has the power of the purse. If it really wants to accomplish something, the money can always be made available . . . spending should never be constrained by arbitrary budget targets or a blind allegiance to so-called sound finance” (p. 4). Fiscal deficits, she argues, are not a problem so long as the deficits do not lead to inflation (more about that shortly). “This book,” Kelton audaciously asserts, “aims to drive the number of people who believe the deficit is a problem closer to zero” (p. 8).

The situation for currency users is very different. The countries that fall into this category have either (1) fixed their exchange rates, “like Argentina did until 2001,” or (2) “taken on debt denominated in a foreign currency, like Venezuela has done,” or (3) abandoned their national currency, as “Italy, Greece and other eurozone countries,” have done (p. 19). Those countries do not have access to the printing press to backstop their debts. Thus, we are told: “The US can’t end up like Greece, which gave up its monetary sovereignty when it stopped issuing the drachma in order to use the euro” (p. 19).

As mentioned, Kelton (pp. 60–63) acknowledges that MMT is guided by the idea of functional finance, developed by Lerner.

²The requirement that the currency of a currency issuer has to be inconvertible (at a fixed exchange rate) into other currencies betrays a lack of understanding by Kelton of international financial arrangements. All major currencies are convertible into other currencies. The imposition of inconvertibility on a currency would undermine the international demand for that currency. The U.S. dollar emerged as the dominant international currency following World War II, in part, because the dollar was the only currency that was convertible (at a fixed exchange rate) into other currencies and gold. On the international use of currencies, see Tavlas (1991).

Let's take a look at what Lerner had to say about functional finance. In his 1944 book, *The Economics of Control*, Lerner argued that the government should not hesitate to incur fiscal deficits required to achieve full employment. If the attainment of this objective entails persistent fiscal deficits (or surpluses), so be it. The size of the national debt, Lerner argued, is not important: "The [size of the] debt is not a burden on posterity because if posterity pays the debt it will be paying the same posterity that will be alive at the time when the payment is made. The national debt is not a burden on the nation because every cent in interest or repayment that is collected from the citizens as taxpayers to meet the debt service is received by the citizens as government bondholders" (Lerner 1944: 303).³ Likewise, argued Lerner, "the interest on the debt is not a burden on the nation" because those payments "are merely transferred to the recipient from taxpayer or from new lenders, and if it should be difficult or undesirable to raise taxes the interest payment can be met, without imposing any burden on the nation as a whole, by borrowing the money or printing it" (Lerner 1944: 303).

Kelton believes that Lerner "turned conventional wisdom on its head," since Lerner showed that the size of a nation's debt and its fiscal position are unimportant. Kelton states: "Instead of trying to force the economy to generate enough taxes to match federal spending, Lerner urged policy makers to think in reverse. Taxes and spending should be manipulated to bring the overall economy into balance" (p. 61). Such a policy might require "sustained fiscal deficits over many years or even decades" (p. 61). So long as inflation remains under control, "Lerner saw this as a perfectly responsible way to manage the government budget" (p. 61). Kelton's depiction of Lerner's argument that fiscal deficits and the size of a nation's debt do not matter is accurate so long as a very important qualification made by Lerner is taken into account. In particular, Lerner made it clear that a necessary condition had to be in place for a nation's fiscal position, including its debt level, not to matter. This condition, which I discuss below, is not taken into account by Kelton.

After describing Lerner's concept of functional finance, Kelton expresses the following view: "Lerner's insights are important to

³In contrast to the present author, Kelton does not quote directly from Lerner (1944).

MMT, but they don't go far enough. . . . We think Lerner's prescriptions will still leave too many people without jobs" (p. 63). Lerner, argues Kelton, thought that attaining what he called full employment would be accompanied by a level of involuntary unemployment, leaving some people out of work. To ensure that everyone has a job, "MMT recommends a federal job guarantee, which creates a nondiscretionary automatic stabilizer that promotes both full employment and price stability" (p. 63).

Here is the way MMT's economic program would work. The Federal Reserve would become subservient to the U.S. Treasury. To ensure that "funding . . . can always be made available" (pp. 234–35) for whatever purpose is deemed worthwhile, "the Federal Reserve carries out an authorized payment on behalf of the Treasury" (p. 235). Specifically, the Fed would print whatever money was needed to finance the government's spending intentions: "[W]e must recognize that the US government can supply all the dollars our private sector needs to reach full employment, *and* it can supply all the dollars the rest of the world needs to build up their reserves and protect their trade flows" (original italics, p. 151). What spending would be worthwhile? In addition to the federal job guarantee, the ability to print currency would provide the fiscal space to fund Medicare for all; free college; middle-class tax cuts; a full Green New Deal; free child care; the cancelation of student debt; affordable housing for everyone; a national high-speed rail; a Civilian Conservation Corps, the responsibilities of which would include fire prevention, flood control, and sustainable agriculture; and expanded Social Security. MMT's program would reduce racial inequalities, decrease poverty, build stronger communities, and more (pp. 229–63).

Under the federal job-guarantee program, if "the economy were to crash the way it did in 2008, the job guarantee would catch hundreds of thousands of people instead of allowing them to fall into unemployment" (p. 67). Moreover, the government would steer spending to preferred sectors of the economy. Thus, Kelton argues that: "With decent jobs guaranteed for all, workers can engage in a public-led *industrial policy* aimed at producing sustainable infrastructure and a wider array of public services" (italics supplied, p. 152). The program would establish a wage floor of "say \$15 per hour" (p. 68). That floor, in and of itself, would help "to stabilize inflation by anchoring a key price [i.e., the wage rate] in the economy" (p. 67). Yet, despite the foregoing menu of spending intentions,

Kelton tells the reader that MMT “is not a plot to grow the size of government” (p. 235).

Apart from the minimum wage’s anchoring role under the federal job program, how would Kelton keep inflation in check? She is very vague: “MMT would make us *safer* [against inflation] because it recognizes that the best defense against inflation is a good offense” (original italics, p. 72). What is a good offense against inflation? She continues: “We want agencies like the CBO helping to evaluate new legislation for potential inflation risk *before* Congress commits to funding new programs so that the risks can be mitigated preemptively.” In this connection, “If the CBO and other independent analysts concluded it [higher spending] would risk pushing inflation above some desired inflation rate, then lawmakers could begin to assemble a menu of options to identify the most *effective* ways to mitigate that risk” (original italics, p. 72). What are these options? Congress would “work *backward*” to identify areas where spending could be cut, thus ensuring “that there is always a check on any new spending. The best way to fight inflation is before it happens” (original italics, pp. 72–73). Essentially, Kelton maintains that there is always spare capacity in the economy such that the aggregate supply curve is flat. Consequently, a large, sustained fiscal stimulus, financed by money creation, would neither crowd out private expenditure nor ignite inflation.

With fiscal policy bearing the heavy work in economic stabilization, Kelton asks: “Can fiscal policy really take over the economic steering wheel? What’s left for monetary policy?” (p. 242). With the Federal Reserve having become an arm of the Treasury, monetary policy—open-market operations, changes in the discount rate, and changes in reserve requirements—merit essentially no role in Kelton’s scheme: “MMT considers *fiscal policy* a more potent stabilizer [than monetary policy]” (original italics, p. 243).

Two points about the financing of the fiscal deficits under MMT are important. First, as indicated in the foregoing synopsis of Kelton’s book, money creation would be the primary means of financing increases in government spending. Kelton argues that money and government bonds are essentially identical assets in private-sector portfolios. Why, then, does the government borrow? Why not just pay for government spending by running the printing press? Kelton states: “[The government] *chooses* to offer people a different kind of government money, one that pays a bit of interest. In other words,

US Treasuries are just interest-bearing dollars” (original italics, p. 36). It follows that the government can make “the national debt disappear” by letting the Federal Reserve purchase all the outstanding government debt (p. 99).⁴ Thus, “the purpose of auctioning US Treasuries—that is, borrowing—isn’t to raise dollars for Uncle Sam” (p. 36).⁵

Second, as is the case with sovereign borrowing, Kelton believes that taxation should not normally be used to finance government spending. Nevertheless, she maintains that taxes serve several key functions within an economic system: (1) “taxes enable governments to provision themselves without the use of explicit force”—without that tax obligation, people would not have a reason to demand the government’s fiat currency (p. 32); (2) should the need arise, taxation provides a means to combat inflation (p. 33); (3) “taxes are a powerful way to alter the distribution of wealth and income” (p. 33); and (4) taxes are a way “to encourage or discourage certain behaviors” (p. 34).

The MMT framework, including Kelton’s rendition of that framework, has been subjected to considerable, tough criticism from leading economists. Kelton’s rendition, for example, has been criticized for (1) its neglect of the 1970s, a period during which expansionary policies led to increases in both inflation and unemployment (Cochrane 2020); (2) the likelihood that it would induce unexpected inflation, reducing the purchasing power of those caught holding “old money” as “new money” is printed (Andolfatto 2020, Dowd 2020); (3) its strong proclivity to increase the size of the government in the economy, thereby diverting resources from productive firms to the quixotic public sector (Coats 2019, Tenreiro 2020); (4) its neglect of the fact that, in the absence of monetary accommodation, fiscal expansion can be a weak policy instrument (Greenwood and Hanke 2019); and (5) its neglect of the literatures on central-bank independence, the term structure of interest rates, and the effects of portfolio-balance decisions by investors on the way that policies interact with

⁴The idea that government bonds, especially short-term Treasury securities, are near-moneys was pushed forward by Simons (1936), who also believed that the Fed could eliminate the entire government debt by purchasing that debt.

⁵Kelton seems to overlook the fact that, because U.S. bank reserves now pay interest, the operation she has in mind might not reduce the interest burden on the debt and could even increase it under certain circumstances.

key economic variables (Edwards 2019).⁶ Krugman (2019) likened MMT to Calvinball, a game under which its adherents keep changing their arguments in response to criticisms; Rogoff (2019) called MMT “nonsense”; and Summers (2019) wrote that it is “a recipe for disaster.” Yet, in the present low-inflation environment, MMT continues to thrive in certain political circles and to be the focus of debate among academics while Kelton has become the leading spokesperson for the movement. After all, if inflation has remained below 2 percent in most major economies despite the highly expansionary monetary and fiscal policies of the past decade, why not take advantage of the benefits that MMT promises to deliver? Let me, therefore, assess MMT through the lens of Kelton’s “poster child” currency user.

Back to the Past, I: The Poster Child

In Kelton’s book, the case of the Greek sovereign-debt crisis is the prime example of what happens when a country gives up its monetary sovereignty to become a currency user. Here are several examples:

- “I never worry about the US ending up like Greece” (p. 81).
- “To cover [its fiscal] deficits [in 2009], Greece had to borrow. The problem is that under the euro, the Greek government no longer had a central bank that could act on its behalf by clearing all its payments” (p. 85).
- “But the US is not like Greece (which borrows in euros)” (p. 91).
- Adoption of the euro “relegates all eurozone members to mere currency users. This point is crucial to understanding Greece’s seemingly endless debt crisis” (p. 145).

What exactly happened in Greece to ignite the financial crisis that erupted in 2009? Kelton maintains that Greece fell victim to the 2008 global financial crisis and the accompanying global recession:

Countries like Greece and Italy, along with the other seventeen members of the eurozone, gave up their sovereign currencies in order to use the euro. Since they can’t issue the

⁶Cochrane (2020) points out that Kelton neglects essentially all the contributions made to the literature on money and inflation since the end of World War II.

euro, member governments must cover fiscal deficits by selling bonds. . . . The problem is that lending to these countries became especially risky once they started promising to repay bondholders in a currency they could no longer issue themselves. This became painfully clear in the wake of the 2008 financial crisis, as the global recession pushed the budgets of Greece and other eurozone countries deeply into deficit. . . . Governments had no choice but to borrow in private financial markets, and they had to pay whatever the market demanded to secure the funding they needed. . . . In Greece, the poster child for the crisis, interest rates on ten-year government bonds skyrocketed from 4.5 percent in September 2008 to nearly 30 percent by February 2012 [p. 124].

Two points made in the above excerpt concerning Greece are important to highlight. First, Kelton argues that it was the global recession associated with the 2008 financial crisis that pushed Greece's budget "deeply into deficit." Second, Kelton identifies the period during which interest rates on Greek government bonds "skyrocketed" as having begun in September 2008. (Recall that September 2008 saw the collapse of Lehman Brothers.)

Kelton's assessment of the origin of the Greek financial crisis begs the question: Why Greece? Germany, Luxembourg, and the Netherlands are members of the euro area. Those countries also gave up their monetary sovereignty to adopt the euro. Yet, those countries—and others in the euro area—did not experience financial crises following the outbreak of the 2008 global recession. As I explain below, there was a reason for this circumstance.

Greece became a member of the euro area in 2001. Why did the country decide to transform itself from a currency issuer to a currency user? To address this issue, let's go farther back in time to the 1980s and the early 1990s, a period during which the country had its own currency, the drachma. Beginning in 1981, the country undertook highly expansionary fiscal policies.⁷ As a result, the fiscal deficit, which was 2.6 percent of GDP in 1980, rose to 8.7 percent of GDP in 1981. During the period 1981 to 1994, the deficit-to-GDP ratio

⁷The year 1981 saw the election of PASOK to power; Andreas Papandreou, the father of George Papandreou, whom we encountered at the beginning of this article, became prime minister.

averaged 11.7 percent.⁸ During most of the period, the Bank of Greece was subservient to the government. Indeed, at one point, the same individual simultaneously held the positions of finance minister and governor of the Bank of Greece. Consequently, monetary policy's role was to help finance the fiscal deficits—annual money growth averaged more than 20 percent during the period 1981 to 1994.⁹ In light of the large fiscal deficits, the debt-to-GDP ratio exploded, from 28.4 percent in 1980 to 107.9 percent in 1994. The key macroeconomic performance indicators during 1981 to 1994 tell the resulting story: the nominal interest rate on the 10-year government bond was consistently near 20 percent, despite the activity of the Bank of Greece in the sovereign-bond market; inflation—something that Kelton says will not happen to a currency issuer—averaged 18 percent; real growth averaged 0.8 percent; and the current account consistently registered deficits in the range of 3 percent to 5 percent of GDP. In those circumstances, the country faced a series of financial crises as the drachma came under attack. Several adjustment programs were undertaken, but they were subsequently abandoned in favor of a reversion to fiscal and monetary expansion.

Having experienced the costs (very low growth, high inflation, financial crises) of overly expansionary macroeconomic policies, in the mid-1990s, Greece committed to putting its economic house in order with the goal of joining the euro area.¹⁰ Fiscal and monetary policies were tightened. The reorientation of policies succeeded, inflation and the fiscal imbalances were reduced, and the country entered the monetary union. The adoption of the euro was expected to produce a low-inflation, low-interest-rate environment conducive to economic growth.

⁸The data for this section are from Garganas and Tavlas (2001), Tavlas (2019), and the Bank of Greece.

⁹The figures refer to M3, a broad measure of the money supply. The Bank of Greece announced annual targets for M3 beginning in the early 1980s.

¹⁰The decision to join the monetary union was underpinned, in part, by a substantial literature—called the “new” theory of optimum currency areas—that showed that countries with histories of high inflation can increase their monetary-policy credibility by having monetary policy determined by a conservative, regional central bank. The increase in credibility, in turn, was said to reduce inflation, interest rates, and the unemployment costs of moving to a low-inflation equilibrium. For a discussion of this literature, see Tavlas (1993) and Dellas and Tavlas (2009).

That is precisely what happened—at least for a while. From 2001 to 2007, inflation moved into the low single digits and real growth averaged about 4 percent a year. In addition, interest rates came down. The spread between 10-year Greek and German sovereigns fell from over 600 basis points in the late-1990s to between 10 and 20 basis points several years after Greece joined the euro area. Greece seemed to have found the magic formula for economic success.

Beneath the surface, however, deep-seated problems emerged. An expanding government sector underpinned economic growth, and the country was internationally uncompetitive. Consider the following developments, mainly during the period from 2001, the year of entry into the euro area, to 2009:

- Fiscal deficits consistently topped 6 percent of GDP, peaking at 15 percent at the end of the period.
- The widening of the fiscal deficits was mainly expenditure driven; the share of government spending in GDP rose by eight percentage points—to 54 percent.
- The ratio of government debt to GDP rose from below 100 percent in the late 1990s to 125 percent at the end of the period. Between 2001 and 2009, the level of government debt almost doubled, from 163 billion euros to 301 billion euros.
- The fiscal expansion was mainly financed by external borrowing. The share of government debt held by non-Greek residents rose from 48 percent in 2001 to 79 percent in 2009.
- Greece's competitiveness, measured in terms of unit labor costs relative to those of its major trading partners, deteriorated by 30 percent.

The loss in competitiveness, and the relatively high growth rates, led to a widening of the current account deficit. Greece entered the euro area with an already-high current account deficit equal to almost 7 percent of GDP. Near the end of the period, that deficit had widened to 15 percent of GDP.¹¹

Despite these large and growing imbalances—which should have sounded loud warning alarms to the financial markets, but did

¹¹ Edwards (2019: 557) pointed out that a core characteristic of MMT is the belief that current account deficits are beneficial phenomena because the rest of the world has no problem in providing the country incurring the deficits with essentially unlimited real resources in exchange for the country's debt.

not—and despite the global financial crisis that erupted in August 2007 (and intensified in September 2008 with the collapse of Lehman), interest rates on Greek sovereigns remained at low levels until 2009. The markets apparently believed that, regardless of the Maastricht Treaty’s no-bail-out clause, if things went wrong, Greece’s euro area partners would be obliged to bail out the country to maintain the cohesion of the monetary union and to prevent negative spillovers to other countries.¹² The markets continued to lend to the Greek government, and the government continued to spend and borrow as though it had unlimited fiscal space.

The turning point came during the elections in the fall of 2009. In October, shortly after the national elections, the new PASOK government announced that the 2009 fiscal deficit would soar to 12.5 percent of GDP, far higher than estimates provided by the former conservative (New Democracy) government (Barber 2009).¹³ The announcement focused market attention on Greece’s large fiscal imbalances, including the amount of government debt owed to non-Greek residents. Figure 1 shows the time profile of the interest rate on 10-year Greek government bonds. As shown in the figure, interest rates spiked upward in October 2009 and continued to rise for the next several years. What caused the Greek financial crisis was fiscal profligacy. Prior to October 2009, financial markets behaved as though the country’s fiscal and external positions were sustainable. Those euro area countries that did not incur fiscal and external imbalances—such as Germany, the Netherlands, and Luxembourg—did not face a crisis.¹⁴ Contrary to what Kelton argues, what caused the Greek financial crisis was not the loss of monetary sovereignty.

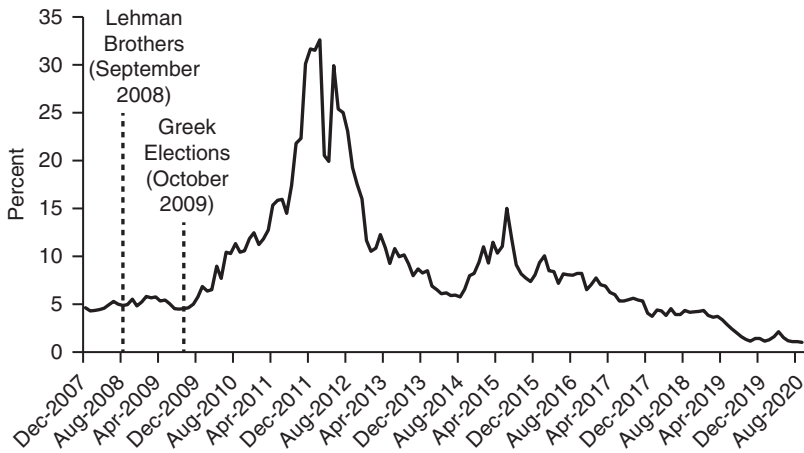
The following conclusions emerge. First, Kelton’s thesis that the 2008 global recession pushed Greece’s fiscal position “deeply into deficit” is incorrect. The 2008 recession exacerbated Greece’s *already large* fiscal deficits. The country ran fiscal deficits on the

¹²Article 25 of the Maastricht Treaty stipulates the euro area member states cannot take on the debts of other member states.

¹³In the event, the 2009 fiscal deficit equaled 15 percent of GDP. An important reason for the large fiscal deficit in 2009 was expansionary fiscal spending. As noted, the year 2009 was an election year in Greece; election years were typically marked by high government spending prior to the election to gain favor with voters.

¹⁴In the several years leading up to 2009, Germany, the Netherlands, and Luxembourg had fiscal positions near balance and maintained current account surpluses.

FIGURE 1
GREEK GOVERNMENT BOND—10-YEAR YIELD



SOURCE: Bank of Greece.

order of 7 percent of GDP throughout the period leading up to 2008.¹⁵ Second, Kelton’s argument that interest rates on Greek government debt “skyrocketed” beginning in September 2008, which, if accurate, would support her contention that the origination of the Greek financial crisis lay with factors external to Greece, is similarly incorrect. As clearly shown in Figure 1, interest rates on Greek sovereigns began their ascent in October 2009, coinciding with the market’s recognition that Greece’s fiscal situation was unsustainable. The interest rate on the 10-year Greek government bond was 5.0 percent in September 2008 (Figure 1).¹⁶ It fell subsequently, reaching 4.5 percent in September 2009. Following the announcement about the size of the 2009 fiscal deficit in October of that year, the interest rate began its upward climb. By January 2010, the rate was at 6.9 percent; by June 2010, it had reached 10.5 percent. The rate peaked at

¹⁵As percentages of GDP, the fiscal deficits between 2003 and 2007 averaged 7.1 percent; the 2008 deficit was 10.2 percent.

¹⁶As indicated above, Kelton (p. 124) reports that the interest rate on the Greek 10-year government bond was 4.5 percent in September 2008, which differs from the 5.0 percent figure that I provide. I conjecture that she has reported the interest-rate spread on Greek 10-year sovereigns over German government bonds, instead of the interest-rate level.

32.6 percent in February 2012. Third, Greece experienced financial crises both in the 1980s and during the period from 2009 to 2015. In the former period, the country was a currency issuer. In the latter period, it was a currency user. Both periods were marked by fiscal profligacy as successive governments behaved as if they had unlimited fiscal space. Fiscal profligacy was the common denominator in the crises.

Back to the Past, II: Kelton, Lerner, and Chicago

Two monetary doctrinal issues shed light on the cohesiveness and soundness of Kelton's framework: (1) Kelton's treatment of Lerner's functional-finance proposal and (2) the similarities and differences between Kelton's framework and the monetary framework developed by Chicago economists in the 1940s. Kelton generously acknowledges the influence of Lerner's formulation of functional finance on her rendition of MMT (pp. 60–63). Like Lerner, Kelton believes that the national debt is not a burden on future generations. Like Lerner, Kelton maintains that the interest paid on the debt is also not a burden on the nation. However, as I mentioned, Lerner believed that an important, third condition had to be in place for the previous two conditions to be valid. What was that third condition? Lerner recognized that the national debt had to be *internally* owned. He wrote:

All this is true, of course, only of internally held national debt. Increasing debt to other countries or to the citizens of other countries *does* indicate impoverishment of the borrowing country and enrichment of the lending country. . . . For a country to borrow from another country may be foolish or wise according to circumstances, just as in the case of individual borrowing. Such debt should be limited because the *repayment will constitute a real burden* on the country just as the borrowing provided a real benefit quite different from any benefit that can accrue from internal borrowing. *When the time comes to make the repayment there may be great inconvenience which could lead to default* [Lerner 1944: 305; italics supplied].

Lerner believed that all externally owned debt was a burden on a country; he did not distinguish between external debt that was denominated in a particular country's currency and external debt that was denominated in foreign currencies.

Kelton does not mention the above qualification made by Lerner to his idea of functional finance.¹⁷ In Kelton's view, the United States performs a service for other countries when it supplies those countries with U.S. dollars, and those countries use the dollars to buy U.S. Treasury bonds. When the United States imports goods from China, for example, and pays for the imports with newly created dollars, Kelton writes: "We're not really borrowing from China [even though the U.S. dollars are a debt obligation of the United States] so much as we're supplying China with dollars to be transferred into a U.S. Treasury security. . . . [Therefore,] terms like *borrowing* [to describe the U.S. side of the transaction] are misleading. So is labeling these securities as national debt. There is no *real* debt obligation" (original italics, p. 83). Why is there no real debt obligation? Kelton argues there is an unlimited demand for the dollar as an international currency. Thus, the United States can simply keep printing dollars to pay for the maturing Treasury securities held by China (pp. 82–83).

A fundamental condition for the international use of a currency is that there has to be confidence in the currency's value. Specifically, high or variable inflation rates add to the costs of using a currency internationally by generating nominal exchange rate depreciation and variability. These factors increase the costs of acquiring information and making efficient calculations about the prices bid and offered for tradable goods and capital assets. Thus, they undermine a currency's use as an international medium of exchange, unit of account, and store of value. Furthermore, inflation increases the costs of holding a currency by eroding its purchasing power, debasing the currency as an international store of value. Kelton's roadmap to the economic promised land considers that money creation will be subservient to the fiscal authorities and politicians' needs and desires. Consequently, money creation would be endogenous. The fiscal authorities and the politicians would see to it, Kelton believes, that money-financed deficits will not create inflation. However, as we have seen in the case of Greece in the 1980s, making money creation subservient to the fiscal authorities can be a roadmap—not to the economic promised land but to inflation, higher long-term interest rates (because of inflation), higher-risk premia, and financial crises.

¹⁷By failing to do so, Kelton does not provide an accurate account of Lerner's views.

I suggest that the dollar's international role, and the seigniorage that the United States derives from that role, would not last long under Kelton's roadmap.

Kelton's policy framework shares several characteristics with the framework developed in the 1940s by Chicago economists, Simons, Mints, and Friedman. Nevertheless, Kelton's—and MMT's—policy proposals were orthogonal to the proposals pushed forward by the Chicagoans. Hence, a comparison of those frameworks will be instructive. I should mention that, unlike Lerner's work, Kelton evinces no awareness of the contributions made by the Chicago economists in the 1940s.¹⁸

Lerner (1944: 303) was indifferent between bond-financed and money-financed fiscal deficits. In contrast, the earlier Chicago economists believed that fiscal deficits should be entirely backed by money creation, a view shared by Kelton.¹⁹ Like Kelton, the Chicagoans believed that the Federal Reserve should be made subservient to the U.S. Treasury.²⁰ Like Kelton, the Chicagoans thought that reliance on open-market operations to conduct monetary policy should be reduced.²¹

There were, however, foundational differences between the framework developed by the earlier Chicagoans and Kelton's framework. First, the Chicagoans believed that money was a crucial variable in the economy and that its quantity needed to be controlled. Otherwise, increases in the quantity of money could lead to self-generating rises in inflation. In contrast, Kelton's framework makes the quantity of money endogenous to the government's fiscal position. Second, the Chicagoans believed that monetary policy uncertainty was a primary cause of disturbances to the economy.

¹⁸ Kelton does refer to Friedman's idea that there is a natural rate of unemployment. Friedman advanced that idea in the 1960s (see Nelson 2020).

¹⁹ See Tavlas (2015) for a discussion of the policy frameworks of the earlier Chicago economists. Friedman's views on monetary policy changed over time. Nelson (2020) provides the definitive analysis of the development of Friedman's views on economics from the early 1930s to the early 1970s.

²⁰ Simons (1936: 175) wrote: "Ultimate control over the value of money lies in fiscal practices—in the spending, taxing and borrowing operations of the central government. Thus . . . the Treasury would be the primary administrative agency."

²¹ Friedman (1948) thought that open-market operations should be discontinued.

Kelton's proposal does not consider the role of policy uncertainty. Third, in light of the first two factors, the Chicagoans thought that monetary policy should be based on a rule that aims to achieve a stable price level.²² Such a rule would, in turn, help keep money creation out of the political process by strictly limiting the quantity of money. Simons (1944: 224) put it as follows: "[O]nly by recognizing and by accepting this [price-level-stabilization] rule can legislatures be made responsible financially or business spared intolerable monetary uncertainty."²³ In contrast, Kelton's framework is discretion based. She believes that her federal job-guarantee proposal would be compatible with price stability under the watchful (i.e., discretionary) eye of the Congressional Budget Office; she does not address the role of monetary uncertainty in the economy.

Under the Chicago framework, the role of the government would be to establish the monetary rule and to ensure that it was followed—that is, the government's role would be constitutional as opposed to administrative. Under that framework, the budget would be balanced over the course of the business cycle, with the aim of limiting the size of the government. Should the automatic stabilizers fail to provide sufficient demand during the trough of the cycle, increases in the size of the fiscal deficit, and, thus, increases in the quantity of money, would be generated through reductions in taxes—and not by increases in government spending.²⁴ The Chicagoans believed that the amount of fiscal space available was subject to strict limits. In contrast, under Kelton's proposal, which would increase the size of the government sector, the government's role would be administrative. Under her proposal, the amount of fiscal space would be unlimited before a vaguely defined inflation constraint kicks in.

²²Friedman (1948) proposed a rule that aimed to stabilize aggregate demand at full employment. Beginning in the mid-1950s, he favored a constant-money-supply-growth rule to stabilize the price level. Taylor (2017) noted that a benefit of the Taylor rule is that it would reduce monetary uncertainty. For a discussion of the rationale underlying monetary rules, see Dorn (2018).

²³To better control the quantity of money and to help moderate the business cycle, the Chicagoans advocated 100 percent reserve requirements on demand deposits. See Tavas (2020).

²⁴However, the Chicagoans believed that the progressive-income-tax system provided the most appropriate means for achieving greater equality.

There was an important reason underlying the Chicagoans' aim to prevent the concentration of power in the government: those economists wanted to preserve individual liberty. In a 1945 paper, Simons expressed this view as follows:

[The] strength [of the competitive system] is in its implied political philosophy. Its wisdom is that of seeking solutions which are within the rule of law, compatible with great dispersion or deconcentration of power, and conducive to extensive supranational organization on a basis that facilitates indefinite peaceful extension. . . . Certainly another kind of system, ruled by authorities, *might be* more efficient and more progressive—if one excludes liberty as an aspect of efficiency and capacity for freedom and responsibility, among individuals and among nations, as a measure of progress. Discretionary authorities, omniscient and benevolent, surely could in some sense do better than any scheme involving democratic, legislative rules and competitive dispersion of power. After any disturbing change they could promptly effect the same arrangements which competition would achieve slowly or with 'unnecessary' oscillations. Indeed, they could probably avoid all real disturbances by anticipating them! But some of us dislike government by authorities, partly because we think they would not be wise and good and partly because we would still dislike it if they were [Simons 1945: 308–9; original italics].

Additionally, the Chicagoans believed that discretionary policies can, in practice, be destabilizing. Specifically, the long and variable lags associated with discretionary policies can mean that counter-cyclical actions will be a source of disruptions. For example, the effects of a policy easing aimed at supporting aggregate demand and raising inflation might not kick in until the expansionary phase of the cycle. Lerner's functional-finance proposal's failure to take account of lags formed the basis of Friedman's criticism of the proposal. In a 1947 article, "Lerner on the Economics of Control," Friedman appraised functional finance as follows:

[Functional finance] conflicts with the hard fact that neither government action nor the effect of that action is instantaneous. There is likely to be a lag between the need for action and government recognition of this need; a further lag

between recognition of the need for action and the taking of action; and a still further lag between the action and its effects. If these time lags were short relative to the duration of the cyclical movements government is trying to counteract, they would be of little importance. Unfortunately, it is likely that the time lags are a substantial fraction of the duration of the cyclical movements. In the absence, therefore, of a high degree of ability to predict correctly both the direction and the magnitude of required action, governmental attempts at counteracting cyclical fluctuations through “functional finance” may easily intensify the fluctuations rather than mitigate them. By the time an error is recognized and corrective action taken, the damage may be done, and the corrective action may itself turn into a further error. This prescription of Lerner’s, like others, thus turns into a exhortation to do the right thing with no advice how to know what is the right thing to do [Friedman 1947: 315–16].

Conclusion

Kelton’s book makes many promises and offers a straightforward way to pay for them: run the printing press. The appeal of the proposal, especially to some politicians, is that it aims to deliver a “people’s economy” at no cost. However, Kelton’s appraisal of her poster child, Greece, for the way not to run an economy is an inaccurate depiction of *monetary* history. Greece’s recent monetary history shows that it was the country’s fiscal profligacy—not its status as either a currency issuer or currency user—that lay at the heart of the country’s financial crises. Greece’s economic waters were turbulent for many years before the country adopted the euro; those waters only stabilized—temporally—by the commitment to fiscal sobriety prior to its entry into the euro area. Moreover, Kelton’s description of Lerner’s idea of functional finance is an inaccurate depiction of *doctrinal* history. In contrast to Kelton, Lerner evinced a deep concern about the consequences of any fiscal program that does not pay attention to externally held government debt. The above comparison of Kelton’s framework with that of Chicago economists in the 1940s shows that, unlike the latter framework, Kelton’s scheme does not aim to provide limits on money creation needed to maintain price stability, assumes essentially unlimited fiscal space, does not account for monetary uncertainty or the destabilizing effects of lags under

discretionary policy, assumes that there is always spare capacity such that the aggregate supply curve is flat, and would increase both the size and the administrative role of government in economic affairs.

An important consideration in evaluating Kelton’s proposal is that the author has ventured into an area—the subject of money—which appears to be uncharted waters for her. She is by no means the first person to have done so; nor is she likely to be the last. Let me, therefore, conclude with something that Chicagoan Mints wrote in his 1950 book, *Monetary Policy for a Competitive Society*:

[M]onetary policy is a subject which requires the exercise of judgement, and this being so, it can hardly be expected that differences of opinion will not arise. For the same reasons it is also nearly inevitable that monetary cranks of many descriptions should appear who have their “sure cures” for the afflictions of society. They see one thing that might possibly be accomplished by monetary measures, but they do not understand the many other effects of their proposals [Mints 1950: v].

References

- Andolfatto, D. (2020) “Modern Monetary Theory.” *MacroMania* (March 14).
- Barber, T. (2009) “Greece Vows Action to Cut Budget Deficit.” *Financial Times* (October 20).
- Coats, W. (2019) “Modern Monetary Theory: A Critique.” *Cato Journal* 39 (3): 563–76.
- Cochrane, J. H. (2020) “‘The Deficit Myth’ Review: Years of Magical Thinking by Stephanie Kelton.” *Wall Street Journal* (June 5).
- Dellas, H., and Tavlas, G. S. (2009) “An Optimum-Currency-Area Odyssey.” *Journal of International Money and Finance* 28 (7): 1117–37.
- Despain, H. G. (2020) “Book Review: *The Deficit Myth: Modern Monetary Theory and the Birth of the People’s Economy* by Stephanie Kelton.” Available at <https://blogs.lse.ac.uk/lseireviewofbooks/2020/06/22/book-review-the-deficit-myth-modern-monetary-theory-and-the-birth-of-the-peoples-economy-by-stephanie-kelton>.
- Dorn, J. A. (2018) “Monetary Policy in an Uncertain World: The Case for Rules.” *Cato Journal* 38 (1) (Winter): 81–108.

- Dowl, K. (2020) "Book Review: *The Deficit Myth: Modern Monetary Theory and the Birth of the People's Economy* by Stephanie Kelton." *Cato Journal* 40 (3): 795–809.
- Edwards, S. (2019) "Modern Monetary Theory: Cautionary Tales from Latin America." *Cato Journal* 39 (3): 529–61.
- Friedman, M. (1947) "Lerner on the Economics of Control." *Journal of Political Economy* 55 (October): 405–16. Reprinted in M. Friedman (ed.), *Essays in Positive Economics* (1953: 303–24). Chicago: University of Chicago Press.
- _____. (1948) "A Monetary and Fiscal Framework for Economic Stability." *American Economic Review* 38: 245–64. Reprinted in *Essays in Positive Economics*, 133–56.
- Garganas, N., and Tavlas, G. S. (2001) "Monetary Regimes and Inflation Performance." In R. C. Bryant, N. C. Garganas, and G. S. Tavlas (eds.), *Greece's Economic Performance and Prospects*, 43–95. Athens and Washington: Bank of Greece and Brookings Institution.
- Greenwood, J., and Hanke, S. H. (2019) "Magical Monetary Theory." *Wall Street Journal* (June 4).
- Hope, K. (2011) "Humiliating End to Greece's Social Reformer." *Financial Times* (November 7).
- Kelton, S. (2020) *The Deficit Myth: Modern Monetary Theory and the Birth of the People's Economy*. New York: PublicAffairs.
- Krugman, P. (2019) "Running on MMT (Wonkish)." *New York Times* (February 25).
- Lerner, A. P. (1944) *The Economics of Control: Principles of Welfare Economics*. New York: Macmillan.
- Mints, L. (1950) *Monetary Policy for a Competitive Society*. New York: McGraw-Hill.
- Nelson, E. (2020) *Milton Friedman and Economic Debate in the United States, 1932–1972*. Chicago: University Chicago Press.
- Papandreou, G. A. (2009) Press conference, 74th Thessaloniki International Fair (September 13).
- Rogoff, K. (2019) "Modern Monetary Nonsense." *Project Syndicate* (March 4).
- Simons, H. C. (1936) "Rules versus Authorities in Monetary Policy." *Journal of Political Economy* 64: 1–30. Reprinted in H. C. Simons (ed.), *Economic Policy for a Free Society* (1948: 160–83). Chicago: University of Chicago Press.

- _____ (1944) “On Debt Policy.” *Journal of Political Economy* 52: 356–61. Reprinted in H. C. Simons (ed.), *Economic Policy for a Free Society* (1948: 220–30). Chicago: University of Chicago Press.
- _____ (1945) “The Beveridge Program: An Unsympathetic Interpretation.” *Journal of Political Economy* 53: 212. Reprinted in *Economic Policy for a Free Society*, 277–312.
- Summers, L. H. (2019) “The Left’s Embrace of Modern Monetary Theory is a Recipe for Disaster.” *Washington Post* (March 5).
- Tavlas, G. S. (1991). *On the International Use of Currencies: the Case of the Deutsche Mark*. Princeton University, Department of Economics, International Finance Section *Essays in International Finance* No. 181 (March).
- _____ (1993) “The ‘New’ Theory of Optimum Currency Areas.” *World Economy* 16 (6): 663–85.
- _____ (2015) “In Old Chicago: Simons, Friedman, and the Development of Monetary-Policy Rules.” *Journal of Money, Credit and Banking* 47 (1): 99–121.
- _____ (2019) *The Theory of Monetary Integration in the Aftermath of the Greek Financial Crisis*. The Seventh Annual Marco Minghetti Lecture, with comments by P. C. Padoan and P. Savona. Rome, Italy: Istituto Luigi Struzzo (January).
- _____ (2020) “On the Controversy Over the Origins of the Chicago Plan for 100 Percent Reserves: Sorry Frederick Soddy, It Was Knight and (Most Probably) Simons!” Hoover Institution, Economics Working Paper No. 20102. Forthcoming in *Journal of Money, Credit and Banking*.
- Taylor, J. B. (2017) “Rules versus Discretion: Assessing the Debate over the Conduct of Monetary Policy.” NBER Working Paper Series No. 24149. Available at www.nber.org/system/files/working_papers/w24149/w24149.pdf.
- Tenreiro, D. (2020) “What the Deficit Myth Lacks.” *National Review* (June 8).

FINANCIAL DEVELOPMENT IN HONG KONG AND CHINA: A HAYEKIAN PERSPECTIVE

Kam Hon Chu

This article draws on certain Hayekian ideas, such as the Hayekian knowledge problem and spontaneous order, to better understand financial development in Hong Kong and China. Ignorance or neglect of such ideas can lead to devastating consequences. A case in point is Hong Kong. If Hong Kong drifts from its grounding in the rule of law and individual freedom, under pressure from the Chinese mainland, its dynamic financial markets may suffer as investors lose trust in Hong Kong's institutions.

Hong Kong as an International Financial Center

Hong Kong's financing and insurance industries account for about 20 percent of GDP and more than 6 percent of the total labor force.¹ These figures do not include the closely related industries like law, accounting, and information technology, not to mention the vital and conducive role of a financial system to the functioning of an economy.

Cato Journal, Vol. 41, No. 1 (Winter 2021). Copyright © Cato Institute. All rights reserved. DOI:10.36009/CJ.41.1.2.

Kam Hon Chu is Professor of Economics at the Memorial University of Newfoundland. He would like to thank an anonymous referee and Jim Dorn for their helpful suggestions and comments on earlier versions of this article. The usual disclaimer applies.

¹The percentage contribution to the labor force is based on the labor statistics for 2019, whereas the contribution to GDP is based on the latest available statistics under the production approach for 2018. See Census and Statistics Department, Hong Kong SAR (2020) for details.

Hong Kong is an indispensable window through which China absorbs foreign investment and new technology. According to statistics released by the Ministry of Commerce of the People’s Republic of China (2020a, 2020b), more than half of the actually utilized foreign investment during the last decade came from Hong Kong (Table 1). Over the years, foreign investment in China was on an upward trend until 2018, and the share of foreign investment coming from Hong Kong had also increased from just below 50 percent in 2009 to 65 percent in 2018, reflecting the increasing reliance of China on Hong Kong as a source of foreign investment. After 2018, China’s foreign investment from Hong Kong and the rest of the world declined noticeably, at least in the first 10 months of 2019. Despite the apparent slowdown in foreign investment in China and the anti-extradition protests in Hong Kong in 2019, Hong Kong has remained the largest source of foreign investment for China and its investment share has become relatively more important at the same time.

TABLE 1
UTILIZED FOREIGN DIRECT INVESTMENT IN
CHINA, 2009–2019 (US\$ BILLION)

Year	From Hong Kong	From All Nations/Regions	Hong Kong’s Share (%)
2009	46.1	94.1	49.0
2010	60.6	114.7	52.8
2011	70.5	124.0	56.9
2012	65.6	121.1	54.2
2013	73.4	123.9	59.2
2014	81.3	128.5	63.2
2015	86.4	135.6	63.7
2016	81.5	133.7	60.9
2017	94.5	136.3	69.3
2018	89.9	138.3	65.0
January–June, 2019	50.1	70.7	70.8
July–October, 2019	29.6	40.0	74.0
January–October, 2019	79.7	110.8	71.9

SOURCES: Data for 2009–2018 are from the Ministry of Commerce, Peoples Republic of China (2020a). Data for 2019 are from the Ministry of Commerce, PRC (2020b).

In addition to foreign investment absorption, Hong Kong plays a pioneering role in the internationalization of the renminbi (RMB). Despite the lack of comprehensive statistics on the volume of offshore RMB transactions, Hong Kong is for sure one of the largest, if not the largest, global centers for offshore RMB businesses. According to the Triennial Central Bank Survey (BIS 2019), for instance, Hong Kong was the largest global offshore RMB foreign exchange market, with an average daily turnover of US\$107.6 billion as of April 2019, considerably higher than the US\$56.7 billion for London and the US\$42.6 billion for Singapore.

For the rest of the world, the OECD countries in particular, Hong Kong not only acts as a springboard for foreign firms' investment and expansion into the Chinese markets, but also benefits foreign countries given Hong Kong's special status under the "one country, two systems" principle of the Basic Law. For instance, the United States has persistently run a merchandise trade surplus with Hong Kong (US\$11.7 billion in 2019), and American financial institutions earn lucrative profits through their investment banking activities and other financial services in Hong Kong.²

Even before the return of Hong Kong to Chinese sovereignty in 1997, all these stakeholders have had economic incentives to maintain and promote Hong Kong's status as an international financial center. In recent years, however, this situation has become more difficult to sustain because of political and ideological divergence and conflicts partly arising from China's promotion of national revival, more generally known as the "Chinese Dream," since 2013.

China's Drive to Create a Global Financial Center

There should be no doubt that China recognizes the importance of an international financial center in fulfilling the Chinese Dream: the nation needs money and finance to accomplish its series of grand projects like Made in China 2025, internationalization of the RMB, the Belt and Road Initiative, as well as development of the Guangdong–Hong Kong–Macau Greater Bay Area. Following economic reform in 1978, the Chinese leaders put forward several

²This does not imply that a trade surplus is always beneficial to the economy as a whole. From a political economy perspective, it is just an example of some foreigners—in this case, U.S. exporters—gaining from the "one-country, two systems" arrangement.

strategic economic plans to develop Shanghai, Shenzhen, and Qianhai into international financial centers.³ Moreover, in December 2019, about six months after the outbreak of anti-extradition bill protests in Hong Kong, there was news, though officially unconfirmed, that President Xi Jinping endorsed the diversification of Macau into a global center for tourism and finance—probably as a contingency plan to replace Hong Kong if the social unrest in Hong Kong deteriorated. In June 2020, China’s State Council released an ambitious master plan to transform Hainan into its largest free-trade port, including liberalization of the financial sector and free exchange of capital, by 2035 (Wong 2020).

However, things are easier said than done; none of these strategic plans have materialized. Undeniably, there are more stock markets and financial institutions than before, but all these Chinese cities are still far from being an international financial center comparable to Hong Kong. The Chinese government seemingly had the confidence that it could design and build up a financial center, or even financial centers, through economic planning. Yet nearly a century ago, in the socialist calculation debate, Mises and Hayek had already correctly pointed out that central economic planning was both erroneous and infeasible in theory and practice.

The Infeasibility of Central Planning

In a planned economy, the central planner will inevitably fail to coordinate production plans and allocate resources efficiently, because of at least two major problems.⁴ First, in socialist or centrally

³For instance, in 1992, the 14th Communist Party of China National Congress proposed to establish Shanghai as an international economic, financial, and trade center. Later, in China’s 11th Five-Year Economic Plan (2006–2010), Shanghai was officially handpicked by the State Council of China to take up the strategic role of developing into an international financial center. There were similar plans and endorsements for Shenzhen and Qianhai in subsequent years, notably the economic plan for the latter announced in 2010, which aimed at developing Qianhai to become the Manhattan of China.

⁴Needless to say, there are other relevant issues like entrepreneurship, property rights, and incentive mechanism. However, the literature on the socialist calculation debate is so voluminous that I am forced to limit my attention here to the two major problems initially raised by Mises and Hayek—namely, the Misesian calculation problem and Hayekian knowledge problem. Admittedly, it may not be necessary to distinguish between these two problems as they are inherently intertwined (see Yeager 1994).

planned economies, goods and services are transferred rather than exchanged via markets; hence, they are either unpriced or irrationally priced. Economic calculation is therefore infeasible in the absence of a market price system where relative prices play both a signaling and rationing function to promote efficient use of resources (Mises [1920] 1990).

Second, economic calculation cannot be made by simulation of markets under market socialism or central planning because of a knowledge problem (Hayek 1937, 1940, 1945). According to Hayek's pathbreaking insight, the knowledge relevant to rational economic calculation exists as dispersed bits of incomplete knowledge possessed by millions of separate individuals rather than in an integrated form given to the central planner. Moreover, individuals may have no or little incentive to convey their knowledge to the central planner. This incentive problem, together with the prohibitively high communication and search costs associated with the available dispersed knowledge at a given moment, implies that it is virtually impossible to make the most effective use of the available dispersed information for efficiently allocating society's scarce resources under central planning. This Hayekian knowledge problem, as Kirzner (1984) argues, cannot become solvable by any planned search for the necessary information, because the planner has only limited knowledge and is likely unaware of his own ignorance. Nevertheless, the knowledge problem can be overcome by the alertness of private firms and entrepreneurs to profit opportunities through the competitive discovery process in a decentralized market economy (Hayek 1978a; Kirzner 1984).

Hayek followed the Austrian tradition of Carl Menger and held that money, like language and law, is one of the complex social institutions whose evolutionary outcomes are the result of human action, not deliberate design (see, e.g., Hayek 1960: chaps. 2 and 4; 1967a; 1973: chaps. 1–2; 1978b: 37–39; and 1988). Moreover, the evolutionary outcomes are inherently path dependent. This notion of society's spontaneous order is equally applicable to the development or evolution of monetary and banking systems.⁵

⁵ See, for example, an earlier study by O'Driscoll (1994), although it addresses the case of money and banking in general instead of the specific case of an international financial center.

Hong Kong's Financial Development as a Spontaneous Order

Hong Kong's rise as an international financial center is an example of a spontaneous order that emerged given limited government and the rule of law. During the 1970s and 1980s, when Hong Kong went through an economic transformation from a manufacturing city into a financial center, the Hong Kong government did not have any economic plan to deliberately turn the city into an international financial center. Hong Kong had, however, certain favorable factors, like its geographical location, a meeting point of East and West, free trade, no exchange control, freedom of press, low tax rates, a credible legal system, a corruption-free government, and its doctrine of "positive non-interventionism."⁶ At the same time, there were the trends of globalization and China's economic reform and opening up its door to the rest of the world.⁷

All these factors acted together, at the right place and at the right time, to help establish Hong Kong as a vibrant global financial center. International investors, entrepreneurs, and financial institutions reacted to the changing market forces and gained from the business opportunities. Metaphorically, it can be said that the invisible hand turned Hong Kong into an international financial center. The role of the Hong Kong government was mainly to provide a stable and fair business environment, and at most it only provided some measures or policies to nudge or facilitate the process of financial development. Even the celebrated linked exchange-rate system, which has provided monetary stability for decades so far, was a response to the political and financial crisis triggered by the uncertainty surrounding the Sino-British negotiations on Hong Kong's future after 1997. It was not deliberately designed and introduced by the Hong Kong government, as reflected by the fact that even the official exchange rate of US\$1=HK\$7.8 was chosen arbitrarily as a rough average of the market exchange rate before and after the crisis rather than an "optimal" rate based on an economic plan or model.

⁶On the doctrine of "positive non-interventionism," see Hanson (2012). Sir John Cowperthwaite, Hong Kong's financial secretary from 1961 to 1971, was the architect of this policy, which can best be stated as "small government, big market."

⁷For details about the rise of Hong Kong as an international financial center, see Jao (1997).

Why China Still Lacks a Global Financial Center

The Chinese government seems to have overlooked Hayek's invaluable insights and Hong Kong's experience.⁸ Against the backdrop of China's rapid surge to become the world's second largest economy and also rising nationalism triggered by the Chinese Dream, the Chinese government appears to have dogmatically believed that it can establish an international financial center on its own through economic planning. This belief was partly reflected by the above-mentioned strategic plans and partly by the editor-in-chief of the *Global Times*, who, in rebutting U.S. Secretary of State Mike Pompeo's criticism of China's National Security Law for Hong Kong, declared: "China's strength determines that there will inevitably be an international financial center on its coastline. It will be wherever China wants it to be, actually" (Hu 2020).

Furthermore, the Chinese government appears to assume that it can simply take over Hong Kong's financial infrastructure and then deploy its own comrades and patriotic financiers to maintain the operation of an international financial center. This is partly reflected by China's increasing participation in Hong Kong's financial markets and also by its tightening grip on Hong Kong.

A Fatal Conceit

China's dream of having a competitive global financial market without creating the institutional infrastructure needed for such a system is what Hayek would call a "fatal conceit." It is a profound and mistaken belief that "man is able to shape the world around him according to his wishes" (Hayek 1988: 27). Without economic freedom, protection of private property rights, and freedom of the press, there is little chance that Beijing can create a credible international financial center that people can trust.

Rule of Law and Freedom of Information

The rise and fall of financial centers can be attributable to many causal factors like war, politics, ideology, infrastructure, technology,

⁸While it may be argued that Hayek's ideas can be overlooked because they originated from the West, it is interesting to note that the idea of spontaneous order can be traced to Taoism in ancient China. Lao Tzu recognized that, if the ruler (sage) left people alone, they would be "transformed of themselves" and "of themselves become rich" (*Tao Te Ching*, §57). See Dorn (1998).

human capital, tax system, and regulatory regime, to name just a few. Two other crucial factors—namely, the rule of law and freedom of information—are highly relevant and merit further analysis here for at least a couple of reasons. First of all, Hong Kong's considerable comparative advantage over China in financial matters can be attributable to many factors, but it is a marked contrast between Hong Kong and China in these two areas.⁹ It is a matter of grave concern that the gradual erosion of the rule of law and freedom of press in Hong Kong in recent years will jeopardize its status as an international financial center. Furthermore, these two crucial factors are important because banking and finance, in both theory and practice, involve structuring, administering, and enforcing financial contracts to overcome the problems arising from imperfect information.

Financial intermediaries have a comparative cost advantage in producing information and monitoring financial transactions to mitigate or to resolve the adverse-selection and moral-hazard problems arising from information asymmetries (see Leland and Pyle 1977; Campbell and Kracaw 1980; and Diamond 1984, to name just a few). Information, both *ex ante* and *ex post*, can have social value if it enables an expansion in the set of enforceable risk-sharing arrangements like tradable securities or state-contingent claim contracts. As an example of the benefit of *ex ante* information, good banks in Hong Kong during the relatively free banking regime of the 1960s used lower deposit rates and higher reserve ratios as signals to depositors about their underlying quality so as to separate themselves from bad banks (Chu 1999). For the benefit of *ex post* information, all parties to a financial contract have to be able to rely on the public *ex post* information so as to observe and validate the actual state occurrence, otherwise they would not be able to enforce the existing contract or to negotiate a new contract. It is well known that the asymmetric information problems can cause markets to break down or even disappear (Akerlof 1970). Therefore, banking and finance is essentially an information industry and no city can hope to become an international financial center if there are serious barriers to the dissemination of news and information.

Financial centers, like money and merchants, emerge as intermediaries of trade and facilitate the expansion of exchange opportunities

⁹These two factors are crucial not only in finance. As Dorn (2019) convincingly argues, the absence of a free market for ideas—because of strict controls of the free flow of information and the lack of limited government under a genuine rule of law—endangers China's future development.

beyond autarky, leading to what Hayek (1973) calls “the extended order of human cooperation.” But how do borrowers, lenders, and financial institutions establish trust? Is a borrower’s or financial institution’s promise to pay on a certain date reliable? How can lenders or investors be sure that their securities are not “lemons” (Akerlof 1970)? These are just some of the problems that have to be overcome in order to achieve this Hayekian state. To a large extent, freedom of information alleviates the asymmetric information problems, but freedom of information alone is insufficient for the trading of financial contracts and also the rise of an international financial center. This is because, as J. R. Hicks (1969: 34) noted, “Trading is trading in promises; but it is futile to exchange promises unless there is some reasonable assurance that the promises will be kept.” Therefore, in order to protect private ownership rights and to punish dishonest traders, trading governance mechanisms, such as a sound rule of law, are also needed for the rise and sustainability of financial centers over time. Besides the rule of law, third-party information dissemination regarding dishonest traders is another form of a trading governance mechanism commonly used in commercial transactions. Yet this mechanism would not be fully effective, or even become infeasible, in the absence of freedom of information.¹⁰

Empirical Evidence

A main objective of this article is to examine the influence of the rule of law and freedom of information on international financial centers. The evolution of financial centers is a complex phenomenon. So instead of resorting to sophisticated econometrics, I look into the available data to identify if there exists a pattern as predicted by theory and, if so, based on the identified pattern, explain the principle behind the complex phenomenon to be explained (Hayek 1967b, 1967c).

From Z/Yen Group’s reports of the Global Financial Centres Index since 2008, the top 15 international financial centers by the five areas of competitiveness as well as the leading financial centers in the major regions are selected.¹¹ Some countries have more than one international financial center, but we count only one of them for

¹⁰ For a variety of trade governance mechanisms, see Aoki (2001: 62–91).

¹¹ The five areas of competitiveness are business environment, human capital, infrastructure, financial sector development, and reputation, whereas the major regions are Asia Pacific, Central Europe and Central Asia, Europe, Latin America, Middle East and Africa, and North America.

each country. For the United States, for example, New York is included but other financial centers like San Francisco and Chicago are omitted. A reason for so doing is that the other data used in this empirical analysis are mainly based on countries rather than on cities. Nonetheless, I obtain a quite representative sample of 54 major international financial centers in the world, ranging from global financial centers like London and New York to regional financial centers like Sydney and Johannesburg. More importantly, these international financial centers are functional centers rather than paper centers or booking centers like Bahamas and the Cayman Islands.¹²

For each selected international financial center, I use estimates of Rule of Law and Voice and Accountability from the World Bank's Worldwide Governance Indicators Database.¹³ The Rule of Law estimate for each financial center captures perceptions of the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence. Meanwhile, the Voice and Accountability estimate is used as a proxy for freedom of information because it captures perceptions of the extent to which a country's citizens are able to participate in selecting their government as well as freedom of expression, freedom of association, and a free media. Each estimate gives a country's score ranging from minus 2.5 to 2.5. A higher positive score denotes stricter adherence to the rule of law (or freedom of information), while a more negative score suggests a stronger violation of the rules or underlying principles or beliefs. Although the scores for each selected country in our sample do not change considerably over time, the average scores for these two estimates over the years 2008 to 2019 are taken as the indicators or indexes in this cross-sectional analysis.

If both the rule of law and freedom of information are not determinants in the evolution of international financial centers, a scatter plot of international financial centers with these two factors would reveal

¹²This distinction is due to McCarthy (1979), according to which a paper center acts as a location for recording financial transactions only, with little or no actual banking or financial business being carried out there, whereas a functional center provides and executes financial services and transactions of all kinds.

¹³By convention, the terminology in this empirical analysis follows from the World Bank Databases, in which the term "countries" (or "country") can refer to sovereign states or to other political entities such as Hong Kong, which is not a country but a special administration region of China.

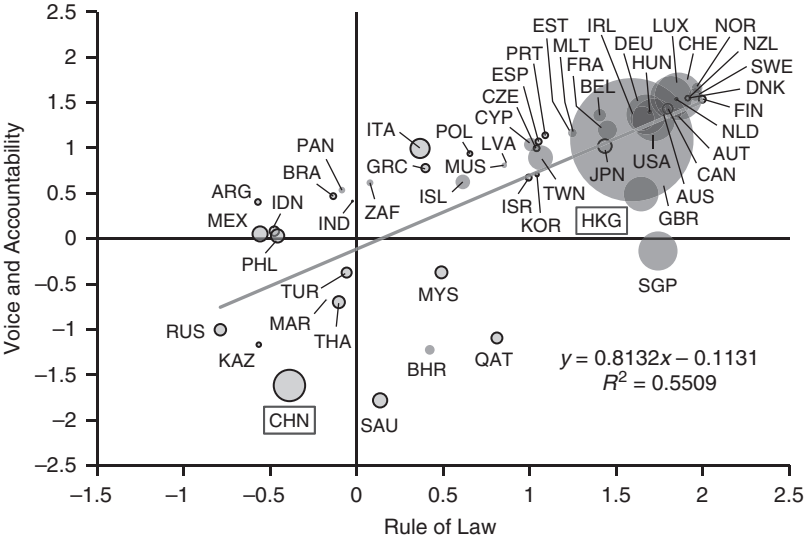
no systematic pattern and international financial centers are more or less randomly distributed regardless of the extent of the rule of law or freedom of information. If only one of them, say, the rule of law, influences the evolution of international financial centers, then there would be an expected systematic pattern such as international financial centers being more likely found in countries that rank high in the rule of law index. Under our maintained hypothesis, both factors are conducive to the evolution of international financial centers, and therefore it is expected that international financial centers are more commonly found in countries that rank high in terms of both the rule of law and freedom of information indexes. Diagrammatically, international financial centers are expected to cluster in the region of the scatter plot where both indexes are high in value.

As shown in the bubble chart (Figure 1), the empirical evidence unambiguously supports our hypothesis. There is a positive correlation between the rule of law and freedom of information indexes, as evidenced by the cross-sectional regression.¹⁴ At the same time, a big cluster of international financial centers is found in the first quadrant (top right) of the diagram, although some international financial centers deviate away from this cluster because of other political and economic factors. These findings indicate that both the rule of law and freedom of information are conducive factors for international financial centers. Furthermore, the cluster shows that it not only contains more international financial centers but also that they are larger in size as indicated by the third dimension of the chart: size of bubbles. Each bubble measures the volume of net exports of insurance and financial services of a country, the data of which are obtained from the World Bank's World Development Indicators Database.¹⁵ The larger the bubble, the higher is the average value of net exports of insurance and financial services over the period under study. As expected, the

¹⁴The cross-sectional regression results show a high R^2 of 0.55 and a statistically significant slope coefficient of 0.81.

¹⁵The database contains data series on insurance and financial services as percentage of commercial service exports and its counterpart as percentage of commercial service imports. These percentages are then converted into values by multiplying the respective commercial service exports and imports (current US dollars). Net exports or imports are then computed as the differences between the two value series. For Taiwan, the data on exports and imports of insurance and financial services are obtained directly from the Central Bank of the Republic of China (Taiwan). Again, the average values over the years 2008–2019 are used in the empirical analysis.

FIGURE 1
INTERNATIONAL FINANCIAL CENTERS
INFLUENCE OF RULE OF LAW AND
FREEDOM OF INFORMATION



- NOTES:
1. Each bubble represents the value (US\$ million) of net exports of insurance and financial services.
 2. A negative bubble is a bubble with a black perimeter and represents the value (US\$ million) of net imports of insurance and financial services.
 3. The larger the size of a bubble (or a negative bubble), the higher the value net exports (or imports) of insurance and financial services.

- SOURCES:
1. Data on Rule of Law and Voice and Accountability estimates are from World Bank (2020a).
 2. Data on imports and exports of insurance and financial services are from World Bank (2020b), except Taiwan.
 3. Data on Taiwan's imports and exports of insurance and financial services are from the Central Bank of the Republic of China (2020).

largest bubbles are found in the advanced industrial countries with the global financial centers, and they are, in order, the United Kingdom (London), Switzerland (Zurich), the United States (New York), Luxembourg, and Germany (Frankfurt), followed by a close tie between two small open economies—Singapore and Hong Kong.

As a matter of fact, these bubbles in the cluster are so large, particularly the one of the United Kingdom, that they bury many countries' bubbles clustered around them and we can see only the country labels in the diagram. Fortunately, Hong Kong's bubble (labeled as HKG in Figure 1) is big enough that we can still see part of it beneath the largest bubble of the United Kingdom.

Conversely, when a country registered net imports of insurance and financial services, it is represented by a negative bubble (with black perimeter) in Figure 1. As can be seen, negative bubbles are more commonly found in countries with low scores of both the rule of law and freedom of information, although there are exceptions. Similarly, the larger the negative bubble, the higher is the value of net imports of insurance and financial services. The largest negative bubbles are, respectively, China (labeled as CHN in Figure 1), Italy, Mexico and Saudi Arabia. Except Italy, these countries, notably China, are among the countries with the lowest scores in both the rule of law and freedom of information. We should not erroneously confuse China's large net imports of insurance and financial services with its huge capital imports during this period, although they can be correlated because, say, foreign investors may have a home bias in favor of dealing with their own countries' transnational banks instead of Chinese banks.

Whatever the true reasons for China's large net imports of insurance and financial services—and despite the gradual increase in Shanghai's ranking, because of China's rapid economic growth and businessmen's optimism about the Chinese markets—the trade statistics reveal that Shanghai has not yet established itself as a global financial center. A financial center is essentially a collection of financial intermediaries that provide a wide range of financial services to their clients. In a closed economy, these financial services, though they have value added, do not appear in the trade balance. In an open economy, however, an international financial center provides financial services not only domestically but also regionally and internationally. China's large net imports of insurance and financial services imply an under-supply of financial services in the domestic financial sector in China and they imply that Shanghai as an international financial center is not comparable to London and New York. The former remains the champion of global net exporter of financial services despite the post-war relative decline of the British domestic economy whereas the latter continues to thrive regardless of the perennial U.S. trade deficit.

From this perspective, Shanghai is not even comparable to Singapore and Hong Kong, two leading net exporters of financial services in the world despite their small open economies with virtually no natural resources. Indeed, China's financial reform in the last four decades has built a large financial sector by establishing large numbers of various types of financial institutions and growing financial assets, but at the expense of efficiency and corporate governance. The negative effects of its repressive financial policies have already begun to take a toll on its economic and financial performance in recent years, and further market-oriented financial reforms are urgently needed to sustain growth and stability in the foreseeable future (Huang and Ge 2019).

Rome was not built in a day. It took decades for Hong Kong to evolve from a subregional financial center to a regional center and ultimately to an international, if not global, financial center. This lengthy evolution process is also found in the emergence and rise of other financial centers (Table 2). There is simply no clear evidence indicating that a financial center can leapfrog ahead of its rival into the top position of a global financial center. But when Rome fell, it burned in one day. Shanghai was the premiere financial center of the Far East before World War II, but it lost not only its glory but also its status after the civil war in China. Other examples abound. Florence is reputed to have been the world's first international financial center in the 13th century, but it was probably also the first one ruined by sovereign or country risk when Edward III, King of England, repudiated his debt a century later. War, politics, and ideology, as history clearly shows, can lead to the rapid fall of financial centers.

Hong Kong's Future

Collectivism and Nationalism

Under collectivism and nationalism, China's rubber-stamp parliament, the National People's Congress (NPC), enacted the National Security Law for Hong Kong on June 30, 2020, without first releasing the full draft of the law to Hong Kong's 7.5 million citizens, let alone having any public consultation. Despite Beijing officials' claim that the law will only strengthen the "one country, two systems" framework, the law is highly contentious and has been strongly criticized by free, democratic nations. The law is generally perceived to further undermine, if not end, Hong Kong's autonomy under the "one country, two systems" principle and also, from a Hayekian perspective, to bring

TABLE 2
FINANCIAL CENTERS IN THE FAR EAST AND AUSTRALASIA

	Subregional	Regional	Global
1919–39	Hong Kong, Singapore	Shanghai	
1945–65	Hong Kong, Singapore		
1965–75		Hong Kong, Singapore, Tokyo	
1975–2000		Hong Kong, Singapore, Sydney	Tokyo
2000–present	Bangkok, Jakarta, Kuala Lumpur, Manila, Seoul, Shanghai, Taipei, Wellington	Sydney	Hong Kong, Singapore, Tokyo

NOTES:

1. The definitions of financial centers follow Jones (1992):
 - (i) Subregional center is one that focuses on international financial activities based on bilateral trade between the center's host economy and the rest of the world;
 - (ii) Regional center is one that supplies financial services to a supranational region; and
 - (iii) Global center is one that supplies a broad range of financial services to the whole world.
2. The entries for 1975–2000 and earlier are mainly based on Jones (1992) with some modifications and corrections:
 - (i) Beirut and Bahrain in Jones (1992) are deleted because they belong to the Middle East;
 - (ii) Shanghai is changed to a regional center during 1919–39 instead of a subregional center as in Jones (1992);
 - (iii) Tokyo is included rather than omitted by Jones (1992).
 - (iv) The years covered by Jones (1992) are extended to 2000.
3. The entries for 2000–present are updates by this author based on the empirical findings of this study. While there may not be a consensus on this classification, a main objective of this table is to show the stages of evolution of financial centers over time.

Hong Kong another step further toward the road to serfdom. That is because Chinese nationalism, like socialism and Germany's National Socialism, champions government control of politico-economic decisionmaking and empowers the state over the individual, which inevitably leads to a loss of individual freedoms (Hayek 1944).

After being rated the freest economy in the world for a quarter of a century, Hong Kong slipped and lost its standing to Singapore (Heritage Foundation 2020). Meanwhile, Hong Kong is no longer the third largest financial center in the world, after New York and London. In March 2020, it fell to sixth place (Z/Yen Group 2020). In addition, Fitch and Moody's have downgraded Hong Kong's credit rating twice since September 2019. Although we should not be overly concerned about these downgrades in the short run, they do point to the difficulties and challenges facing Hong Kong in maintaining its status as an international financial center. The lower ratings are commonly attributed to the ongoing protests in Hong Kong triggered by the Hong Kong government's abortive attempt to introduce the extradition bill. The key issue is whether Hong Kong can continue to maintain its autonomy under "one country, two systems."

Fitch expressed openly that one of the reasons for the downgrade of Hong Kong's credit rating was its increasing linkages to mainland China. Although another credit rating agency, Standard and Poor's, did not downgrade Hong Kong's credit rating over the last couple of years, it has already stated quite openly that it would do so if Hong Kong failed to maintain its autonomy under "one country, two systems."

As far as financial development is concerned, the National Security Law for Hong Kong has also increased concern over the future of Hong Kong as an international financial center. Indeed, Hong Kong will lose its current special status when it is deemed by other countries to have lost its autonomy and become part of China. President Trump's Executive Order on Hong Kong Normalization (White House 2020) has already revoked Hong Kong's preferential status of being treated separately from China, in accordance with the United States–Hong Kong Policy Act of 1992. The newly passed Hong Kong Autonomy Act (Gilroy, Owens, and Tovar 2020) can potentially hamper Hong Kong's status as an international financial center because it empowers the U.S. president to impose sanctions on financial institutions that do business with entities that undermine

Hong Kong's autonomy under "one country, two systems." Even though no provision related to Hong Kong's currency board system has been made yet, any departure from the link to the U.S. dollar would no doubt be detrimental to Hong Kong's status as a global financial center.

Conclusion

It is questionable if the National Security Law for Hong Kong can actually promote national security or economic and social stability, particularly given the growing tensions in the Sino-U.S. relations following the outbreak of the Covid-19 pandemic. It is also questionable whether the law is really in the interests of both China and Hong Kong. Without Hong Kong's special status as an international financial center, as well as a window to absorb foreign investment and new technologies for China, it is unlikely China can continue to experience robust growth. Empirical evidence from numerous cross-country studies has clearly indicated that such factors as rule of law, civil liberties, capitalism, investment, and openness to trade are good for economic growth, whereas factors like wars and market distortions are bad. Without the good factors, nationalism alone cannot make the Chinese Dream a reality.

References

- Akerlof, G. (1970) "The Market for Lemons: Quantitative Uncertainty and the Market Mechanism." *Quarterly Journal of Economics* 84: 488–500.
- Aoki, M. (2001) *Toward a Comparative Institutional Analysis*. Cambridge, Mass.: MIT Press.
- Bank for International Settlement (BIS) (2019) *BIS Quarterly Review* (December). Basel, Switzerland: BIS.
- Campbell, T., and Kracaw, W. A. (1980) "Information Production, Market Signalling, and the Theory of Financial Intermediation." *Journal of Finance* 35: 863–82.
- Census and Statistics Department, Hong Kong SAR (2020) *Hong Kong Monthly Digest of Statistics* (June).
- Central Bank of the Republic of China (2020) *National Statistics, Republic of China*. Available at <https://eng.stat.gov.tw/mp.asp?mp=5>.

- Chu, K. H. (1999) "Free Banking and Information Asymmetry." *Journal of Money, Credit and Banking* 31: 748–62.
- Diamond, D. (1984) "Financial Intermediation and Delegated Monitoring." *Review of Economic Studies* 51: 393–414.
- Dorn, J. A. (1998) "China's Future: Market Socialism or Market Taoism?" *Cato Journal* 18 (1): 131–46.
- (2019) "China's Future Development: Challenges and Opportunities." *Cato Journal* 39: 173–88.
- Gilroy, T.; Owens, I.; and Tovar, A. (2020) "President Trump Signs into Law the Hong Kong Autonomy Act." *SanctionsNews*, BakerMcKenzie blog (July 14). Available at <https://sanctionsnews.bakermckenzie.com/president-trump-signs-into-law-the-hong-kong-autonomy-act>.
- Hanson, D. (2012) "Positive Non-Interventionism: The Policy that Unleashed Hong Kong." *AEIdeas* (March 22).
- Hayek, F. A. (1937) "Economics and Knowledge." *Economica* 4 (13): 33–54.
- (1940) "Socialist Calculation: The Competitive 'Solution.'" *Economica* 7 (26): 125–49.
- (1944) *The Road to Serfdom*. Chicago: University of Chicago Press.
- (1945) "The Use of Knowledge in Society." *American Economic Review* 35 (4): 519–50.
- (1960) *The Constitution of Liberty*. Chicago: University of Chicago Press.
- (1967a) "The Result of Human Action but Not of Human Design." In *Studies in Philosophy, Politics and Economics*, 96–105. Chicago: University of Chicago Press.
- ([1955] 1967b) "Degrees of Explanation." In *Studies in Philosophy, Politics, and Economics*, 3–21. Chicago: University of Chicago Press.
- ([1964] 1967c) "The Theory of Complex Phenomena." In *Studies in Philosophy, Politics, and Economics*, 22–42. Chicago: University of Chicago Press.
- (1973) *Law, Legislation and Liberty. Volume I: Rules and Order*. London: Routledge.
- (1978a) "Competition as a Discovery Procedure." In *New Studies in Philosophy, Politics, Economics and the History of Ideas*, 179–90. Chicago: University of Chicago Press.

- _____ (1978b) *Denationalization of Money: The Argument Refined*. London: Institute of Economic Affairs.
- _____ (1988) *The Fatal Conceit: The Errors of Socialism*. Chicago: University of Chicago Press.
- Heritage Foundation (2020) Miller, T.; Kim, A. B.; and Roberts, J. M. (eds.), *2020 Index of Economic Freedom*. Washington: Heritage Foundation.
- Hicks, J. R. (1969) *A Theory of Economic History*. Oxford: Oxford University Press.
- Hu, X. (2020) “Pompeo Makes Arrogant, Hysterical Statements about HK Autonomy.” *Global Times* (May 28).
- Huang, Y., and Ge, T. (2019) “Assessing China’s Financial Reform: Changing Roles of the Repressive Financial Policies.” *Cato Journal* 39: 65–85.
- Jao, Y. C. (1997) *Hong Kong as an International Financial Centre: Evolution, Prospects, and Policies*. Hong Kong: City University of Hong Kong Press.
- Jones, G. (1992) “International Financial Centers in Asia, the Middle East, and Australia: A Historical Perspective.” In Y. Cassis (ed.), *Finance and Financiers in European History, 1880–1960*, 405–28. Cambridge: Cambridge University Press.
- Kirzner, I. (1984) “Economic Planning and the Knowledge Problem.” *Cato Journal* 4 (2): 407–18.
- Leland, H. E., and Pyle, D. H. (1977) “Information Asymmetries Financial Structure, and Financial Intermediation.” *Journal of Finance* 32: 371–81.
- McCarthy, I. (1979) “Offshore Banking Centers: Benefits and Costs.” *Finance and Development* 16: 45–48.
- Ministry of Commerce, People’s Republic of China (2020a) *Statistical Bulletin of FDI in China 2019*.
- _____ (2020b) *News Release of National Assimilation of FDI*, various issues. Available at <http://mofcom.gov.cn>.
- Mises, L. V. ([1920] 1990) *Economic Calculation in the Socialist Commonwealth*. Auburn: Ludwig von Mises Institute.
- O’Driscoll, Jr., G. P. (1994) “An Evolutionary Approach to Banking and Money.” In J. Birner and R. Zijp (eds.), *Hayek, Co-ordination and Evolution: His Legacy in Philosophy, Politics, Economics, and the History of Ideas*, 126–40. London: Routledge.

- White House (2020) “The President’s Executive Order on Hong Kong Normalization” (July 14). Available at <https://www.whitehouse.gov/presidential-actions/presidents-executive-order-hong-kong-normalization>.
- Wong, D. (2020) “Hainan FTZ Masterplan Released to Establish China’s Biggest Free Trade Port by 2035.” *China Briefing*, Dezan Shira & Associates (June 5).
- World Bank (2020a) *Worldwide Governance Indicators Database*. Washington: World Bank. Available at <https://databank.worldbank.org/source/worldwide-governance-indicators>.
- _____ (2020b) *World Development Indicators Database*. Washington: World Bank. Available at <https://databank.worldbank.org/source/world-development-indicators>.
- Yeager, L. B. (1994) “Mises and Hayek on Calculation and Knowledge.” *Review of Austrian Economics* 7 (2): 93–109.
- Z/Yen Group (2020) *Global Financial Centres Index 27*, and also reports in previous years since 2008. London: Long Finance.

U.S. TRADE POLICY TOWARD CHINA: LEARNING THE RIGHT LESSONS

Scott Lincicome

Labor market and cultural disruptions in the United States are real and important, as is China's current and unfortunate turn toward illiberalism and empire. But pretending today that there was a better trade policy choice in 2000—when Congress granted China “permanent normal trade relations” (PNTR) status and paved the way for broader engagement—is misguided. It assumes too much, ignores too much, and demands too much. Worse, it could lead to truly bad governance: increasing U.S. protectionism; forgiving the real and important failures of our policymakers, CEOs, and unions over the last two decades; and preventing a political consensus for real policy solutions. Indeed, that is happening now.

China entered the World Trade Organization (WTO) on December 11, 2001, and President George W. Bush signed a proclamation on December 27, that extended PNTR status to the People's Republic of China (White House 2001). The benefits of granting PNTR to China were widely understood at the time. Nicholas Lardy, a respected China scholar, summarized those benefits as follows:

A positive vote [for PNTR with China] would give U.S. companies the same advantages that would accrue to companies from Europe, Japan, and all other WTO member states when

Cato Journal, Vol. 41, No. 1 (Winter 2021). Copyright © Cato Institute. All rights reserved. DOI:10.36009/CJ.41.1.3.

Scott Lincicome is a Senior Fellow in Economic Studies at the Cato Institute and a Senior Visiting Lecturer at Duke University Law School. He thanks Erin Partin for her research assistance. This article is based on the author's Cato Policy Analysis No. 895 (Lincicome 2020).

China enters the World Trade Organization. It would also provide an important boost to China's leadership, that is taking significant economic and political risks in order to meet the demands of the international community for substantial additional economic reforms as a condition for its WTO membership. A positive vote would strengthen bilateral economic relations more generally [Lardy 2000: 1].

The 20th anniversary of the U.S. law implementing PNTR with China and the Chinese government's numerous recent offenses have caused a bipartisan chorus of American politicians and pundits to question the wisdom of that law. In particular, these critics allege that the Clinton administration and Congress rubberstamped both the law and China's entry into the WTO, and that these events fueled China's rise and the now-famous "China Shock"—the period between 1999 and 2011 during which a sizable increase in Chinese imports supposedly produced the loss of approximately 2.4 million U.S. jobs. In turn, critics have used the "mistake" of past economic engagement with China to justify grand rethinks of current U.S. foreign and economic policy, including withdrawal from the WTO itself.

However, a proper accounting of the relevant literature reveals most of this criticism to be mistaken on the law, economics, and history of PNTR, China's WTO accession, and the China Shock. This analysis instead reveals that new or continued U.S. restrictions on Chinese imports would not have saved most of the U.S. manufacturing jobs destroyed between 1999 and 2011; that China would have joined the WTO and become an export powerhouse regardless of PNTR; that U.S. engagement with China in the 1990s was a gradual and pragmatic policy decision based on numerous supporting facts at the time, and with no better alternatives; and that myriad U.S. policy failures since PNTR actually did enable China to harm American companies and workers. Policymakers should focus on these errors, not PNTR, lest they risk enacting new policies that do nothing to address America's real and serious challenges, including China, and might actually make things worse.

U.S.-China Trade: Amplified Costs and Ignored Benefits

A central flaw in the anti-PNTR thesis is that it ignores the benefits of increased U.S. trade with China over the last two decades—for American consumers, manufacturers, and workers. Economists have

found, for example, that Chinese import competition between 2000 and 2007—the peak of the China Shock—had substantial procompetitive effects on U.S. firms and generated over \$202 billion in consumer benefits via lower prices. That equals \$101,250 in benefits to U.S. consumers per manufacturing job lost (Jaravel and Sager 2018). Another study concludes that Chinese imports “significantly reduced inflation,” cutting the price index for consumer goods by 0.19 percentage points per year between 2004 and 2015 through lower prices and an increased variety of goods available (Bai and Stumpner 2019; see also Amiti et al. 2018). Consumer benefits of trade, already tilted toward America’s poor and middle class, were even more so for Chinese imports because those consumers frequently shop at places like Target and Walmart that carry such goods (Broda and Romalis 2008; Fajgelbaum and Khandelwal 2014). One can argue that those consumer benefits are cold comfort to someone who lost a job because of Chinese import competition, but they are nevertheless real, widespread, and important.

Chinese imports have also been found to generate substantial benefits for many American companies and their workers. Import competition encouraged many American manufacturing firms to invest and innovate more and ultimately led to net welfare benefits (Gutiérrez and Philippon 2017; Caliendo, Dvorkin, and Parro 2019). After accounting for manufacturing supply chains and intermediate inputs, the effect of the China Shock on American jobs and wages has been quite positive overall, and while the China Shock produced losses for certain groups of Americans, it generated overall gains in social welfare (Wang et al. 2018; Galle, Rodriguez-Clare, and Yi 2017). In fact, researchers have estimated that about 56 cents of every dollar that Americans spent on “Made in China” imports in 2019 actually went to American firms and workers (Hale et al. 2019). Such benefits make sense: data show millions more “blue collar” American jobs that might benefit from Chinese imports—in transportation, logistics, construction, maintenance, and repair, for example—than in manufacturing (BLS 2018).

This assessment erroneously assumes, moreover, that all U.S. manufacturing jobs are potentially hurt by Chinese import competition. Yet approximately one-third of all Chinese imports were intermediate goods used by American companies to produce globally competitive products. These imports have helped, not hurt, U.S. manufacturing workers. In fact, U.S. manufacturing firms

that increased direct imports from China between 1997 and 2007 experienced growing or steady employment, likely because of the importers' ability to lower prices and raise output (even as nonimporting competitors suffered) (Antràs, Fort, and Tintelnot 2017). With respect to these types of complex value chains, the WTO estimates that China in 2015 was the third-largest user—behind only Mexico and Canada—of “Made-In-America” manufacturing inputs and the largest source of inputs for American manufacturers (WTO 2020).

Then there are the benefits that American farmers and workers have derived from exporting to China, still the United States' third largest export destination (U.S. Census Bureau 2020). According to the U.S.-China Business Council, exports to China in 2019 supported over 1.1 million American jobs in a wide range of manufacturing, logistics, and services industries (USCBC 2019). The U.S.-China “Phase One” agreement's heavy emphasis on American agriculture sales, as well as the massive expansion of federal farm subsidies during the countries' trade dispute, shows just how much U.S. farmers have benefited from selling to the growing Chinese market.

Beyond the benefits of trade with China, a proper accounting of the China Shock also requires proper context. There is evidence, for example, that many U.S. manufacturers adapted during the shock, and ended up *hiring* many Americans and *increasing* output. For example, Fort, Pierce, and Schott (2018: 18–21) find that declines in “manufacturing firm workers” employed in “manufacturing plants” between 1977 and 2012 were more than offset by contemporaneous increases in employees in “non-manufacturing plants” that were owned by many of the same “manufacturing firms.”

The evolution of American manufacturing—driven by trade, automation, or other factors—raises further concerns about attempting to isolate the effects of Chinese import competition on low-skill American manufacturing employment. For example, researchers find that the decline in manufacturing employment during the 2000s was a substantial cause of rising American unemployment, especially for less-educated prime-age workers (Charles, Hurst, and Schwartz 2018). However, they also find that these declines were caused by a mix of both import competition and nontrade factors. They further speculate that persistently depressed low-skilled manufacturing employment in the United States was likely caused by *nontrade issues* like a skills mismatch in the U.S. manufacturing sector (which

is becoming more skilled compared to other low-skill professions like retail and construction) and declining cross-region mobility among U.S. workers during the 2000s compared to earlier periods. As a result, “imposing trade barriers against the rest of the world is unlikely to substantially increase the employment prospects of workers with lower levels of accumulated schooling” (ibid.: 63).

Studies have similarly found it difficult to distinguish the employment effects of trade from those of technology. After documenting the evolution of American manufacturers in their aforementioned paper, for example, Fort, Pierce, and Schott (2018: abstract) acknowledge that the “data provide support for both trade- and technology-based explanations of the overall decline of [manufacturing] employment over this period, while also highlighting the difficulties of estimating an overall contribution for each mechanism.”

Additional China Shock context is provided by Eriksson et al. (2019), who show that the China Shock was so “shocking,” not because of China or PNTR, but because of *when* it hit the United States: during regional shifts in the U.S. production of certain labor-intensive goods. In particular, “late stage” industries—with now-standardized processes and low technologies that are susceptible to global competition (particularly in developing countries)—had moved out of high-education/innovation U.S. regions to places with less education and innovative capacity, thus explaining “why the shock hurt in these areas to the extent that it did.” This timing adds to the uniqueness of the China Shock, as the authors find that previous U.S. trade shocks—involving Japan and the Asian Tigers, for example—had no such dynamic (and thus far more limited labor market effects). The analysis also shows that these “late-stage” industries were well on their way out of the United States *regardless of the China Shock*.

Many other experts have questioned whether the China Shock literature tells the whole story about Chinese imports, U.S. manufacturing jobs, and related issues. Economists have found substantial net benefits for the United States when more fully accounting (e.g., through a general equilibrium model) for Chinese import competition. In particular, the China Shock was found to cause fewer manufacturing job losses (Caliendo, Dvorkin, and Parro 2019); to be accompanied by offsetting job gains in U.S. manufacturing exports and services (Feenstra, Ma, and Xu 2017; Feenstra and Sasahara 2017); and to result in a net loss of only 300,000 U.S. jobs—just

0.22 percent of nonfarm employment (DeLong 2017). Others found one-third *fewer* manufacturing job losses and different regional effects when using value-added trade flows to measure the China Shock and 20–30 percent fewer job losses when accounting for booms and busts in the U.S. housing market (Xu, Ma, and Feenstra 2019).

Others disagree with the China Shock authors' general conclusions about the impact of Chinese imports on U.S. jobs. Alan Reynolds (2016), for example, argues that the China Shock's "micro-economic model designed for local 'commuting zones' cannot properly be extended to the entire national economy," and therefore misses important macroeconomic effects of U.S.-China trade liberalization like increased U.S. exports (to China and other countries), and that extending the period beyond 2011, during which the U.S. economy was still affected by the Great Recession, causes half of the job loss attributed to the China Shock to "disappear."

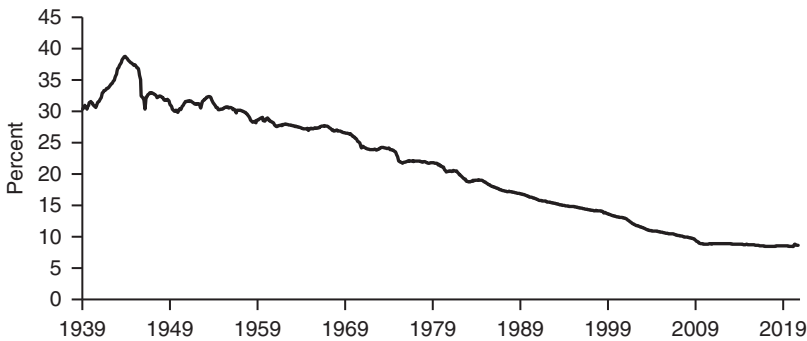
Charles Freeman, who ran the U.S. Trade Representative's Office of China Affairs during the George W. Bush administration, recalls, "We just didn't see any profound direct US job losses in sectors exposed to new direct competition from China" (Freeman 2019). Phil Levy, a member of the George W. Bush administration's Council of Economic Advisers, adds that the fungibility of Chinese and other developing country imports undermines the argument that Chinese imports—as opposed to imports more generally—were to blame for some of the manufacturing job losses that occurred during the China Shock period (Levy 2016). Levy concludes from this experience that it "calls into question the premise of [the China Shock] analysis. If the alternative to imports from China was imports from other developing nations, then the impact of China on U.S. workers was negligible."

The data tend to corroborate Freeman's and Levy's claims. First, Figure 1 shows only a modest change in trend for manufacturing jobs as a share of the U.S. workforce before and after PNTR passed and China entered the WTO.¹

Second, data indicate that, pace Levy, Chinese imports may have replaced other imports more than they did domestic production

¹ Economic Research from the Federal Reserve Bank of St. Louis, "All Employees: Manufacturing/All Employees: Total Nonfarm Payrolls," <https://fred.stlouisfed.org/graph/?g=mcsO>.

FIGURE 1
MANUFACTURING SHARE OF TOTAL U.S. EMPLOYMENT



SOURCE: Federal Reserve Bank of St. Louis.

before, during, and after the China Shock. According to the Congressional Research Service, for example, the total share of imports into the United States from Pacific Rim countries between 1990 and 2017 remained relatively constant at 47.1 percent, but “the role of China as a supplier of U.S. manufactured products among Pacific Rim countries increased sharply, while the relative importance of the rest of the Pacific Rim (excluding China) for these products sharply decreased,” a result “partly due to many multinational firms shifting their export-oriented manufacturing facilities from other countries to China” (Morrison 2018: 10–11).

Economists at the San Francisco Federal Reserve Bank also found that Americans’ total import consumption, as measured by 2017 personal consumption expenditures (PCE), remained relatively steady during the China Shock period. According to Hale et al. (2019): “The fact that the overall import content of U.S. consumer goods has remained relatively constant while the Chinese share has increased demonstrates that Chinese gains have come, in large part, at the cost of other exporters, namely Japan.”

That economists repeatedly and openly express reservations—supported by various trade and employment data—about blaming China trade alone for massive declines in U.S. manufacturing employment should foment similar levels of caution among U.S. politicians and pundits.

Finally, there is the matter of putting the China Shock's effects into proper perspective. For example, the 97,000 annual manufacturing job losses attributable to Chinese imports (on average) represents less than 20 percent of total involuntary job losses in manufacturing and less than 5 percent of total involuntary job losses (Lawrence 2014). Autor himself (Clement 2016) has called his estimate of 2 million jobs lost an "upper bound" (the more likely central estimate was about half that number), which includes around 1 million non-manufacturing jobs (Dubner 2017).

Overall, the aforementioned literature review reveals that the claims of harm from Chinese trade are likely wildly overstated while the substantial economic benefits are usually ignored, and that the China Shock issues are more uncertain and complex than the caricature painted by PNTR/China critics. Moreover, current calls to overhaul the U.S. economy based on the China Shock are unsupportable when placing the jobs lost to Chinese import competition in proper perspective, and when considering that the disruptions likely caused by the China Shock would still have happened in its absence.

The Reality of China's WTO Accession and Export Competitiveness

PNTR was not responsible for first opening the United States to Chinese imports. Prior to the law's enactment, China held "most favored nation" (MFN) trade status since 1980, meaning the country faced no greater trade barriers than most other U.S. trading partners. Only once between 1990 and 2001 was China's MFN/NTR status truly in doubt: in 1992, when a presidential veto was needed to maintain it. As a result, Chinese imports to the United States increased more than sixfold in the decade *preceding* PNTR, and by the late 1990s, the rational expectation of most U.S. importers was more of the same.

Nevertheless, there is evidence that the certainty of "permanent" trade relations accelerated the growth of Chinese imports into the United States. In particular, researchers have found a substantial connection between PNTR, Chinese imports in sectors that would have faced high tariffs in the absence of MFN/NTR, and U.S. jobs (Pierce and Schott 2016; Handley and Limão 2017). Other experts, however, question the magnitude of the PNTR "uncertainty driver." For example, 2019 research found that the annual MFN/NTR votes

actually increased Chinese imports into the United States as a result of importers' front-running (increasing shipments in advance of) any potential tariff increases (Alessandria, Khan, and Khederlarian 2019), and that the probability of NTR denial averaged only about 5.5 percent between 1990 and 2001, reaching a mere 1.4 percent in 2001. Thus, the trade-dampening effects of MFN/NTR uncertainty had evaporated by the late 1990s.

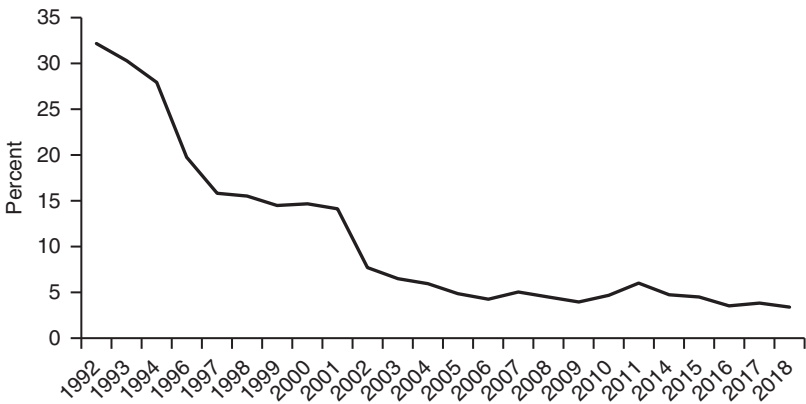
Regardless of which expert is ultimately correct, the *Congressional Record* and Chinese trade flow data contradict the popular assertion that an isolated U.S. policy choice in 2000 first exposed the United States to Chinese import competition. At most, PNTR merely accelerated a bilateral economic integration that was already well underway.

More importantly, ample evidence shows that PNTR was *not* the only driver of the China Shock. Multiple studies have found, for example, that the reduction in trade policy uncertainty accounted for only about one-third of the growth in Chinese exports to the United States in the 2000s, with the remaining two-thirds attributable to *China's own market reforms* (Handley and Limão 2017; Amiti et al. 2018). These include privatization, trading rights, and import liberalization, often in response to new WTO commitments, which themselves were major contributors to China's export competitiveness in the late 1990s and 2000s (Autor, Dorn, and Hanson 2018; Jakubik and Stolzenburg 2018). Indeed, as shown in Figure 2, average Chinese tariffs declined from more than 30 percent in the 1990s to less than 9 percent in 2006, and even lower on a trade-weighted scale.

Thus, PNTR probably accelerated Chinese exports to the United States, but China's own reforms—far beyond the control of Washington policymakers—also fueled the China Shock.

Furthermore, China's WTO accession was not “shocking” to anyone watching U.S. trade policy in the 1990s. China first applied to join the WTO's predecessor, the General Agreement on Tariffs and Trade (GATT), in 1985, then reapplied in 1995 when the WTO came into being, and finally acceded to the body in 2001. China's accession over this time involved dozens of bilateral and multilateral (“working party”) meetings, negotiating texts, disclosures, and internal reforms. China's final accession package—a “Working Party Report” and “Protocol of Accession,” plus liberalization schedules for goods and services—contained hundreds of pages of commitments (by far the

FIGURE 2
CHINA TARIFF RATES, 1992–2018



NOTE: Applied weighted mean, all products (percent).

SOURCE: World Bank.

most of any acceding member up to that point and still today considered to be some of the deepest ever). This included many “WTO-plus” commitments that the United States and other members dictated (via bilateral accession agreements) and have since been used, for example, to challenge Chinese laws through dispute settlement or to restrict Chinese imports (Chemutai and Escaith 2017).

The United States also did not rubber-stamp China’s WTO accession or base it on Pollyannaish dreams of Chinese democratization. Instead, the United States was the final holdout among large industrialized nations to approve China’s WTO accession via bilateral negotiations, demanding ever more concessions from the Chinese government over a contentious 13-year negotiation (Nakatsuji 2001). U.S. trade representatives for multiple presidents from each major party also frequently consulted with Congress and the private sector, including labor unions, at every step of the process (as required by U.S. law).

Furthermore, creating a liberal democracy in China was not a primary reason for the United States government’s approval of China’s WTO accession. Instead, key Clinton administration speeches and policy documents demonstrate that U.S.–China engagement “was a

balancing act with multiple objectives”—most of them pragmatic—including “increasing bilateral dialogues, . . . WMD nonproliferation in East and South Asia, preventing the nuclearization of the Korean Peninsula, cooperating on disease and environmental issues, better market access [for U.S. companies] and intellectual property rules, fighting organized crime, ensuring stability in the Taiwan Strait, and WTO accession on ‘commercial terms,’ among others” (Thomas 2019). Democratization, on the other hand, was rarely mentioned.

In reality, Washington policymakers had no good alternatives to granting PNTR to China (a move that every other WTO member had done years earlier), especially based on the facts available at that time. The two actual alternatives to PNTR—letting China in the WTO but continuing the *annual* NTR process (or even raising tariffs on Chinese goods), or keeping China from the WTO entirely—were inferior, in terms of both the economics and geopolitics, to granting PNTR. Neither would have prevented China’s rise, given the aforementioned realities of China’s export competitiveness (driven mainly by internal reforms), the multipolar nature of today’s global economy (offering economic options beyond the United States), and the pervasiveness of global supply chains (allowing Chinese inputs to enter the United States via finished goods made in third countries) (Levy 2018).

Regardless, marshaling the necessary WTO-wide consensus to deny more than one billion people in a modernizing economy access to an open multilateral trade organization—one that already included communist Cuba and for decades had tolerated Eastern Bloc command-and-control economies and “socialist” countries with pervasive state-owned industries—was not realistic, especially given what policymakers could have known at the time about China’s relatively liberal leadership and impressive economic reforms. China was entering the WTO, whether critics liked it or not.

Ignoring Government Failures and Market Realities

Blaming PNTR for Chinese economic malfeasance and the problems of the American working class ignores many key facts about federal policy over the last several decades.

First, the U.S. government has missed several opportunities to check China’s actions and support domestic labor markets since

Beijing undertook the economic liberalization and legal commitments required of WTO accession. Most of those actions—for example on industrial subsidies and intellectual property—are covered by WTO rules and can be litigated through dispute settlement (Bacchus, Lester, and Zhu 2018). Contrary to popular wisdom, moreover, such litigation has proven effective. For example, the United States was undefeated at the WTO when challenging Chinese trade practices between 2002 and 2018 (Schott and Jung 2019) and has won several more cases since that time. And when China loses WTO disputes, it tends to comply with the decisions. Chinese compliance is not perfect (nor is any other WTO member's), but it is arguably better than that of the United States, which has famously shirked WTO rulings on subsidies, antidumping rules, and internet gambling.

The refusal of the United States and other WTO members to pursue more disputes against China—or open “compliance proceedings” when China does not fully comply—is a policy worth criticizing, but says *nothing* about the original decision to admit China to the WTO in the first place. Indeed, to claim that China's WTO accession terms were weak and that the WTO cannot discipline China's unfair trade practices without fully utilizing the sole means of imposing such discipline (dispute settlement) is declaring defeat without firing a shot.

Other policy mistakes since the passage of PNTR also deserve scrutiny. Among these are the United States' (1) withdrawal from the Trans-Pacific Partnership (TPP), which was designed in part to counterbalance China's economic and geopolitical ambitions; (2) failure to reform tax, trade, and immigration policies that inhibit American companies' global competitiveness; (3) failure to modernize adjustment assistance and worker retraining programs intended to mitigate trade, technological, or cultural disruptions; or (4) continued imposition of tax, education, occupational licensing, criminal justice, zoning, and other policies that discourage labor adjustment and economic dynamism. Such policies are indeed worthy of criticism and debate but cannot inform the decision to pass PNTR, allow China to join the WTO, or otherwise “normalize” trade with China. And blaming China for these policies' inevitable failures relieves them of the scrutiny that they deserve.

Second, critics of engagement with China also often ignore the United States' own history of market interventions and their failures

to help companies and workers—even during the China Shock. Most importantly, the only apparent alternative to freer trade—protectionism—has repeatedly proven both costly and ineffective. For example, International Monetary Fund economists in 2018 examined data for 151 countries over 51 years (1963–2014) and found that “tariff increases lead, in the medium term, to economically and statistically significant declines in domestic output and productivity” as well as more unemployment and higher inequality (Furceri et al. 2019). The same is true for American protectionism, which has been found to impose immense economic costs that outweigh any benefits; to fail to protect American firms and workers; and to breed corruption, cronyism, and political dysfunction (Lincicome 2017).

Nevertheless, the United States has maintained or implemented (1) significant tariffs and tariff-rate quotas on imports of “sensitive” products like trucks, apparel, footwear, and food (Lincicome 2016); and (2) a relatively high number of non-tariff trade interventions (Global Trade Alert 2020), including hundreds of special restrictions on “unfair” Chinese imports (imposed via antidumping and countervailing duty laws, and “Section 337” actions that remedy intellectual property rights violations), as well as numerous other “national security” restrictions on Chinese inbound investment or on U.S. technology exports to China.

Once again, however, these measures have proven ineffective. For example, President Barack Obama’s 2009 imposition of 35 percent “special safeguard” tariffs on Chinese tires was found to result, even under the best assumptions, in a handful of American jobs saved at an annual cost to U.S. consumers of over \$900,000 per job, plus a substantial increase in non-China imports instead of new U.S. production (Hufbauer and Lowry 2012). Today, the industry’s prospects are no better. A 2017 review of all U.S. antidumping investigations against Chinese imports between 1998 and 2006 revealed that the duties simply caused Chinese imports to be replaced by other imports instead of bolstering U.S. producers (Lee, Park, and Saravia 2017).

The U.S. government also has long provided financial and other support to favored industries and workers, for example, through auto bailouts, steel industry bailouts, alternative energy subsidies, manufacturing tax credits, Export–Import Bank loans and other export assistance, procurement preferences like Buy American and the Davis–Bacon Act, shipping restrictions like the Jones Act and the Passenger Vessel Services Act, and the billions of other taxpayer

dollars that the United States has doled out to “blue collar” industries and workers over the last few decades at the federal level alone. As I documented in a 2012 article on global subsidies and anti-subsidy disciplines, “despite the obvious economic, legal, and political problems associated with domestic subsidies, the United States remains one of the world’s largest subsidizers” (Lincicome 2012).

The United States government has also repeatedly tried to fund and retrain workers, most notably through the Trade Adjustment Assistance (TAA) program, which offers generous subsidies to U.S. workers affected by import competition. Unfortunately, TAA has proven to be a “notorious failure” and, “multiple studies commissioned by the Labor Department have found that TAA participants are worse off, as measured by future wages and benefits, than similarly situated jobless individuals outside the program” (Lincicome 2016).

All of these restrictions and programs refute the common claim that U.S. policymakers simply passed PNTR and walked away from the American working class out of some sort of “market fundamentalism” or rigid adherence to “laissez faire ideology.” The real problem was that the interventions just did not work very well. As a 2013 Congressional Research Service report concluded about the state of American manufacturing, “Congress has established a wide variety of tax preferences, direct subsidies, import restraints, and other federal programs with the goal of retaining or recapturing manufacturing jobs, [but] only a small proportion of US workers is now employed in factories” (Levinson 2013).

Finally, those seeking to blame PNTR or Chinese imports for the current plight of the American working class ignore the many places in the United States that were affected by Chinese import competition but did adjust and have thrived economically—often with the help of trade and foreign investment. Indeed, the fact that the longer-term effects of Chinese import competition vary dramatically from place to place, even in states or regions that faced similar import competition, undermines the notion that the China Shock was a national trade problem necessitating national protectionism, as opposed to a local adjustment problem necessitating local solutions (see Caliendo, Dvorkin, and Parro 2019).

Many cities and towns in America that were once known for low-skill manufacturing and recently faced intense import competition—places like Pittsburgh and Bethlehem, Pennsylvania; Greenville–Spartanburg, South Carolina; Hickory, North Carolina;

Warsaw, Indiana; and Waterloo, Iowa—have since adapted and thrived. Several studies show that most U.S. regions ended up better off following the China Shock. Furthermore, a 2018 report finds that 115 of the 185 U.S. counties identified as having a disproportionate share of manufacturing jobs in 1970 had “transitioned successfully” away from manufacturing by 2016, and that, of the remaining 70 “older industrial cities,” 40 exhibited “strong” or “emerging” economic performance between 2000 and 2016 (Berube and Murray 2018).

Anecdotal evidence reiterates these findings: towns that once depended on low-skill manufacturing are now home to thriving companies that succeeded by adapting to the market, including through international trade and investment. Anyone doubting such successes need only take a drive down Interstate 85 from Charlotte, North Carolina, to Montgomery, Alabama, to see firsthand the impressive economic development—especially the multinational automotive facilities highlighted in a recent Federal Reserve Bank of New York study of U.S. manufacturing job growth (Abel and Deitz 2019).

The contrast between now-thriving American towns and those still reeling from a trade shock that ended a decade ago again indicates that the problem the “China Shock” revealed was not import competition but many communities’ inability to adjust to economic changes. Thus, those commentators and politicians who blame China trade for the difficulties of the American working class should stop asking, “Why did elites normalize trade with China in the 1990s?” and instead ask, “What did many American towns, companies, and workers do right in the face of intense import competition, and how can local, state, and federal policies encourage that important improvement?”

Conclusion

The historical record and numerous academic analyses of the China Shock provide a straightforward explanation for recent U.S. trade policy toward China. Engagement and liberalization were a pragmatic and bipartisan policy choice with few, if any, legitimate alternatives. This engagement produced real economic benefits for most Americans while bolstering the multilateral trading system and removing historical inequities under the previous, more protectionist U.S. trade policy regime.

This assessment of the China Shock, China's WTO accession, and PNTR thus answers two counterfactual questions that critics seemingly never ask:

- Would U.S. manufacturing jobs have disappeared in the absence of the China Shock?
- Would denying PNTR have stopped China's rise?

The weight of the evidence tilts strongly toward “yes” on the former and “no” on the latter: non-China imports and other disruptive forces would have eventually destroyed the U.S. manufacturing jobs whose loss is frequently blamed on China, and China was bound to emerge as a global trading power with or without PNTR.

The resulting economic disruption and adjustment were difficult for some U.S. regions and workers—more difficult than many experts expected—and certainly post-liberalization policy mistakes were made. With the benefit of hindsight, one can legitimately claim that certain specific “WTO-plus” rules should have been drafted differently during China's accession.

That said, the facts do not support assertions from American politicians and pundits that engagement with China in the 1990s and 2000s was a mistake, and that denying China's WTO admission would have improved U.S. economic and geopolitical standing today; or that the labor and cultural issues in America are the fault of “Washington elites” who pursued normalized trade with China to benefit corporate donors and democratize communist China, while refusing to support the working class.

References

- Abel, J. and Deitz, R. (2019) “Where Are Manufacturing Jobs Coming Back?” *Liberty Street Economics*, Federal Reserve Bank of New York (February 6).
- Alessandria, G. A.; Khan, S. Y.; and Khederlarian, A. (2019) “Taking Stock of Trade Policy Uncertainty: Evidence from China's Pre-WTO Accession.” NBER Working Paper No. 25965.
- Amiti, M.; Dai, M.; Feenstra, R. C.; and Romalis, J. (2018) “How Did China's WTO Entry Affect U.S. Prices?” NBER Working Paper No. 23487.
- Antràs, P.; Fort, T. C.; and Tintelnot, F. (2017) “The Margins of Global Sourcing: Theory and Evidence from U.S. Firms.” *American Economic Review* 107 (9): 2514–64.

- Autor, D.; Dorn, D.; and Hanson, G. (2018) “When Work Disappears: Manufacturing Decline and the Falling Marriage-Market Value of Young Men.” NBER Working Paper No. 23173.
- Bacchus, J.; Lester, S.; and Zhu, H. (2018) “Disciplining China’s Trade Practices at the WTO.” Cato Policy Analysis No. 856 (November 15).
- Bai, L., and Stumpner, S. (2019) “Estimating US Consumer Gains from Chinese Imports.” *American Economic Review* 1 (2): 209–24.
- Berube, A., and Murray, C. (2018) “Renewing America’s Economic Promise Through Older Industrial Cities.” Washington: Brookings Institution (April).
- Broda, C., and Romalis, J. (2008) “Inequality and Prices: Does China Benefit the Poor in America?” University of Chicago Working Paper (March 10).
- Bureau of Labor Statistics (BLS) (2018) “Labor Force Statistics from the Current Population Survey,” Household Data. Available at www.bls.gov/cps/cpsaat11.htm.
- Caliendo, L.; Dvorkin, M.; and Parro, F. (2019) “Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Shock.” *Econometrica* 87 (3): 741–835.
- Charles, K. K.; Hurst, E.; and Schwartz, M. (2018) “The Transformation of Manufacturing and the Decline in U.S. Employment.” NBER Working Paper No. 24468 (March).
- Chemutai, V., and Escaith, H. (2017) “Measuring World Trade Organization (WTO) Accession Commitments and their Economic Effects.” *Journal of International Commerce, Economics and Policy* 8 (2): 1–27.
- Clement, D. (2016) “Interview with David Autor.” Federal Reserve Bank of Minneapolis (September 7). Available at www.minneapolisfed.org/article/2016/interview-with-david-autor.
- DeLong, J. B. (2017) “NAFTA and Other Trade Deals Have Not Guttled American Manufacturing—Period.” *Vox* (January 24).
- Dubner, S. (2017) “Did China Eat America’s Jobs?” *Freakonomics* (January 25). Available at <https://freakonomics.com/podcast/china-eat-americas-jobs>.
- Eriksson, K.; Russ, K.; Shambaugh, J. C. ; and Xu, M. F. (2019) “Trade Shocks and the Shifting Landscape of U.S. Manufacturing.” NBER Working Paper No. 25646 (March).

- Fajgelbaum, P. D., and Khandelwal, A. K. (2014) “Measuring the Unequal Gains from Trade.” NBER Working Paper No. 20331 (July).
- Feenstra, R.C.; Ma, H.; and Xu, Y. (2017) “US Exports and Employment.” NBER Working Paper No. 24056 (November).
- Feenstra, R. C., and Sasahara, A. (2017) “The ‘China Shock’, Exports and U.S. Employment: A Global Input-Output Analysis.” NBER Working Paper No. 24022 (November).
- Fort, T. C. ; Pierce, J. R.; and Schott, P. K. (2018) “New Perspectives on the Decline of U.S. Manufacturing Employment,” NBER Working Paper No. 24490 (April).
- Freeman, C. (2019) Twitter Thread (January 15, 1:58 p.m.). Available at https://twitter.com/AsiaPac_Freeman/status/1085249972767997954.
- Furceri, et al. (2019) “Macroeconomic Consequences of Tariffs.” International Monetary Fund Working Paper No. 19/9.
- Galle, S.; Rodríguez-Clare, A.; and Yi, M. (2017) “Slicing the Pie: Quantifying the Aggregate and Distributional Effects of Trade.” NBER Working Paper No. 23737 (August).
- Global Trade Alert (2020) “Independent Monitoring of Policies That Affect World Commerce.” Available at www.globaltradealert.org.
- Gutiérrez, G., and Philippon, T. (2017) “Declining Competition and Investment in the U.S.” NBER Working Paper No. 23583 (July).
- Hale, G.; Hobijn, B.; Nechio, F.; and Wilson, D. (2019) “How Much Do We Spend on Imports?” Federal Reserve Bank of San Francisco, *Economic Letter* (January 7).
- Handley, K., and Limão, N. (2017) “Policy Uncertainty, Trade, and Welfare: Theory and Evidence for China and the United States.” *American Economic Review* 107 (9): 2731–83.
- Hufbauer, G. C., and Lowry, S. (2012) “US Tire Tariffs: Saving Few Jobs at High Cost.” Policy Brief 12–9 (April). Washington: Peterson Institute for International Economics.
- Jakubik, A., and Stolzenburg, V. (2018) “The ‘China Shock’ Revisited: Insights from Value Added Trade Flows.” World Trade Organization Staff Working Paper ERSD-2018–10 (October 26).
- Jaravel, X., and Sager, E. (2018) “What Are the Price Effects of Trade? Evidence from the U.S. and Implications for Quantitative Trade Models.” U.S. Bureau of Labor Statistics Working Paper No. 506 (September).

- Lardy, N. R. (2000) “PNTR for China.” Policy Brief No. 58 (May). Washington: Brookings Institution.
- Lawrence, R. (2014) “Adjustment Challenges for U.S. Workers.” In C. F. Bergsten, G. C. Hufbauer, and S. Miner (eds.), *Bridging the Pacific: Toward Free Trade and Investment between China and the United States*. Washington: Peterson Institute for International Economics.
- Lee, M.; Park, D.; and Saravia, A. (2017) “Trade Effects of U.S. Antidumping Actions Against China.” *Asian Economic Journal* 31 (1): 3–16.
- Levinson, M. (2013) “‘Hollowing Out’ in U.S. Manufacturing: Analysis and Issues for Congress.” Washington: Congressional Research Service. Available at <https://fas.org/sgp/crs/misc/R41712.pdf>.
- Levy, P. (2016) “Did China Trade Cost the United States 2.4 Million Jobs?” *Foreign Policy* (May 8).
- _____ (2018) “Was Letting China into the WTO a Mistake?” *Foreign Affairs* (April 2).
- Lincicome, S. (2012) “Countervailing Calamity: How to Stop the Global Subsidies Race.” Cato Policy Analysis No. 710 (October 9).
- _____ (2016) “The Truth about Trade.” *National Review* (April 11).
- _____ (2017) “Doomed to Repeat It: The Long History of America’s Protectionist Failures.” Cato Policy Analysis No. 819 (August 22).
- _____ (2020) “Testing the ‘China Shock’: Was Normalizing Trade with China a Mistake?” Cato Policy Analysis No. 895 (July 8).
- Morrison, W. M. (2018) “China–U.S. Trade Issues.” Congressional Research Service (July 30). Available at <https://fas.org/sgp/crs/row/RL33536.pdf>.
- Nakatsuji, K. (2001) “Essence of Trade Negotiation: A Study on China’s Entry for WTO.” Available at www.ritsumei.ac.jp/acd/cg/ir/college/bulletin/vol14-1/015-34_nakatsuji.pdf.
- Pierce, J. R., and Schott, P. K. (2016) “The Surprisingly Swift Decline of U.S. Manufacturing Employment.” *American Economic Review* 106 (7): 1632–62.
- Reynolds, A. (2016) “Did the U.S. Lose 2.4 Million Jobs from China Imports?” *Cato at Liberty* (September 15).

- Schott, J., and Jung, E. (2019) "In U.S.-China Trade Disputes, the WTO Usually Sides with the United States." *Peterson Institute for International Economics* (March 12).
- Thomas, N. (2019) "Matters of Record: Relitigating Engagement with China." *Macro Polo* (September 3).
- United States Census Bureau (2020) "Top Trading Partners." Available at www.census.gov/foreign-trade/statistics/highlights/toppartners.html.
- U.S.-China Business Council (USCBC) (2019) "2019 State Export Report: Goods and Services Exports by U.S. States to China over the Past Decade." Available at www.uschina.org/reports/2019-state-export-report.
- Wang, Z.; Wei, S. J.; Yu, X.; and Zhu, K. (2018) "Re-Examining the Effects of Trading with China on Local Labor Markets: A Supply Chain Perspective." NBER Working Paper No. 24886 (October).
- White House (2001) "President [George W.] Bush Grants Permanent Normal Trade Relations Status to China." Washington: Office of the Press Secretary (December 27). Available at <https://georgewbush-whitehouse.archives.gov/news/releases/2001/12/20011227-2.html>.
- World Trade Organization (WTO) (2020) "Trade in Value-Added and Global Value Chains: Statistical Profiles," Profile of the United States. Available at www.wto.org/english/res_e/statis_e/miwi_e/countryprofiles_e.htm.
- Xu, Y.; Ma, H.; and Feenstra, R. C. (2019) "Magnification of the 'China Shock' through the U.S. Housing Market." NBER Working Paper No. 26432 (November).

IMPACT OF INSTITUTIONS IN THE AFTERMATH OF NATURAL DISASTERS

Diego A. Díaz and Cristian Larroulet

The number and impact of natural disasters are increasing because of climate change and more people living in urban areas (Sanderson and Sharma 2016). The mechanism is simple, at least when considering climatic events: higher temperatures lead to higher rates of water evaporation, which increases the chance of flooding events (Wallace et al. 2014; IPCC 2001). The number of hot days has increased and the number of cold days has decreased in land areas, with model projections indicating that extreme precipitation events will continue to increase, resulting in more floods and landslides. At the same time, mid-continental areas will get dryer, which will increase the chance of droughts and wildfires (Van Aalst 2006). The course of action taken by humanity in the next decades will likely play a pivotal role since extreme differences in projections are expected if global temperatures rise 2°C in comparison to 1.5 °C above preindustrial levels (Allen et al. 2019). What are the economic impacts of natural disasters? This question has been addressed to a large extent in the literature, but it still does not have a conclusive response.

Cato Journal, Vol. 41, No. 1 (Winter 2021). Copyright © Cato Institute. All rights reserved. DOI:10.36009/CJ.41.1.4.

Diego A. Díaz is a graduate student at the Harris School of Public Policy at the University of Chicago and a Research Associate at the School of Business and Economics at Universidad del Desarrollo. Cristian Larroulet is a Professor at the School of Business and Economics at Universidad del Desarrollo. We want to thank Juan Pablo Couyoumdjian and Maria de los Angeles Figueroa for useful discussions and Andrés Marticorena for exceptional research assistance. We are also thankful for comments received from participants at workshops at the School of Business and Economics and School of Government at Universidad del Desarrollo.

The seemingly natural reasoning that destruction cannot lead to a net benefit for society was explained almost two centuries ago by Bastiat (1850) in his famous broken window fallacy. A shopkeeper's son, Bastiat relates, breaks a pane of glass in his father's store. The father, angry due to the boy's careless action, is offered consolation by the spectators, who claim that the event is positive for the economy since it provides labor to glaziers. While Bastiat acknowledges that the accident brings trade to the glazier since the shopkeeper has to replace the window, regarding the event as wealth-increasing conveys a narrow perspective. The shopkeeper ends up poorer since he cannot spend the same money elsewhere, and if the boy had not broken the window, then the labor and other materials that were used to repair the damage would have been used elsewhere, potentially making the tangible wealth of the community grow.

As reasonable as Bastiat's argument might sound, it is not clear if natural disasters, or any destructive event for that matter, should affect economic growth and in which direction it might do so. There is a difference in the predictions that an economist would obtain with neoclassical growth models and certain endogenous growth models. While the former theorizes that a natural disaster—that is, a negative shock to capital or both to capital and labor—decreases output in the short run, it does not affect the steady-state of the economy in the long run. Neoclassical models also predict that the destruction of the capital stock will temporarily accelerate growth immediately after the disaster by increasing the marginal return on capital. On the other hand, some endogenous growth models based on Schumpeterian creative destruction can predict an overall higher growth rate produced by an accelerated replacement of the capital stock with more productive capital, as is the case in vintage capital models (Hallegatte and Dumas 2009).

Given the extensive debate regarding natural disasters and economic growth, in an attempt to shed light on the mechanisms by which these destructive events affect the economy, we ask whether economic institutions and, in particular, economic freedom, is relevant for obtaining a higher rate of economic growth, independent of whether this growth is positive or negative. There are not many articles that have studied the relationship between institutions and economic growth after natural disasters. In our review of the literature, we found the following published articles: Felbermayr and Gröschl

(2014), Barone and Mocetti (2014), Raschky (2008), and Kahn (2005). Our objective is to expand on this literature by emphasizing the role of economic freedom as a relevant factor in explaining the economic recovery process in the aftermath of disasters, which is increasingly relevant today given the climate change projections for the next decades. As the frequency of natural disasters increases and their economic impact in society can be ameliorated with better institutions, it becomes imperative to find out, with as much precision as possible, which institutions are relevant and what policies can be put in place to lessen the destruction.

In the following section, we discuss the relevant literature regarding the impact of natural disasters on output and the role of institutional measures. Then we describe our methods and data, present the regression results, and offer some concluding comments.

Literature Review

Natural disasters are a topic that has been studied extensively in the scientific literature. Over 6,000 articles matching the topic “natural disasters” are registered in the Web of Science core collection. The body of literature that focuses on economic growth is much smaller, with 136 articles matching search results for “natural disasters” and “economic growth” at the same time.¹ Instead of attempting to provide a general perspective regarding this literature, we focus on the few articles that study the effects of institutions and their relationship with economic activity after a natural disaster (Felbermayr and Gröschl 2014; Barone and Mocetti 2014; Raschky 2008; Kahn 2005) and discuss their findings.

Felbermayr and Gröschl (2014), in a growth model that shows that disasters are negatively correlated with growth, included two institutional variables in their analysis: measures of democracy (polity index) and trade openness. They found that only the latter is significant in reducing the negative economic impact of natural disasters. Barone and Mocetti (2014) focused on whether the quality of institutions affect subsequent growth in GDP per capita after an earthquake by performing a case study of two Italian regions. They construct what

¹The number of articles indexed in the Web of Science core collection is 136 when searching the terms “natural disasters” and “economic growth” up until July 25, 2020.

they call an “overall institutional quality index,” which they estimate through a principal component analysis based on four local variables: corruption levels, the share of politicians involved in scandals, electoral turnout, and newspaper readership. The authors found that the region with the higher institutional quality index experiences faster growth after the disaster.

In a cross-country setup, Raschky (2008) studied the relationship between economic development and vulnerability to natural disasters, suggesting that although economic development reduces the number of disaster victims and the amount of economic losses, increasing wealth inverts the relationship and causes higher losses. He concluded that higher income does not necessarily lead to better protection against disasters. As a secondary research question, he addressed the importance of institutional quality as a socioeconomic factor that provides protection against natural hazards, for which he used data on governmental stability and the investment climate, which are institutional measures from the International Country Risk Guide (ICRG). Raschky found that both measures, government stability and investment climate, have a significant impact on reducing the death toll and economic losses from natural disasters. In a similar line, Kahn (2005) studied whether institutions are relevant for predicting the death toll from disasters by examining the impact of the Polity index from Systemic Peace and the World Governance Indicators from the World Bank. He determined with borderline significance that *ceteris paribus*, democracies experience less death from disasters, which he attributes to the fact that democracies have greater accountability and less corruption. As for the World Governance Indicators, which include measures of property rights, democracy, regulatory quality, voice and accountability, rule of law, and control of corruption, Kahn finds an overall positive correlation with less death by disasters.

The importance of institutions for economic growth has been established since North (1991), who was the first to explain how institutions provide the incentive structure of an economy (see also Acemoglu, Johnson, and Robinson 2005). Knowing the rules of the game and being confident that those rules will be enforced is a key component for growth, and factors such as property rights, regulation, inflation, civil liberties, political rights, freedom of the press, government expenditures, and trade barriers have all been linked to growth in empirical studies (Talbot and Roll 2001).

The structure of property rights is a critical economic institution and, therefore, an essential factor underlying growth. With poor property rights, individuals will have little incentive to invest in physical or human capital, which is a larger problem after a natural disaster since a portion of the capital stock has been destroyed and, therefore, the incentives to increase it are higher. It is also likely that a society with more disregard for property rights will have a higher incidence of looting behavior after a disaster, which is detrimental to the recovery process.

A mechanism commonly presented in the empirical literature that is used to explain the positive effects of natural disasters is the accelerated replacement of the existing capital stock with more productive capital, which causes a faster embodiment of new technologies and temporarily accelerates growth (Albala-Bertrand 1993; Skidmore and Toya 2002). In Hallegatte and Dumas (2009), this mechanism is referred to as the “productivity effect” and is investigated in detail in a Solow-like model with embodied technological change. Without this mechanism, one way to explain higher economic growth after a natural disaster is by the higher amount of government spending in the aftermath of the disaster. An economic boom can occur driven by the construction sector (Benson and Clay 2004). However, spending will be easier to finance if institutions that value sound fiscal policy are present. If a country has an unbalanced budget with a high deficit and sustained debt, then financing the costs of reconstruction will increase the deficit and debt level even more, which is associated with lower levels of growth (Checherita-Westphal and Rother 2012).

Entrepreneurship activities have been increasingly linked to natural disasters, with recent evidence suggesting that they decrease start-up activity in the next two years after a disaster (Boudreaux, Escaleras, and Skidmore 2019), although they have also been shown to be correlated with an increase on entrepreneurial intention (Monllor and Murphy 2017).

The relationship between entrepreneurship and disasters is relevant given the link between entrepreneurial activity, economic growth, and institutions. An atmosphere that promotes strong economic institutions is also favorable for entrepreneurship (Larroulet and Couyoumdjian, 2009), which is a fundamental driver for growth, especially after a disaster in which there is significant destruction of the capital stock.

Methods and Data

We use publicly available data from two different sources for our regression model, GeoMet for economic and disaster variables (Felbermayr and Gröschl 2014), and the Fraser Institute's Economic Freedom Index for institutional quality, which includes the size of government, property rights, sound money, freedom to trade internationally, and regulation.² The reason behind using these institutional variables as opposed to, for example, the World Governance Indicators by the World Bank, is that the Fraser Institute's variables start further back in time, beginning on a quinquennial basis since 1970 and yearly since 2000, while the data from the World Governance Indicators starts at 1996. We also include the Polity index in our analysis as a measure of the democracy level, for which yearly data are available throughout the entire period in question. Table 1 shows a description of the variables used as well as the sample mean and standard deviation for each.

Following Felbermayr and Gröschl (2014), we use a standard growth regression framework. The units of observation are country-year combinations. We include lagged GDP per capita to estimate a dynamic model, a disaster measure, several control variables, and the economic freedom institutional indexes. Our basic specification takes the following form:

$$(1) \Delta \ln y_{i,t} = (\rho - 1) \ln y_{i,t-1} + \alpha I_{i,t-1} + \beta D_{i,t} + \gamma X_{i,t-1} + \epsilon_{i,t},$$

where $y_{i,t}$ is the GDP per capita of country i at time t , $\Delta \ln y_{i,t}$ is the growth rate of GDP per capita at time t , and $I_{i,t-1}$ is the set of institutional variables included, that is, the Polity index and the Economic Freedom Index, indexed for country and year and lagged one period. $D_{i,t}$ is a yearly measure of disaster intensity, encompassing the disasters that occurred in the country i at time t , as constructed by Felbermayr and Gröschl and available in GeoMet.

²Data from the Fraser Institute can be downloaded from the institute's website at www.fraserinstitute.org/sites/default/files/efw-2019-master-index-data-for-researchers.xlsx. Data from GeoMet can be downloaded from Ifo Institute's website at www.cesifo-group.de/de/dms/ifodoc/docs/IfoGAME/IfoGAME_balanced_panel.dta. Both sources checked online on July 26, 2020.

TABLE 1
DESCRIPTIVE STATISTICS

Variable	Mean	Std. Dev.	Source	Description
Economic Freedom	6.183	1.265	Gwartney et al. (2018)	Degree to which the people are free to enjoy their civil liberties
Size of Government	5.960	1.402	Gwartney et al. (2018)	Index comprising taxation, government spending, and government investment
Property Rights	5.105	1.884	Gwartney et al. (2018)	Degree of impartiality of courts, judicial independence, and the enforcement of contracts
Sound Money	7.245	2.109	Gwartney et al. (2018)	Measures the degree to which money is a reliable store of value
Freedom to Trade Internationally	6.520	1.944	Gwartney et al. (2018)	Degree of openness that comprises tariffs and other barriers
Regulation	6.224	1.286	Gwartney et al. (2018)	Degree of interest rate controls and labor market regulations
Logarithm of GDP per Capita	8.217	1.349	Felbermayr & Gröschl (2014)	Logarithm of GDP per capita
Logarithm of Population	9.489	1.393	Felbermayr & Gröschl (2014)	Logarithm of population
Polity	0.638	0.347	Felbermayr & Gröschl (2014)	Polity index, normalized between 0 and 1
Trade Openness	0.715	0.460	Felbermayr & Gröschl (2014)	Exports plus imports divided by GDP

(Continued)

TABLE 1 (Continued)
DESCRIPTIVE STATISTICS

Variable	Mean	Std. Dev.	Source	Description
Interest Rate	0.007	0.022	Felbermayr & Gröschl (2014)	Real interest rate
Domestic Credit	0.604	0.512	Felbermayr & Gröschl (2014)	Domestic credit in banking sector (share of GDP)
Gross Capital Formation	0.053	0.249	Felbermayr & Gröschl (2014)	Gross capital formation (share of growth)
Foreign Direct Investment	0.027	0.057	Felbermayr & Gröschl (2014)	Net inflows (share of GDP)
Consumer Price Index	0.038	0.335	Felbermayr & Gröschl (2014)	Inflation, consumer prices
Account Balance	-0.025	0.080	Felbermayr & Gröschl (2014)	Account balance as share of GDP
Disaster Index (GeoMet)	0.032	0.164	Felbermayr & Gröschl (2014)	Measure of disaster intensity

NOTES: The dataset contains the combined data from Felbermayr and Gröschl (2014) and Gwartney et al. (2018), including a total of 108 countries that have data on natural disasters between the years 1979 and 2010. The total number of observations is 3,208.

Finally, $X_{i,t-1}$ include the following group of control variables: the logarithm of population, trade openness, the real interest rate, domestic credit in the banking sector (as a share of GDP), gross capital formation (as a share of growth), foreign direct investment (as a share of GDP), the logarithm of the inflation rate, and the current account balance (as a share of GDP).³

We use the extended version of the Emergency Events Database (EM-DAT), GeoMet, for disasters data between the years 1979 and 2010, and the Fraser Institute's Economic Freedom Index database, which has data from 1970 onward. Combining these datasets, we are left with a panel of 90 countries to run our main regression model. As the data on economic freedom before the year 2000 are available only once every five years until 1970, the data are interpolated to be used in our regression model for the period 1979–2010.

Most disaster studies use the EM-DAT Database, but as has been noticed by Felbermayr and Gröschl (2014), there might be an inclusion bias in the database (see Strobl 2012). The problem is that the probability of inclusion could be correlated with income. Even more troublesome is the fact that direct damage might not be a reliable measure of disaster intensity. Monetary damage, or variables correlated to it, are higher in a richer economy after a disaster, which is why they are not reliable variables to be used in growth regressions. Felbermayr and Gröschl (2014) put together GeoMet to account for this issue, which is why their measure of disaster intensity, $D_{(i,t)}$, the same one we use in our regressions, does not depend on estimations of monetary damage but on direct exogenous measures of disaster intensity, as Richter magnitude for earthquakes, wind speed for storms, precipitation level for floods, and temperature level for extreme temperatures events.

Results: Regression Analysis

We present the results of our regression model in Table 2.⁴ The disaster index variable is an unweighted sum of the physical intensity measures of disasters that happened in a specific country in a specific

³As our dependent variable is yearly growth in GDP per capita, the time series of income growth could be stationary, so we do not include a time trend in our main specification. However, we decided to include it in one of our specifications for robustness (third column of Table 2) to capture possible linear growth trends in the data.

⁴Replication files for our statistical analysis are available at the following repository: <https://github.com/diodz/disasters-economic-freedom>.

TABLE 2
REGRESSION RESULTS, 1979–2010

Dependent Variable	$\Delta \ln$ GDP per capita ($N = 1,490$; Countries = 87)	$\Delta \ln$ GDP per capita ($N = 1,546$; Countries = 90)	$\Delta \ln$ GDP per capita ($N = 1,490$; Countries = 87)
Institutional Variables			
Polity $i, t - 1$	0.0101 (0.0094)	0.0155 (0.010)	0.011 (0.0094)
Size of Government $i, t - 1$	0.000003 (0.0020)		0.000003 (0.002)
Property Rights $i, t - 1$	0.0064** (0.0021)		0.0064** (0.0021)
Sound Money $i, t - 1$	0.0002 (0.0013)		0.0002 (0.0012)
Freedom to Trade Internationally $i, t - 1$	0.0052** (0.0017)		0.0052** (0.0017)
Regulation $i, t - 1$	0.0052* (0.0029)		0.0052 (0.0029)
Economic Freedom Aggregated Index $i, t - 1$		0.0172*** (0.0034)	
GeoMet Disaster Index i, t	-0.0556* (0.0263)	-0.0638* (0.0281)	-0.0556* (0.0263)
Controls			
\ln GDP per capita $i, t - 1$	-0.0986*** (0.0097)	-0.1042*** (0.0100)	-0.0986*** (0.0097)
\ln Population $i, t - 1$	-0.0197 (0.0194)	-0.052*** (0.0195)	-0.0197 (0.0195)
Trade Openness $i, t - 1$	0.0379*** (0.0100)	0.0396*** (0.0104)	0.0379*** (0.0100)
Interest Rate $i, t - 1$	-0.2980* (0.1289)	-0.214 (0.1344)	-0.2980* (0.1290)
Domestic Credit $i, t - 1$	-0.0350*** (0.0072)	-0.0287*** (0.0075)	-0.0350*** (0.0071)
Gross Capital Formation $i, t - 1$	0.0338*** (0.0074)	0.0327*** (0.0075)	0.0338*** (0.0074)
Foreign Direct Investment $i, t - 1$	0.0557* (0.0383)	0.0444 (0.0411)	0.0558 (0.0383)

(Continued)

TABLE 2 (*Continued*)
REGRESSION RESULTS, 1979–2010

Dependent Variable	$\Delta \ln$ GDP per capita	$\Delta \ln$ GDP per capita	$\Delta \ln$ GDP per capita
	($N = 1,490$; Countries = 87)	($N = 1,546$; Countries = 90)	($N = 1,490$; Countries = 87)
ln Inflation $i, t - 1$	−0.0024 (0.0016)	−0.0016 (0.0016)	−0.0024 (0.0016)
Current Account Balance $i, t - 1$	0.005 (0.0284)	0.0286 (0.0298)	0.005 (0.0284)
Time trend			0.0028*** (0.0007)
Adj R squared	0.2834	0.2316	0.2834
F-statistic	15.67	14.988	15.67

NOTES: Robust standard errors reported in parentheses. ***, **, and * denote significance at the levels of 0.1 percent, 1 percent, and 5 percent, respectively. Country and time fixed effects included but not reported. Column (1) uses the economic freedom subindexes, while column (2) uses the aggregated index only. Column (3) includes a time trend.

year divided by the log of the area of the affected country to account for country size. The empirical analysis indicates that a country with a higher quality of institutions, which is expressed in the economic freedom indexes as higher values of all variables except for size of government, are more favorable for economic growth. Countries with higher economic freedom will be able to respond more effectively to an exogenous shock to capital and labor, achieving higher economic growth immediately after a natural disaster.

Our results regarding the disaster index in growth in GDP per capita are in line with Felbermayr and Gröschl (2014), which is expected since we use the same disaster index variable. Quantitatively, looking at our main specification in column (1) of Table 2, since the estimated coefficient for the disaster index is −0.0556 and the sample mean of the index is 0.032, in a year in which the index is equal to the mean, economic growth will be lower by 0.18 percent. Considering a stronger natural disaster (i.e., a disaster that is one standard deviation higher in the GeoMet

index), we can tell by the estimated coefficient that growth in the year of the disaster, *ceteris paribus*, will be 1.1 percent lower.

Concerning the institutional variables from the Fraser Institute, we find significant effects for property rights and freedom to trade internationally at 0.1 percent significance, and borderline significance for the regulation component of the index at the 10 percent level. The property rights component of the index has a sample mean of 5.105, which means that, on average, a country with the mean level of property rights will grow by 3.27 percent, and a 1 point decrease will cause growth to decline to 2.63 percent. Regarding freedom to trade internationally, a country with the sample mean value of 6.520 will have an average yearly growth of 3.39 percent, and a 1 point decrease from the mean level will decrease growth to 2.87 percent. Lastly, regulation has a regression coefficient of 0.0052, which means that based on its sample mean level of 6.224, the yearly growth after a natural disaster is 3.23 percent, and a 1 point decrease will cause growth to decline to 2.72 percent.

In column 3 of Table 2, we include a time trend to account for linear growth in the time series. The coefficient is statistically significant at the 0.1 percent level and positively impacts growth, indicating that countries are on average increasing their economic growth in the data. Comparing our three regression specifications in Table 2: with or without the individual economic freedom components, and with or without a time trend, we find that our main results hold, that is, that economic freedom is relevant and causes a positive and statistically significant effect in economic growth after natural disasters.

Conclusion

Performing a fixed-effects regression clustered at the country level and using the growth of GDP per capita as the dependent variable, we found evidence that natural disasters reduce growth in real GDP per capita in the year that they occur and that institutions, specifically property rights and freedom to trade internationally, reduce their negative impact.

The regression coefficients express the magnitude of the effect of institutional variables on GDP per capita. We find that a 1 point drop in the property rights index causes growth to decline by 0.64 percent

after a natural disaster. Regarding freedom to trade internationally, a 1 point drop decreases growth in income by 0.52 percent. We do not find the level of democracy to be relevant for economic growth after a natural disaster according to the Polity index as its coefficient is not significant in our specifications, nor are the economic freedom components related to the size of government or sound money.

The analysis suggests policy recommendations regarding the importance of institutions. Increasing economic freedom, property rights or freedom to trade internationally attenuates the negative impact of a natural disaster. In the same way, it may improve the indirect positive impact of fiscal spending in economic growth after a disaster.

References

- Acemoglu, D.; Johnson, S.; and Robinson, J. A. (2005) "Institutions as a Fundamental Cause of Long-Run Growth." *Handbook of Economic Growth* 1 (Part A): 385–472.
- Albala-Bertrand, J. M. (1993) *Political Economy of Large Natural Disasters: With Special Reference to Developing Countries*. Oxford: Clarendon Press.
- Allen M. R. et al. (2019) "Technical Summary." In Masson-Delmotte et al. (eds.), *Global Warming at 1.5°C. An IPCC Special Report on the Impacts of Global Warming of 1.5°C above Pre-Industrial Levels . . . and Efforts to Eradicate Poverty*. Geneva: World Meteorological Organization. Available at www.ipcc.ch/site/assets/uploads/sites/2/2019/02/SR15_TS_High_Res.pdf.
- Barone, G., and Mocetti, S. (2014) "Natural Disasters, Growth and Institutions: A Tale of Two Earthquakes." *Journal of Urban Economics* 84 (November): 52–66.
- Bastiat, F. ([1850] 1964) "What Is Seen and What Is Not Seen." In *Selected Essays on Political Economy*, chap. 1. Edited by G. B. de Huszar; translated from the French by S. Cain. Irvington-on-Hudson, N.Y.: Foundation for Economic Education.
- Benson, C., and Clay, E. J. (2004) *Understanding the Economic and Financial Impacts of Natural Disasters*. Washington: World Bank.
- Boudreaux, C. J.; Escaleras, M. P.; and Skidmore, M. (2019) "Natural Disasters and Entrepreneurship Activity." *Economics Letters* 182: 82–85.

- Centre for Research on the Epidemiology of Disasters (2015) EM-DAT: The OFDA/CRED International Disaster Database. Brussels, Belgium: Catholic University of Leuven.
- Checherita-Westphal, C., and Rother, P. (2012) “The Impact of High Government Debt on Economic Growth and Its Channels: An Empirical Investigation for the Euro Area.” *European Economic Review* 56 (7): 1392–405.
- Fellbermayr, G., and Gröschl, J. (2014) “Naturally Negative: The Growth Effects of Natural Disasters.” *Journal of Development Economics* 111 (November): 92–106.
- Gwartney, J.; Lawson, R.; Hall, J.; and Murphy, R. (2018) *Economic Freedom of the World: 2018 Annual Report*. Vancouver, B.C.: Fraser Institute.
- Hallegatte, S., and Dumas, P. (2009) “Can Natural Disasters Have Positive Consequences? Investigating the Role of Embodied Technical Change.” *Ecological Economics* 68 (3): 777–86.
- IPCC (2001) “Summary for Policymakers.” In R. T. Watson et al. (eds.), *Climate Change 2001: Synthesis Report*. New York: Cambridge University Press for the Intergovernmental Panel on Climate Change.
- Kahn, M. E. (2005) “The Death Toll from Natural Disasters: The Role of Income, Geography, and Institutions.” *Review of Economics and Statistics* 87 (2): 271–84.
- Larroulet, C., and Couyoumdjian, J. P. (2009) “Entrepreneurship and Growth: A Latin American Paradox?” *Independent Review* 14 (1): 81–100.
- Monllor, J., and Murphy, P. J. (2017) “Natural Disasters, Entrepreneurship, and Creation after Destruction: A Conceptual Approach.” *International Journal of Entrepreneurial Behavior and Research* 23 (4): 618–37.
- North, D. C. (1991) “Institutions.” *Journal of Economic Perspectives* 5 (1): 97–112.
- Raschky, P. A. (2008) “Institutions and the Losses From Natural Disasters.” *Natural Hazards and Earth System Sciences* 8 (4): 627–34.
- Sanderson, D., and Sharma, A. (2016) *World Disasters Report 2016: Resilience: Saving Lives Today, Investing for Tomorrow*. Geneva, Switzerland: International Federation of Red Cross and Red Crescent Societies (IFRC).

- Skidmore, M., and Taya, H. (2002) “Do Natural Disasters Promote Long-Run Growth?” *Economic Inquiry* 40 (4): 664–87.
- Strobl, E. (2012) “The Economic Growth Impact of Natural Disasters in Developing Countries.” *Journal of Development Economics* 97 (1): 130–41.
- Talbott, J., and Roll, R. (2001) “Why Many Developing Countries Just Aren’t.” The Anderson School at UCLA, Finance Working Paper No. 19–01.
- Van Aalst, M. K. (2006) “The Impacts of Climate Change on the Risk of Natural Disasters.” *Disasters* 30 (1): 5–18.
- Wallace, J. M.; Held, I. M.; Thompson, D. W.; Trenberth, K. E.; and Walsh, J. E. (2014) “Global Warming and Winter Weather.” *Science* 343 (6172): 729–30.

ECONOMIC WELL-BEING UNDER PLAN VERSUS MARKET: THE CASE OF ESTONIA AND FINLAND

Anna Bocharnikova

Historically, Estonia and Finland shared similar cultural, political, and economic characteristics apart from the almost 50 years of Soviet rule (1944–1991) in Estonia (Raun 1987). During that time, Estonia lacked political freedom and became a centrally planned economy. Meanwhile, Finland remained a democracy with a market economy. After Estonia gained independence in 1991, the resemblance of the institutional landscape in the two neighboring countries started to be restored.

This article investigates the dynamics of individual economic well-being in Estonia and Finland over three periods: (1) 1923–1938, when both countries were similarly situated; (2) 1960–1988, during which Estonia was under Soviet control; and (3) 1992–2018, after Estonian independence. Economic well-being is calculated using the purchasing power of wages in terms of the affordability of a minimal food basket.

The results show that, in 1938, the purchasing power of wages in Estonia was 4 percent lower than in Finland; in 1988, it was 42 percent

Cato Journal, Vol. 41, No. 1 (Winter 2021). Copyright © Cato Institute. All rights reserved. DOI:10.36009/CJ.41.1.5.

Anna Bocharnikova is a graduate student at the University of Glasgow. She thanks Andrei Illarionov for supervising her research and the enormous contribution to it, Alexander Hammond for his edits of earlier drafts of this article, and James Dorn for his advice and edits. The author also acknowledges the assistance of staff members at the University of Tartu's Library, the offices of Statistics Finland and Statistics Estonia, the National Library of Estonia, the University of Glasgow, and especially, the Research Institute of the Finnish Economy.

lower; and, by 2018, the gap had fallen to 17 percent. Consequently, as measured by the purchasing power of wages, well-being in Estonia and Finland was similar before the Soviet occupation, widely diverged during Soviet rule, and converged after Estonian independence, with the transition from plan to market.

Background

There has been a long-standing debate on the economic impacts of central planning versus free markets. Traditionally, the most illustrative examples are countries that differ only in their economic and political institutions—such as East and West Germany, or North and South Korea. Such examples offer an opportunity to examine the effects of two different systems in similar environments. While some pairs are more popular for analysis, some are less obvious, especially if their characteristics became too dissimilar after the introduction of different policies. One of these less obvious examples is Estonia and Finland.

Estonia is often compared to other post-Soviet states, while Finland is normally examined within the Nordic countries. However, historically, they share more similarities than it first seems. In addition to geographical proximity and cultural affinities, such as religious (Lutheranism) and linguistic (Uralic languages), Estonia and Finland had very similar political and economic developments before the Soviet Union occupied Estonia (see *Nordic Estonia*).

The two neighboring countries had a similar legacy after the Scandinavian rule and the rule of the Russian Empire. They both underwent wars of liberation and gained independence in the early 20th century. As two new sovereign countries, both became parliamentary democracies under the rule of law and remained market economies.

Similar agrarian reform in Estonia and Finland led to the formation of a family farming system that played a big role in the economies of both countries (Lapping 1993; Jörgensen 2006). In 1920 and 1921, Finland and Estonia, respectively, became members of the League of Nations and developed strong economic and political connections with the international community. Estonia and Finland were relatively late with industrialization. Indeed, until the 1930s, both were predominantly agrarian economies. In the early 1930s, both nations suffered from the Great Depression but had

their economies expand anew in the late 1930s, with a growing industrial sector. Under the Russian Empire, Estonia and Finland had strong connections with the Russian market, but the Soviet regime disturbed those linkages by 1924. Both countries then reoriented toward the West and began actively trading with European nations.

There is a common opinion that Estonia never became a Nordic country because it was occupied by the USSR when the Nordic Council was formed. In fact, the Nordic identity in modern Estonia is still very strong. Toomas Hendrik Ilves, Estonian foreign minister, who later served as a president of Estonia, delivered an influential speech, “Estonia as a Nordic Country,” in 1999 (Ilves 1999). In 2015, the Estonian prime minister Taavi Rõivas defined Estonia as a “New Nordic Country.” In the same year, the Swedish ambassador to Estonia, Anders Ljunggren, stated that Estonia would have been considered a Nordic country by the other Nordic countries (Aavik 2015).

As advocates for the Nordic affiliation of Estonia often pointed out, more than a thousand years of Estonian and Finnish shared history was eclipsed by about 50 years of the Soviet occupation. The similarities between the two countries became disturbed during that time, when Estonia experienced a totalitarian regime and centrally planned economy.

The Soviet rule in Estonia began in 1940, was interrupted in 1941 during the German occupation of Estonia, reestablished in 1944, and lasted until 1991. Already in 1940–1941, the Soviet Union nationalized all land in Estonia, banks, and large industrial enterprises. When the USSR regained control over Estonia in 1944, it quickly and fully integrated the country into the Soviet centrally planned structure. The Estonian economic landscape changed dramatically. Its trade connections with non-Soviet markets were cut; private ownership of the means of production was liquidated; and the economy was refocused toward the extraction of natural resources, much of which was sent to other Soviet republics. Political institutions experienced a major change as well: the democratic regime was replaced by a one-party system with almost unlimited political power.

In the Soviet Union, analysis of the Estonian transition to a planned economy had a highly ideological character and was described as a big success. However, it was widely criticized later and is still a very debatable matter. Whether the Soviet period helped or hindered Estonia in achieving its economic policy aims, it is clear that

Estonian development experienced a radical twist under Soviet central planning, compared to what it was before and relative to its neighbor Finland. Such a change cannot but affect normal people's lives.

Meanwhile, Finland defended its independence and remained a market economy with private property rights, the rule of law, limited power of the executive, and relatively free trade. In 1973, Finland signed a free-trade agreement with the European Community. In the late 1980s, like other Nordic countries, Finland eased economic regulations, privatized some state-owned enterprises, and altered tax policy. In 1995, the country joined the European Union and adopted the euro. We can never know if Estonia would have done the same if it had remained independent from the Soviet Union but similar to Finnish developments, which were popular among other European countries as well.

When Estonia gained independence from the Soviet Union during 1988–1991, its institutional structure began to converge with Finland's. Estonia became a parliamentary democracy, the size of the Estonian government was reduced dramatically, subsidies and state-owned enterprises were abolished, the economy reopened for international trade and foreign direct investment, and macroeconomic stability was achieved via a strict monetary policy and balanced budgets. Estonia left almost no sign of the Soviet system and joined the European Union in 2004.

This article examines how the economic well-being of the Estonian people was affected by the institutional change during and after the Soviet regime compared to Finland. The main contribution is the use of hard-to-find statistical data to construct relative measures of well-being using an indicator for the purchasing power of wages. As such, this article should contribute to the long-lasting debate over the efficacy of markets versus planning by providing new data for a case of two similar countries with different experiences.

Assuming that changes in government policies affect individual economic well-being, the economic situation in Estonia and Finland is expected (1) to be similar when the countries had similar economic and political institutions; (2) to differ widely during the Soviet occupation of Estonia; and (3) to start converging after the Estonian independence.

Research Design

Such measures of national well-being as gross domestic product (GDP) per capita do not cover the entire period of interest, but Estonian and Finnish historical statistics contain enough information to construct an indicator to grasp it. Data on retail prices and average wages provide information about purchasing power, since they can help in determining the quantity of goods and services that can be bought with a wage unit. The purchasing power of wages (PPW) is calculated at a specific point in time for a specific place to capture people's economic well-being. This procedure is used to determine how individual economic well-being in Estonia and Finland has been changing over time.

PPW is measured using retail prices of basic foodstuffs over the entire period of interest in both countries. The higher the PPW, the more things of an average price can be purchased, and the freer people are financially. While the conceptual definition of PPW includes goods and services, the *operational* definition used in this study means the quantity of foodstuffs that can be bought with a wage unit.

PPW was found according to the minimal consumer basket of food normally used for minimum wage legislation. The minimal food basket (Table 1) respects the minimum amount of calories required in order to function normally—a standard of 2,100 to 2,300 calories, according to the Food and Agriculture Organization and World Health Organization (MONSTAT 2018).

Since the exact minimal baskets for Finland and Estonia were not available, and the methodology does not change much internationally, the model used is taken from the Russian Federation Labour Code (2012). Although there are probably some differences in consumption habits among the countries, the format of the minimal basket is suitable to make conclusions due to similar climates, agriculture trends, and a comparably similar gastro-cultural background.

The PPW indicator will show how many minimal baskets an average wage in each country could buy in different points in time. The cost of the basket is calculated from the average retail prices and minimal required consumption per year. Based on the cost of the basket, PPW was counted using annual average wages to establish how many baskets per year can be bought for an average wage in each nation.

PPW is calculated for three base years: 1938, when both countries had similar institutions; 1988, after almost 50 years of the Soviet

TABLE 1
MINIMAL FOOD BASKET

Description	Measurement	Consumption (average for one person per annum)
Bread Products	kg	126.5
Potatoes	kg	100.4
Vegetables and Gourds	kg	114.6
Fruits	kg	60
Sugar	kg	23.8
Meat Products	kg	58.6
Fish and Seafood	kg	18.5
Milk	kg	290
Eggs	piece	210
Butter, Margarine, and Other Fats	kg	11

SOURCE: See Statistical Appendixes.

occupation in Estonia; and 2018, after the Estonian transition to a democracy and market economy. The Estonian data were primarily drawn from the statistics provided by the National Bureau of Statistics before 1940, the Central Board of Statistics of the Estonian SSR until 1989, and Statistics Estonia for current data. The Finnish data mostly come from statistical yearbooks published by Statistics Finland (Finnish: *Tilastokeskus*; named the Central Statistical Agency before 1971). Also, *European Historical Statistics, 1750–1970* by B. R. Mitchell (1975) was used for this article. Finally, some data for the 2018 base year were extracted from the OECD statistics.

Previous Research

A considerable amount of literature has been published to identify the difference in the economic impacts of central planning and free markets. The most common examples are cases such as East and West Germany, as mentioned earlier. There is also much research on how similar countries change under different regimes in terms of economic well-being, culture, and outlooks.

Normally, the resemblance between Finland and Estonia is described in the context of language and culture. However, reflecting on Estonia after the fall of the Soviet Union, academics have compared Estonia and Finland based on their socioeconomic performance. For instance, in 1993, the Research Institute of the Finnish Economy, jointly with the Institute of Economics of Estonian Academy of Science and the Government Institute for Economic Research, published *Estonia and Finland: A Retrospective Socioeconomic Comparison* (Lugus and Vartia 1993). This collaborative work of Estonian and Finnish researchers compared the demographics, consumption, housing, natural recourses, and other aspects for the two countries. The book highlighted the socioeconomic affinity of Estonia and Finland in 1938, before the Soviet regime in Estonia, and stressed the difference in 1988, after almost 50 years of different development models—namely, central planning versus a market economy. That perspective raised the question of the significant divergence in purchasing power in the two countries that emerged after Estonia became part of the Soviet bloc. Yet, very little information was provided on the calculations made.

Limitations

Despite a long history of recordkeeping in both countries, the earlier statistics do not always provide all the necessary information, so this article also uses information from unofficial sources, such as newspapers from the late 1930s. Because of the dearth of data in the earlier periods, an accurate comparison of wages was possible by using the average wages of male workers in industrial enterprises (Hagfors and Kuus 1993).

Although Estonia and Finland differ in some policies, this study is primarily interested in examining how the transition from plan to market in Estonia affected economic well-being relative to Finland. What matters for this study is the existence or absence of the rule of law, trade openness, competitive political participation, private ownership—and whether prices are determined by supply and demand or fixed by government.

Results

We can now move to examine the three periods of interest: the interwar period, the postwar period, and the post–Cold War period.

Interwar Period

After the Russian Revolution in 1917, Estonia and Finland gained their independence from the Russian Empire, which was an industrial capitalist economy. At that time, the Baltic cousins were similarly situated:

- For both nations, their trade relationship with Russia collapsed and they reoriented the markets toward Europe;
- Both recognized private property and adhered to the rule of law;
- Both were parliamentary democracies;
- Both economies largely focused on agriculture.

This article calculates PPW in 1938 in both countries. According to the calculations, the purchasing power of an Estonian worker's wage in 1938 was 96 percent of the average Finnish worker, or almost same (Table 2).

Figure 1 shows the course of PPW over time using food price and wage indexes from historical statistics. Estonian performance was typically close behind the Finnish experience and even overtook it in 1932–1936 by an average of 8 percent. Thus, overall development in both countries went hand in hand.

Postwar Period

Since Finland had defended its independence, the country's economic progress continued in a similar matter. For example, Finland

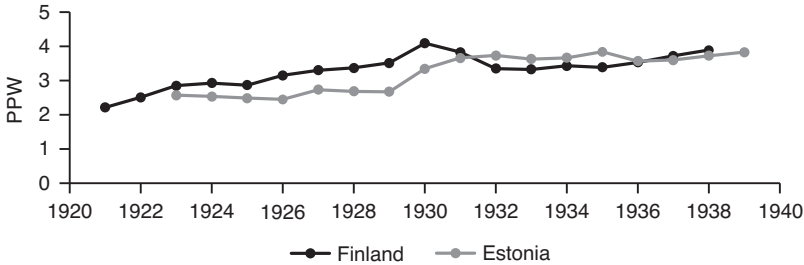
- Remained a democracy with limited power of the executive;
- Continued to be a market economy;
- Recognized and guaranteed private property;
- Remained oriented to international trade and signed a free-trade agreement with the European Community in 1973.

TABLE 2
PURCHASING POWER IN ESTONIA AND FINLAND, 1938

	Estonia (kroons)	Finland (markkas)
Annual Cost of the Food Basket	257.48	4,044.93
Annual Average Wage	959.70	15,738.87
Purchasing Power of Wages	3.73	3.89

SOURCE: See Statistical Appendixes.

FIGURE 1
PURCHASING POWER OF WAGES IN ESTONIA, 1923–1939,
AND FINLAND, 1921–1938



In contrast, after almost 50 years in the Soviet bloc, the Estonian economy had experienced a series of changes, such as:

- Lack of competitive political participation;
- Controlled economy with a focus on natural resource extraction;
- Liquidation of private ownership of the means of production;
- Price fixing;
- An abrupt end of international trade.

After these changes, the geographical proximity did not match the economic distance between Estonia and Finland. PPW in the two countries diverged substantially compared to where it had been before the Soviet occupation of Estonia. In 1988, the Estonian PPW, with regard to foodstuffs, was 58 percent of the Finnish counterpart, falling by 38 percent since 1938 (Table 3).

Using food price and wage indexes to measure differences in PPW in Estonia and Finland in the postwar period, Figure 2 shows that the situation in Estonia was significantly worse than in Finland.

Since purchasing power is measured with regard to food, it is important to remember that prices in Finland reflected market prices, based on supply and demand, while prices in Soviet Estonia were fixed by the central government and kept artificially low. Therefore, Estonians' food affordability during Soviet occupation is distorted by false prices and PPW appears more promising than the overall financial situation during that period (Hagfors and Kuus 1993). The food scarcity, which had been increasing at the turn of the century, should also be taken into account (Ellman 2000).

TABLE 3
PURCHASING POWER OF WAGES IN ESTONIA
AND FINLAND IN 1988

	Estonia (roubles)	Finland (markkas)
Annual Cost of the Food Basket	550.76	7,576.41
Annual Average Wage	2,748	64,965
Purchasing Power of Wages	4.99	8.57

SOURCE: See Statistical Appendixes.

Post–Cold War Period

After the fall of the Soviet Union, Finland still had the same characteristics as described above, and Estonia

- Became a democracy again and limited the power of the government;
- Transitioned to a market economy;
- Opened for trade and foreign direct investment;
- Joined the EU single market in 2004.

When Estonia underwent the reforms, its economic performance skyrocketed. Starting with two minimal baskets per year in 1992, after the fall of the USSR, Estonia’s purchasing power reached almost 12 baskets by 2018 (Table 4). It was 22 percent of the Finnish

FIGURE 2
PURCHASING POWER OF WAGES IN ESTONIA, 1960–1988,
AND FINLAND, 1955–1989



TABLE 4
PURCHASING POWER OF WAGES IN
ESTONIA AND FINLAND, 2018

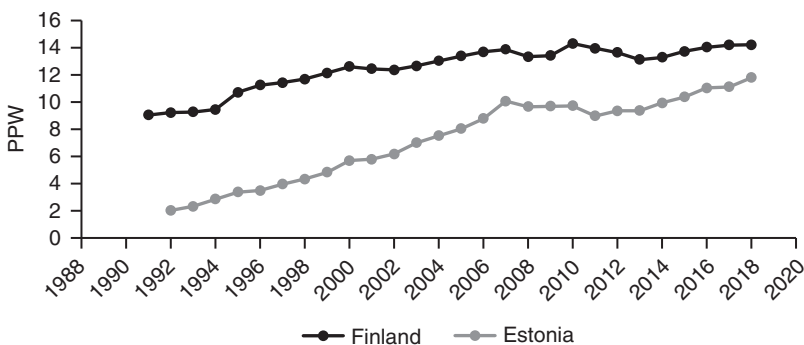
	Estonia (euros)	Finland (euros)
Annual Cost of the Food Basket	1,426.75	2,967.89
Annual Average Wage	16,795	42,122
Purchasing Power of Wages	11.77	14.19

SOURCE: See Statistical Appendixes.

figure in 1992 and increased to 83 percent of the Finnish performance in 2018 (Figure 3).

Arguably, the Estonian economy experienced a J curve after gaining independence (Bremmer 2006, Laar 2008). Since Estonia inherited a collapsing economic “dinosaur,” the country needed to carry out often unpopular but necessary policies to get rid of it. Estonia had a fresh start: it instituted a comprehensive monetary reform by adopting a currency board, cutting spending, eliminating subsidies, abolishing state-owned enterprises, renewing international trade, and opening the market to foreign investment. This was bitter medicine, but the reforms helped Estonia achieve robust economic growth. Indeed, the wide gap between Estonian and Finnish PPW has been shrinking rapidly, unlike during Soviet occupation.

FIGURE 3
PURCHASING POWER OF WAGES IN ESTONIA, 1992–2018,
AND FINLAND, 1991–2018

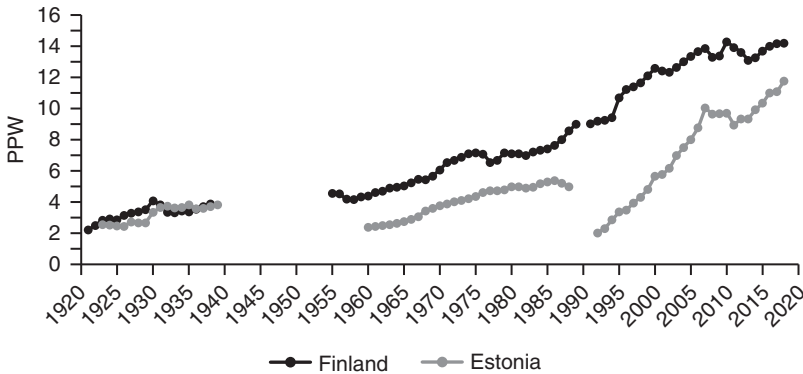


Conclusion

The aim of this article was to examine how the Estonian people’s economic well-being was affected under central planning versus a market economy—compared to Finland, which maintained its market-oriented and democratic system over the entire period of consideration. Figure 4 shows that PPW in Estonia and Finland (1) developed similarly when both countries shared a similar institutional structure; (2) had a considerable gap during the Soviet occupation of Estonia, when the country experienced central planning and a one-party state; and (3) started to converge to the Finnish PPW when Estonia began to undertake structural changes similar to those of Finland.

When Estonia was a part of the Soviet Union, it lagged behind Finland for nearly 30 years. It is not surprising that, given the state of the Soviet economy before the collapse, economic well-being in Estonia suffered and diverged from the level found in Finland. The decrease in Estonia’s PPW during the Soviet occupation meant that, once the transition from plan to market began, economic performance would see a sharp increase and converge toward Finland’s PPW. Indeed, after the transition to a market economy, Estonian PPW went through a J curve and started to rapidly rise. Although Estonia’s economic performance still falls short of Finland’s, the upward trend is promising. In 1992, Estonia’s PPW was only 22 percent of Finland’s, but by 2018, it was 83 percent.

FIGURE 4
PURCHASING POWER OF WAGES IN ESTONIA AND FINLAND



From the results of this study, it is evident that institutions supporting freer markets and greater political freedom have benefited the well-being of the Estonian people.

Statistical Appendixes

Appendix A discusses the sources of retail prices for foodstuffs and calculation of the cost of the food basket. Appendix B covers the wage data. Appendix C provides a table showing PPW in Estonia and Finland from 1932 to 2018.

Appendix A: Cost of Food

The average retail prices in Estonia and Finland (Appendix Table 1) are extracted from official statistical sources, except for the price of fruit in 1938 and margarine in 2018 in Estonia. While the prices for eggs and potatoes are fairly straightforward, several products, such as vegetables and gourds, include different products' prices and are presented as an average. The prices are presented in local currencies for each given year and in the same units as in the minimal basket. The annual cost was calculated based on consumption per annum.

Interwar Period

Base Year, 1938. Average retail prices in Finland were taken from the *Statistical Yearbook of Finland* (Central Statistical Agency 1938). Average retail prices in Estonia were extracted from the *Monthly Estonian Statistics* (National Bureau of Statistics 1938). The price of fruit was taken from several newspapers from the late 1930s (*Daily* newspaper 1938; *New Estonia* newspaper 1938).

Estonia, 1923–1939. The index of food prices (Appendix Table 2) was gathered from different issues of the *Monthly Statistics* (National Bureau of Statistics 1937, 1938, 1939, 1940) and was applied for the cost of the 1938 basket to find the annual cost of food in the years between 1923 and 1939.

Finland, 1921–1938. In the *Finnish Statistical Yearbook* for 1939, there are two food price indexes: one for 1921 to 1936, with 1914 as the base year, and another from 1935 to 1938, with 1935 as the base year (Central Statistical Agency 1939: 345). The second index includes 1939, the year that has the cost of the food basket, and both indexes have two matching years, 1935 and 1936. The index in

APPENDIX TABLE 1				
AVERAGE RETAIL PRICES AND ANNUAL COST OF FOOD				
Description	Estonia (kroons)		Finland (markkas)	
	Retail Price	Cost per Year	Retail Price	Cost per Year
Total Cost of Food per Year		257.48		4,044.93
Bread Products				
Rye bread	0.20		4.51	
Wheat bread	0.39		8.97	
Average	0.30	37.95	6.74	852.61
Potatoes				
Average	0.04	4.02	0.74	74.30
Vegetables and Gourds				
Beets	0.12		2.03	
Cabbage	0.10		1.94	
Carrots	0.12		1.96	
Average	0.11	12.61	1.98	217.74
Fruits				
Bananas	1.75		n/a	
Oranges	1.12		n/a	
Apples	0.32		n/a	
Watermelons	1.50		n/a	
Melons	1.40		n/a	
Mixed fruits	n/a		14.12	
Average	1.13	67.60	14.12	853.20
Sugar				
Average	0.47	11.19	6.06	144.23
Meat Products				
Pork, 3 types	0.78		n/a	
	0.93		n/a	
	1.00		n/a	
Average	0.90	52.74	19.29	1130.39
Fish and Seafood				
Pike	0.48		10.52	
Perch	0.39		7.14	
Herring, salted	0.55		4.00	
Average	0.47	8.70	7.22	133.57

APPENDIX TABLE 1 (*Continued*)
AVERAGE RETAIL PRICES AND ANNUAL COST OF FOOD

Description	Estonia (kroons)		Finland (markkas)	
	Retail Price	Cost per Year	Retail Price	Cost per Year
Milk				
Average	0.12	34.80	0.56	162.4
Eggs				
Average	0.06	12.60	1.07	224.7
Butter, Margarine, and Other Fats				
Butter	1.80		30.80	
Margarine	0.98		14.99	
Average	1.39	15.29	22.90	251.79

APPENDIX TABLE 2
FOOD PRICE INDEX AND ANNUAL COST OF
FOOD IN ESTONIA, 1923–1939

Year	Food Price Index	Annual Price of the Food Basket (kroons)
1923	115	213.72
1924	106	196.99
1925	118	219.29
1926	118	219.29
1927	112	208.14
1928	120	223.01
1929	126	234.16
1930	103	191.42
1931	90	167.26
1932	80	148.67
1933	77	143.10
1934	76	141.24
1935	77	143.10
1936	89	165.40
1937	96	178.41
1938	101	257.48
1939	103	191.42

APPENDIX TABLE 3
FOOD PRICE INDEX AND ANNUAL COST OF
FOOD IN FINLAND, 1921–1938

Year	Food Price Index	Annual Price of the Food Basket (markkas)
1921	127	5,154.42
1922	116	4,695.4
1923	110	4,435.68
1924	111	4,491.57
1925	116	4,711.42
1926	113	4,552.39
1927	113	4,582.39
1928	117	4,726.63
1929	114	4,616.91
1930	99	3,991.05
1931	88	3,571.07
1932	91	3,687.36
1933	91	3,673.80
1934	89	3,599.01
1935	93	3,745.30
1936	92	3,708.73
1937	99	4,007.47
1938	100	4,044.93

Appendix Table 3 is calculated from these data and applied to the 1938 basket to find the cost of food in the 1923–1938 period.

Postwar Period

Base Year, 1988. Appendix Table 4 shows average retail prices in 1988 for Estonia and Finland. Retail prices in Estonia are drawn from the *National Economy of the Estonian SSR* in 1988 (State Statistical Committee of the Estonian SSR 1989: 264), and in Finland from the 1990 *Statistical Yearbook of Finland* (Central Statistical Agency 1990).

Estonia, 1960–1988. In Estonia’s 1987 statistical yearbook, *National Economy of the Estonian SSR*, there is a price index for 1986–1988, with a 1980 base year (State Statistical Committee of the Estonian SSR 1988: 280). The index for 1961 to 1988 is calculated from the price index for 1975, the 1980–1984 price index (with a base year 1970=100), and the price index for each fifth year between 1960 and 1975 and 1976 (with 1960=100) (State Statistical

APPENDIX TABLE 4
AVERAGE RETAIL PRICES AND ANNUAL COST OF FOOD

Description	Estonia (roubles)		Finland (markkas)	
	Retail Price	Cost per Year	Retail Price	Cost per Year
Total Cost of Food per Year		550.76		7,576.41
Bread Products				
Rye bread	n/a		13.26	
French bread	n/a		15.85	
Average	0.47	59.46	14.55	1,841.21
Potatoes				
Average	0.13	13.05	4.50	451.80
Vegetables and Gourds				
Carrots	n/a		9.65	
Onions	n/a		9.09	
Average	0.71	81.37	9.37	1,073.80
Fruits				
Average	2.05	123.00	6.23	373.80
Sugar				
Average	0.80	19.04	8.00	190.40
Meat Products				
Meat and chicken	1.75		n/a	
Sausages	2.65		31.66	
Beef, minced	n/a		43.83	
Pork flank	n/a		30.75	
Average	2.20	128.92	35.41	2,075.22
Fish and Seafood				
Herring	1.30	24.05	8.20	149.85
Milk				
Average	0.25	72.50	3.53	1,023.70
Eggs				
Average	0.01	2.10	1.070	224.70
Butter, Margarine, and Other Fats				
Butter	3.44		19.15	
Margarine	1.52		12.12	
Average	2.48	27.28	15.63	171.93

Committee of the Estonian SSR 1977: 217). The data for the interim missing years were calculated using the percentage of average growth. The resulting index was applied for the price of the food basket in 1988 (Appendix Table 5).

Finland, 1955–1989. Appendix Table 6 shows the annual costs of the food basket based on the cost of the food basket in 1988 and the CPI inflation rate, with the base year 2015 taken from the OECD data.

APPENDIX TABLE 5
FOOD PRICE INDEX AND ANNUAL COST OF
FOOD IN ESTONIA, 1960–1988

Year	Food Price Index	Annual Price of the Food Basket (roubles)
1960	90	409.34
1961	91	413.36
1962	92	417.37
1963	93	421.38
1964	93	425.39
1965	94	429.40
1966	94	429.65
1967	94	429.89
1968	94	430.14
1969	95	430.38
1970	95	430.63
1971	95	432.09
1972	95	433.56
1973	96	435.02
1974	96	436.49
1975	96	437.95
1976	96	437.95
1977	97	442.26
1978	98	446.56
1979	99	450.87
1980	100	455.18
1981	102	462.50
1982	106	481.01
1983	106	484.03
1984	106	482.31
1985	107	486.95
1986	108	491.59
1987	116	528.00
1988	121	550.76

APPENDIX TABLE 6
FOOD PRICE INDEX AND ANNUAL COST OF
FOOD IN FINLAND, 1955–1989

Year	Inflation (CPI), Food 2015 = 100	Annual Price of the Food Basket (markkas)
1955	5.7	620.41
1956	6.5	708.39
1957	7.3	801.93
1958	7.7	845.13
1959	7.8	854.07
1960	8.1	890.52
1961	8.3	907.71
1962	8.6	946.91
1963	9.1	995.04
1964	10.2	1,114.69
1965	10.8	1,185.52
1966	11.2	1,225.41
1967	11.7	1,283.50
1968	13.0	1,428.02
1969	13.4	1,472.98
1970	13.6	1,494.39
1971	14.2	1,558.62
1972	15.5	1,705.57
1973	17.4	1,909.09
1974	20.2	2,213.65
1975	24.3	2,671.13
1976	28.3	3,107.01
1977	33.3	3,651.03
1978	34.8	3,813.81
1979	36.2	3,973.54
1980	40.9	4,484.69
1981	46.1	5,061.51
1982	51.8	5,686.23
1983	55.4	6,072.21
1984	59.6	6,537.06
1985	63.9	7,006.01
1986	66.3	7,277.49
1987	67.9	7,447.97
1988	69.1	7,576.41
1989	71.5	7,849.06

Post–Cold War Period

Base Year, 2018. Appendix Table 7 shows the average retail prices for foodstuffs in Estonia and Finland in 2018. Average retail prices for Estonia are published on the website of the Estonian Institute of Economic Research. The margarine and butter prices are taken from the website of Estonian supermarket Selver Heamõte. The average retail prices for foodstuffs in Finland in 2018 are available on the Statistics Finland website.

Estonia, 1991–2018. The government agency Statistics Estonia provides the food cost index for 1998–2015, with 2015 as the base year. The food index can be created between 1991–1997 and 2016–2018 by using two tables of CPI changes in relation to previous years published on the agency’s website. Appendix Table 8 shows the food price index in Estonia for 1991–2018.

Finland, 1991–2018. Appendix Table 9 shows the food price index in Finland for 1991–2018. Data for the CPI inflation rate (food) in 1991–2015 and the percentage change on the same period of the previous year for the 2016–2018 period were drawn from the OECD data and applied for the annual cost of the food basket in 2018.

Appendix B: Average Wages

Interwar Period

Due to the lack of data, an accurate comparison of wages in the interwar period was only possible by using the average wages of male workers in large industrial enterprises in both countries.

Wages in Estonia. In *Statistics Estonia Monthly*, No. 223, June 1940, there are data on average hourly wages of male industrial workers from 1927 through 1939 (National Bureau of Statistics 1940). Wages in 1923–1926 were counted according to the nominal index of wages (National Bureau of Statistics 1940). Also, there are statistics on working-day duration for each year from 1932 through 1939; and, as working hours did not change much over time, the working hours for 1927–1931 are the same as in 1932. Appendix Table 10 shows the annual average wages of male workers in large industrial enterprise in Estonia over the 1923–1939 period.

APPENDIX TABLE 7
AVERAGE RETAIL PRICES AND ANNUAL COST OF FOOD

Description	Estonia (euros)		Finland (euros)	
	Retail Price	Cost per Year	Retail Price	Cost per Year
Total Cost of Food per Year		1,724.41		4,111.04
Bread Products				
Average	1.79	226.44	5.59	707.14
Potatoes				
Average	0.36	36.14	0.91	91.36
Vegetables and Gourds				
Tomatoes	1.67		3.05	
Cucumbers	1.22		2.68	
Carrots	0.55		1.98	
Cabbages	0.46		3.47	
Average	0.98	111.74	2.80	320.31
Fruits				
Apples	1.21	72.60	2.42	167.70
Sugar				
Average	0.80	19.04	0.83	19.75
Meat Products				
Average	5.42	317.61	8.25	483.45
Fish and Seafood				
Salmon	12.33		21.05	
Herring	3.54		20.93	
Trout	13.62		21.33	
Average	9.83	181.86	21.10	390.35
Milk				
Cottage cheese	4.36		6.95	
Milk	0.58		0.99	
Cream cheese	2.26		10.75	
Average	2.40	696.00	6.23	1,806.70
Eggs				
Average	0.01	2.10	0.31	65.10
Butter, Margarine, and Other Fats				
Butter	7.89		6.06	
Margarine	3.18		4.70	
Average	5.54	60.89	5.38	59.18

APPENDIX TABLE 8
FOOD PRICE INDEX AND ANNUAL COST OF
FOOD IN ESTONIA, 1991–2018

Year	Food Price Index 1991=100	Annual Price of the Food Basket (euros)
1992	1,021	251.2
1993	1,737	427.3
1994	2,293	564.1
1995	2,666	656.0
1996	3,231	795.1
1997	3,406	838.1
1998	3,610	888.4
1999	3,473	854.7
2000	3,557	875.2
2001	3,851	947.7
2002	3,968	976.3
2003	3,899	959.5
2004	4,063	999.7
2005	4,207	1,035.1
2006	4,419	1,087.4
2007	4,832	1,188.9
2008	5,517	1,357.6
2009	5,295	1,303.0
2010	5,456	1,342.5
2011	5,984	1,472.5
2012	6,214	1,529.0
2013	6,471	1,592.4
2014	6,473	1,592.7
2015	6,438	1,584.1
2016	6,438	1,584.1
2017	6,805	1,674.4
2018	7,015	1,726.3

APPENDIX TABLE 9
FOOD PRICE INDEX AND ANNUAL COST OF
FOOD IN FINLAND, 1991–2018

Year	Inflation (CPI), Food 2015 = 100	Annual Price of the Food Basket (euros)
1991	76.3	3,141.13
1992	76.2	3,138.69
1993	75.8	3,118.74
1994	75.9	3,124.72
1995	69.9	2,877.43
1996	68.9	2,836.02
1997	69.2	2,848.08
1998	70.5	2,903.45
1999	70.5	2,902.61
2000	71.3	2,933.58
2001	74.4	3,063.52
2002	76.6	3,151.72
2003	77.0	3,170.62
2004	77.6	3,195.24
2005	78.1	3,213.31
2006	79.1	3,257.28
2007	80.8	3,324.71
2008	87.7	3,610.55
2009	89.5	3,683.49
2010	86.4	3,557.58
2011	91.9	3,781.10
2012	96.6	3,976.19
2013	101.7	4,188.41
2014	101.9	4,194.72
2015	100.0	4,116.48
2016	98.9	4,071.03
2017	98.0	4,034.39
2018	99.9	4,111.04

APPENDIX TABLE 10
ANNUAL AVERAGE WAGES OF MALE WORKERS IN LARGE
INDUSTRIAL ENTERPRISES IN ESTONIA, 1923–1939

Year	Nominal Index of Wages	Hourly Wage, cents	Working Hours per Day	Annual Average Wage (kroons)
1923	91.1			752.92
1924	82.9			685.15
1925	90.5			747.96
1926	89.2			737.22
1927	94.4	34.4	8.40	780.19
1928	97.9	36.2	8.40	821.02
1929	102.1	37.9	8.40	859.57
1930	104.2	38.7	8.40	877.72
1931	101.3	37.0	8.40	839.16
1932	95.5	34.5	8.18	761.50
1933	93.1	33.2	7.95	712.64
1934	93.6	33.0	7.97	710.13
1935	97.3	34.7	8.05	754.20
1936	104.1	37.1	8.09	810.38
1937	112.9	40.1	8.13	880.24
1938	122.1	43.4	8.19	959.70
1939	128.2	45.4	8.20	1,005.16

Wages in Finland. This study uses average hourly wages of male workers in the paper and pulp industry, sawmill industry, and metal industry taken from Vattula (1983: 421–41). The average annual wages of male industrial worker were calculated in two ways. Since the sawmill industry was a major sector in Finland at the time (Korpelainen 1957; Ahvenainen 2011), the first average wage count is based on wages in this industry. The second wage count was adjusted for 25 percent of wages in paper and pulp industry, 25 percent of metal industry, and 50 percent of sawmills in 1938, due to the larger employment in the sawmill sector. The 1921–1928 period is based on the wage index from Mitchell (1975: 186). Working hours are the same as in Estonia. Average wages in 1925–1926 in the sawmill industry are based on the trend in the metal industry (coeff: Sawmill/Metal); 1927 and 1935 are

APPENDIX TABLE 11
ANNUAL AVERAGE WAGES OF MALE WORKERS IN LARGE
INDUSTRIAL ENTERPRISES IN FINLAND, 1921–1938

Year	Wage Index 1929 = 100	Wages of Male Industrial Workers (markkas per hour)				Average Annual Wage (markkas) (1)	Average Annual Wage (markkas) (2)
		Sawmill Industry	Metal Industry	Paper and Pulp	Sawmill/ Metal coeff.		
1921	68	5.03	4.98		1.011	11,410	11,141
1922	73	5.19	4.86		1.068	11,773	11,960
1923	81	5.57	5.13		1.086	12,633	13,270
1924	84	5.80	5.49		1.056	13,143	13,762
1925	88	5.95	5.64			13,496	14,417
1926	93	6.32	5.99			14,341	15,236
1927	96	6.67				15,131	15,728
1928	100	7.02	9.10	7.42	0.946	15,921	16,383
1929	100	7.15		7.51		16,216	16,383
1930	97	7.20	9.46	7.7	0.935	16,330	15,892
1931	85	6.03	8.23	6.69	0.901	13,676	13,926
1932	82	5.45	7.68	6.16	0.885	12,361	13,434
1933	93	5.54	7.81	6.38	0.868	12,228	15,236
1934	83	5.76		6.77	0.850	12,352	13,598
1935	84	5.89				12,680	13,762
1936	89	6.03	7.09	6.74		13,106	14,581
1937	97	6.83	8.15	7.57		14,919	15,892
1938	103	7.17	8.73	7.68		15,739	16,875

interim figures. For the comparison with Estonia, I used the first approach, but the second approach is still relevant. Appendix Table 11 shows the annual average wages of male workers in large industrial enterprises in Finland for the 1921–1938 period.

Postwar Period

Wages in Estonia. Average annual wages in the years between 1960 and 1988 (Appendix Table 12) were collected from different issues of the *National Economy of the Estonian SSR: Statistical Yearbook* (State Statistical Committee of the Estonian SSR 1968: 199; 1976: 40; 1980: 212; 1989: 229).

APPENDIX TABLE 12
ANNUAL AVERAGE WAGES IN ESTONIA, 1960–1988

Year	Average Annual Wage (roubles)
1960	975.6
1961	1015.2
1962	1,045.2
1963	1,078.8
1964	1,120.8
1965	1,183.2
1966	1,239.6
1967	1,321.2
1968	1,478.4
1969	1,551.6
1970	1,623.6
1971	1,683.6
1972	1,749.6
1973	1,790.4
1974	1,846.8
1975	1,917.6
1976	2,028.0
1977	2,088.0
1978	2,112.0
1979	2,160.0
1980	2,264.4
1981	2,313.6
1982	2,356.8
1983	2,407.2
1984	2,499.6
1985	2,581.2
1986	2,652.0
1987	2,748.0
1988	2,748.0

Wages in Finland. Appendix Table 13 shows annual average wages in Finland from 1955 through 1989. The average annual wage in 1988 was drawn from the *Yearbook of Nordic Statistics 1991* (Nordic Council of Ministers and the Nordic Statistical Secretariat 1991). The index of wages for 1955–1989 was taken from the *Statistical Yearbook of Finland 2005*.

APPENDIX TABLE 13
ANNUAL AVERAGE WAGES IN FINLAND, 1955–1989

Year	Index of Wages, 1938 = 100	Average Annual Wage (markkas)
1955	2,180	2,827.78
1956	2,470	3,203.95
1957	2,590	3,359.61
1958	2,720	3,528.24
1959	2,850	3,696.87
1960	3,010	3,904.41
1961	3,240	4,202.76
1962	3,430	4,449.21
1963	3,760	4,877.27
1964	4,260	5,525.85
1965	4,620	5,992.82
1966	4,960	6,433.85
1967	5,400	7,004.59
1968	6,000	7,782.88
1969	6,440	8,353.62
1970	6,980	9,054.08
1971	7,870	10,208.54
1972	8,790	11,401.92
1973	10,130	13,140.10
1974	12,130	15,734.39
1975	14,780	19,171.83
1976	16,990	22,038.52
1977	18,430	23,906.41
1978	19,690	25,540.82
1979	21,940	28,459.40
1980	24,580	31,883.87
1981	27,740	35,982.85
1982	30,650	39,757.55
1983	33,830	43,882.47
1984	37,020	48,020.37
1985	40,140	52,067.47
1986	42,939	55,698.18
1987	45,959	59,615.57
1988	50,083	64,965.00
1989	54,527	70,729.52

Post–Cold War Period

Wages in Estonia. Data on monthly average wages in Estonia in the 1992–2000 period is available on the Statistics Estonia agency website. The 2000–2018 period was drawn from the OECD data. Appendix Table 14 shows annual average wages in Estonia from 1992 through 2018.

APPENDIX TABLE 14
ANNUAL AVERAGE WAGES IN ESTONIA, 1992–2018

Year	Average Annual Wage (euros)
1992	420.00
1993	816.00
1994	1,332.00
1995	1,824.00
1996	2,292.00
1997	2,736.00
1998	3,168.00
1999	3,408.00
2000	4,107.72
2001	4,522.69
2002	4,979.85
2003	5,545.41
2004	6,200.45
2005	6,860.80
2006	7,880.47
2007	9,866.77
2008	10,818.04
2009	10,422.89
2010	10,774.01
2011	10,900.07
2012	11,786.88
2013	12,307.80
2014	13,063.60
2015	13,550.27
2016	14,428.19
2017	15,347.50
2018	16,795.39

APPENDIX TABLE 15
ANNUAL AVERAGE WAGES IN FINLAND, 1991–2018

Year	Average Annual Wage (euros)
1991	2,0471.00
1992	20,868.00
1993	20,853.00
1994	21,291.00
1995	22,227.00
1996	22,976.00
1997	23,450.00
1998	24,465.00
1999	25,399.00
2000	26,640.00
2001	27,490.00
2002	28,100.00
2003	28,924.66
2004	29,988.91
2005	30,993.62
2006	32,131.13
2007	33,241.35
2008	34,688.76
2009	35,609.68
2010	36,693.04
2011	38,005.38
2012	39,099.90
2013	39,628.54
2014	40,147.58
2015	40,694.04
2016	41,188.06
2017	41,285.46
2018	42,122.28

Wages in Finland. The data on annual average wages in Finland in the 1991–2018 period (Appendix Table 15) were drawn from OECD.Stat.

Appendix C: Purchasing Power of Wages, 1923–2018

The purchasing power of wages is the quantity of goods and services that can be bought with a wage unit. Its evolution is linked to that of prices and salaries. The ratio of PPW in Estonia to PPW in Finland, from 1923 through 2018, is shown in Appendix Table 16.

APPENDIX TABLE 16
PURCHASING POWER OF WAGES IN
ESTONIA AND FINLAND, 1923–2018

Year	Estonia, PPW	Finland, PPW	Ratio (%)
1923	2.57	2.85	90
1924	2.54	2.93	87
1925	2.49	2.86	87
1926	2.45	3.15	78
1927	2.73	3.30	83
1928	2.68	3.37	80
1929	2.68	3.51	76
1930	3.34	4.09	82
1931	3.66	3.83	96
1932	3.73	3.35	111
1933	3.63	3.33	109
1934	3.67	3.43	107
1935	3.84	3.39	113
1936	3.57	3.53	101
1937	3.60	3.72	97
1938	3.73	3.89	96
1939	3.83		
1940			
1941			
1942			
1943			
1944			
1945			
1946			
1947			
1948			
1949			
1950			
1951			
1952			
1953			
1954			
1955		4.56	
1956		4.52	
1957		4.19	
1958		4.17	

(Continued)

APPENDIX TABLE 16 (*Continued*)
PURCHASING POWER OF WAGES IN
ESTONIA AND FINLAND, 1923–2018

Year	Estonia, PPW	Finland, PPW	Ratio (%)
1959		4.33	
1960	2.38	4.38	54
1961	2.46	4.63	53
1962	2.50	4.70	53
1963	2.56	4.90	52
1964	2.63	4.96	53
1965	2.76	5.05	55
1966	2.89	5.25	55
1967	3.07	5.46	56
1968	3.44	5.45	63
1969	3.61	5.67	64
1970	3.77	6.06	62
1971	3.90	6.55	59
1972	4.04	6.69	60
1973	4.12	6.88	60
1974	4.23	7.11	60
1975	4.38	7.18	61
1976	4.63	7.09	65
1977	4.72	6.55	72
1978	4.73	6.70	71
1979	4.79	7.16	67
1980	4.97	7.11	70
1981	5.00	7.11	70
1982	4.90	6.99	70
1983	4.97	7.23	69
1984	5.18	7.35	71
1985	5.30	7.43	71
1986	5.39	7.65	70
1987	5.20	8.00	65
1988	4.99	8.57	58
1989		9.01	
1990			
1991		6.52	
1992	1.67	6.65	
1993	1.91	6.69	29
1994	2.36	6.81	35

(*Continued*)

APPENDIX TABLE 16 (*Continued*)
PURCHASING POWER OF WAGES IN
ESTONIA AND FINLAND, 1923–2018

Year	Estonia, PPW	Finland, PPW	Ratio (%)
1995	2.78	7.72	36
1996	2.89	8.10	36
1997	3.27	8.23	40
1998	3.57	8.43	42
1999	3.99	8.75	46
2000	4.70	9.08	52
2001	4.78	8.97	53
2002	5.11	8.92	57
2003	5.79	9.12	63
2004	6.21	9.39	66
2005	6.64	9.65	69
2006	7.26	9.86	74
2007	8.31	10.00	83
2008	7.98	9.61	83
2009	8.01	9.67	83
2010	8.03	10.31	78
2011	7.41	10.05	74
2012	7.72	9.83	78
2013	7.74	9.46	82
2014	8.21	9.57	86
2015	8.56	9.89	87
2016	9.12	10.12	90
2017	9.18	10.23	90
2018	9.74	10.25	95

References

- Aavik, M. (2015) “Mis on Põhjamaade edu saladus?” *Postimees Arvamus* (June 12).
- Ahvenainen, J. (2011) “The Competitive Position of the Finnish Sawmill Industry in the 1920s and 1930s.” *Scandinavian Economic History Review* 33 (3): 173–92.
- Bremmer, I. (2006) *The J Curve: A New Way to Understand Why Nations Rise and Fall*. New York: Simon & Schuster.

- Central Statistical Agency (1938) *Statistical Yearbook of Finland* 1938. New Series 36. Helsinki: Valtioneuvoston kirjapaino.
- _____ (1939) *Statistical Yearbook of Finland* 1939. Vol. 85. Helsinki: Valtioneuvoston kirjapaino.
- _____ (1990) *Statistical Yearbook of Finland* 1990. Vol. 90. Helsinki: Valtioneuvoston kirjapaino.
- Ellman, M. (2000) “The 1947 Soviet Famine and the Entitlement Approach to Famines.” *Cambridge Journal of Economics* 24 (5): 603–30.
- Estonian Institute of Economic Research (2018) “Average Retail Prices of Groceries in Estonian Stores.” Available at www.ki.ee/index.html.
- Hagfors, R., and Kuus, T. (1993) “Income: Structure and Distribution.” In Lugus and Vartia (eds.), *Estonia and Finland: A Retrospective Socioeconomic Comparison*, 302–30. Research Institute of the Finnish Economy (ETLA), Series B, No. 86. Helsinki: ETLA.
- Iives, T. H. (1999) “Estonia as a Nordic Country.” Minister of Foreign Affairs, to the Swedish Institute for International Affairs (December 14). Available at <https://vm.ee/et/uudised/valisminister-ilvese-loeng-rootsi-valispoliitika-instituudis-eesti-kui-pohjamaa-inglise>.
- Jørgensen, H. (2006) “The Interwar Land Reforms in Estonia, Finland, and Bulgaria: A Comparative Study.” *Scandinavian Economic History Review* 54 (1): 64–97.
- Korpelainen, L. (1957) “Trends and Cyclical Movements in Industrial Employment in Finland, 1885–1952.” *Scandinavian Economic History Review* 5 (1): 26–48.
- Laar, M. (2008) “Leading a Successful Transition: The Estonian Miracle.” *European View* 7 (1): 67–74.
- Lapping, M. B. (1993) “The Land Reform in Independent Estonia: Memory as Precedent: Toward the Reconstruction of Agriculture in Eastern Europe.” *Agric Hum Values* 10: 52–59.
- Lugus, O., and Vartia, P., eds. (1993) *Estonia and Finland: A Retrospective Socioeconomic Comparison*. Research Institute of the Finnish Economy (ETLA), Series B, No. 86. Helsinki: ETLA.
- Mitchell B. R. (1975) *European Historical Statistics, 1750–1970*. London: Macmillan.
- MONSTAT (Montenegro Statistical Office) (2018) “Methodological Guidelines: Minimal Consumer Basket.” Podgorica, Montenegro.

- National Bureau of Statistics (1937) *Monthly Estonian Statistics* 182 (1) and 186 (5). Tallinn: Olion.
- _____ (1938) *Monthly Estonian Statistics* 193 (12) and 198 (5). Tallinn: Olion.
- _____ (1939) *Monthly Estonian Statistics* 217 (12). Tallinn: Olion.
- _____ (1940) *Monthly Estonian Statistics* 223 (6). Tallinn: Olion.
- Nordic Council of Ministers and the Nordic Statistical Secretariat (1991) *Yearbook of Nordic Statistics 1991*. Vol. 29. Stockholm: Norstedts Tryckeri.
- Nordic Estonia. "Estonia as a Nordic Country That Got Occupied by USSR." Available at www.nordicestonia.com/nordic/introduction.
- OECD.Data. "Prices: Inflation (CPI)." Available at <https://data.oecd.org/price/inflation-cpi.htm>.
- OECD.Stat. "Databases: Labour: Earnings: Average Annual Wages: Estonia, Finland." Available at https://stats.oecd.org/Index.aspx?DataSetCode=AV_AN_WAGE#.
- Raun, T. (1987) "Finland and Estonia: Cultural and Political Relations, 1917–1940." *Journal of Baltic Studies* 18 (1): 5–20.
- Russian Federation Labour Code (2012) "On the Consumer Basket in General in the Russian Federation Act, Article 2. N227-FL." Available at <https://legalacts.ru/doc/federalnyi-zakon-ot-03122012-n-227-fz-0>.
- State Statistical Committee of the Estonian SSR (1968) *National Economy of the Estonian SSR in 1967. Statistical Yearbook 1967*. Tallinn: Olion.
- _____ (1976) *National Economy of the Estonian SSR in 1975. Statistical Yearbook 1976*. Tallinn: Olion.
- _____ (1977) *National Economy of the Estonian SSR in 1976. Statistical Yearbook 1977*. Tallinn: Olion.
- _____ (1980) *National Economy of the Estonian SSR in 1979. Statistical Yearbook 1980*. Tallinn: Olion.
- _____ (1988) *National Economy of the Estonian SSR in 1987. Statistical Yearbook 1988*. Tallinn: Olion.
- _____ (1989) *National Economy of the Estonian SSR in 1988. Statistical Yearbook 1989*. Tallinn: Olion.
- Vattula K. (1983) *Economic History of Finland, Vol. 3: Historical Statistics*. Helsinki: Tammi.

BARBER LICENSING IN ARKANSAS: PUBLIC HEALTH OR PRIVATE GAIN?

Tanner Corley and Marcus M. Witcher

In Arkansas, the barber profession has been regulated and licensed for more than 80 years, and until recently, the issue was mostly absent from the political debate. During a regular session of Arkansas's 92nd General Assembly in 2019, however, state Sen. John Cooper presented a bill to "repeal the [1937] Arkansas Barber Law" and to "abolish the State Board of Barber Examiners" (Briggs 2019). The average Arkansan probably was not aware of the bill, but occupational licensing reformers saw this as a great opportunity for Arkansas to pave the way for other states to reform their own license laws. If Cooper's bill had passed, Arkansas's economy would have likely benefited (Timmons and Thornton 2010, 2018). By removing restrictive requirements to becoming a barber, the bill would have allowed more Arkansans to enter the profession. This reform would have

Cato Journal, Vol. 41, No. 1 (Winter 2021). Copyright © Cato Institute. All rights reserved. DOI:10.36009/CJ.41.1.6.

Tanner Corley is an Undergraduate Fellow with the Arkansas Center for Research in Economics (ACRE) at the University of Central Arkansas. Marcus M. Witcher is an ACRE Scholar-in-Residence and affiliated with the Department of History at the University of Central Arkansas. They thank Mitch Mitchell, Thomas Snyder, and the entire ACRE team for their support of this project. They also thank Alicia Plemmons for reading an early draft of the paper and all the participants at the Knee Center Occupational Licensing Conference for their feedback. Finally, they would like to thank Harry David for his careful reading and suggestions.

provided people with more economic opportunities, increased competition, and benefited consumers.¹

Although the bill ultimately failed, it forced the Board of Barber Examiners to justify its existence. The board denounced the bill as a mistake. In a hearing hosted by the Board of Barber Examiners, the director used the history of the board's creation to justify opposing the bill. The director proclaimed that "the State Board of Barber Examiners was founded 82 years ago to prevent unlicensed barbers from coming to Arkansas and working in unclean and unsanitary conditions. The minutes of that first meeting indicate it was out of control" (Arkansas Board of Barber Examiners 2019).

The board presented a story in which a public health crisis made licensing regulations necessary to ensure that Arkansans got a safe shave. The minutes from the first meetings of the board, however, do not support that claim. Indeed, the minutes consist of mundane items and certainly nothing that indicated the profession was "out of control." Modern-day justifications for the existence of the board rest on its protection of public health. But the origin of the Board of Barber Examiners is much more complicated and less noble than the director wanted Arkansans to believe.

Much work has been done in recent years on occupational licensing and its effects on the economy (Timmons and Konieczny 2018; Thornton and Timmons 2015; Zapletal 2019). The commonly held view in the literature is that occupational licensing limits the number of people who can participate in an occupation, thus restricting supply and raising the cost for consumers. Moreover, the regulatory apparatus that is created by licensing laws often leads to rent seeking and sometimes to capture by the industry (Kleiner 2000). Scholars also argue that restrictive occupational licensing disproportionately

¹ Although scholars disagree about the effects of delicensing barbers, we believe the economic impact would be positive. In support of our position, Hall and Pokharel (2016) demonstrate that the number of exams required for barber licensure negatively correlates with the number of barbershops in a state. Deyo (2017) finds that states with licensing exams have less competition, diminishing returns from licensure, and overall lower quality of service. Likewise, Carrol and Gaston (1983) demonstrate a strong negative association between occupational licensing and quality of service. Plemmons (2019) demonstrates that states with strict occupational licensing requirements have less employment and fewer firms. Although they examine cosmetology and not barbering, Adams, Jackson, and Ekelund (2002) find that increases in licensing regulations result in a smaller workforce, wage gains for those who are licensed, and deadweight losses.

harms minorities who are denied access to labor markets (Bernstein 1994; Klein, Powell, and Vorotnikov 2012). As more and more occupations have become licensed in America, the negative effects of licensure have become more evident and important (Kleiner 2000). While most of the research on occupational licensing focuses on its effects on the modern economy, very little work has emphasized how these regulations came about in the first place. Indeed, this may be the first article to examine the origins of barber licensure in any state. By looking at a specific industry in a specific state, we uncover the tactics of barbers and unions in seeking licensing laws.

Arkansas adopted regulations licensing the barber industry in 1937 and was one of the last states to adopt such regulations. Accordingly, the tactics used by the Journeyman Barber International Union of America (JBIUA) to lobby for licensing were well developed by the time the union came to Arkansas in the hope of achieving legislative reforms. After a couple years of organizing, and in the midst of the Great Depression, the Board of Barber Examiners was created and given the task of enforcing licensure laws on all of the barbers in the state. Not long after the formation of the board, letters came pouring in from local barbers and others who had campaigned for the regulations and were seeking the enforcement positions created by the new law. In this sense, the story of the origin of barber licensure in Arkansas resembles many other stories in which an entrenched interest lobbies the government to create a regulatory framework to exclude potential competitors. Once that framework has been created, it is often the established professionals that capture the regulatory body and use it to perpetuate their market share (Downs 1957; Olson 1965; Stigler 1971; Hilton 1972; Posner 1974; Peltzman 1976; McChesney 1987; Leaver 2009).²

Arkansas did not lack regulations governing barbers prior to the creation of the board, although it did lack licensing. As early as 1913, Arkansas introduced sanitation codes that covered various trades, including barbers. Act No. 96, introduced in 1913, gave the State Board of Health the power to introduce sanitation regulations that

²There are also a good number of historical examples. Gabriel Kolko (1963) argues that many of the regulatory reforms in the Progressive Era were sought by the very corporations that were to be regulated and that the corporations were able to stamp out free competition through those reforms. See also High and Coppin (1988), Mahony (2001), and Bluestone (1991).

mirrored the regulations advocated in 1937. The purpose of the 1913 act was to prevent the spread of disease in unsanitary barber-shops (Arkansas State Board 1913). Despite the existence of these sanitary regulations, barbers lobbied for a barber board that would impose licensing requirements. It seems clear that sanitation was not the sole—and perhaps not the primary—motivating factor for such lobbying.

During the 1930s, Arkansas barbers cited public health as their main justification for the creation of a barber board. In reality, they also had a powerful private interest in creating a regulatory apparatus that would give barbers more control of the profession. To push for licensure laws, barbers created a local Journeyman Barber Union that fell under the umbrella of the JBIUA (*Hope Star* 1936: 1; *Northwest Arkansas Times* 1938: 1). While motivations for wanting to become licensed may have varied from barber to barber, the union presented some common reasons. If we look at the history of the international union, it becomes clear that the barbers used long-standing tactics to capture the regulatory apparatus through creating a board run by them.

The story of barber licensure in Arkansas is a story full of rent seeking. In establishing licensing laws, barbers were motivated by the prospects of limiting competition and increasing their wages. Perhaps unsurprisingly, after the Board of Barber Examiners was created, many barbers sought jobs on the board and favors from it. Rent seeking occurs when a private interest seeks favorable legislation from the government to increase its wages and exclude competition (Tullock 1967; Krueger 1974; Tullock 1993; Tullock, Seldon, and Brady 2002). Although most of the General Assembly in Arkansas supported the law creating the board, enforcing the new rules was not smooth sailing. In fact, some barbers opposed the law, and organizing such a large industry was difficult. Most importantly, perhaps, the creation of the board placed new power in the hands of the few men in charge. Employment on the Board of Barber Examiners was widely sought, especially given that the board was created during the Great Depression. The opportunity for nepotism and corruption arose, and some barbers attempted to use their political leverage to become part of the new regulatory apparatus. Perhaps unsurprisingly to public choice economists, once the board was created and the license law went into effect, many barbers sought even more regulation and further restrictions

to entering the profession. As happens with most regulatory frameworks, once in place, the regulatory framework governing barbers expanded and became more burdensome.

Finally, underlying the motivations for the Arkansas Barber License Law was a Progressive Era (1890–1920) desire to professionalize the industry.³ Most barber license laws in the United States were enacted during the Progressive Era. Though Arkansas lagged behind most states, the law followed a similar pattern to the previous ones. Like other professions, such as doctors and lawyers, barbers thought of themselves as highly professional and skilled laborers. Barbers wanted the same respect as other highly professional laborers and adhered to the idea that not just anyone could be a barber. Barbers in Arkansas wanted only skilled professionals to work in barbershops, and by establishing examinations and other barriers to entry, they hoped to restrict barbering to respectable professionals. As is often the case, the new barriers to entry empowered men with racist attitudes, who used their newfound power to limit racial minorities' entry into the profession.⁴

History of Barber Licensing

The fight to professionalize barbers and pass legislation regulating the industry began long before the 1937 law (Act 313) passed the General Assembly. In fact, in a letter to the governor of Arkansas, one barber, Carl Bailey, boasted that “we have been trying for fifteen years to get this measure passed but have been unable to do so until Arkansas was blessed with having you as our Governor” (Owen 1937). It is likely that barbers in Arkansas wanted licensing requirements earlier, but the issue gained traction during the 1930s.

Although the adoption of barber licensing laws took place later in Arkansas, it followed the pattern of professionalization and reform adopted in most other states during the Progressive Era. In Lawrence Friedman's “Freedom of Contract and Occupational Licensing 1890–1910: A Legal and Social Study,” he analyzes the

³No single volume has been published on the rise of professionalism during the Progressive Era, but it is discussed in numerous books, including Wiebe (1967), Hofstadter (1955), Burrow (1977), Bledstein (1978), Geison (1983), Chandler (1977), Furner (1975), Haskell (2000), and Leonard (2016).

⁴Although we find no explicit evidence of this in newspapers from the time, Quincy T. Mills (2013) details the struggles of black barbers.

period in which occupational licensing laws came into effect and how those laws survived challenges in the courts. According to Friedman, it was from 1890 to 1910 that “occupational licensing first achieved a firm foothold in the statute books of most American states” (Friedman 1965: 489). Other occupations for skilled laborers were licensed well before this period, including plumbers, doctors, barbers, and funeral directors. In general, occupational licensing laws during the Progressive Era were met with indifference. Courts and the public tended to regard occupational licensing law, unlike other areas of notable labor law, as “a quieter, blander area of constitutional law” (Friedman 1965: 489). Not all occupations found it easy to achieve their own license laws. For instance, horseshoers sought licensing, but it was to no avail because “the distinction between a doctor and the horseshoer was that the doctors were professional men” while the horseshoers were an “ordinary trade” (Friedman 1965: 493). Various occupations sought licensing during this period, but some sense of professionalism and a need to protect the public was needed for the courts to acquiesce.

The occupations that wanted license laws tended to already be organized and sometimes faced no opposition in pursuing their goal. In fact, “the success of the licensing movement depended in part on the absence of a strong and coherent pressure group of employers united in opposition” (Friedman 1965: 487). Most barbers were self-employed and for that reason faced little opposition. The barbers organized in the form of unions and sought to regulate the trade through unionization. Many unions found, however, that organizing had its limits and that the enforcement mechanisms of the state government were needed for them to fully achieve their goals. As in many other occupations, “licensing was an attempt to enforce, through legal mechanisms, goals which the trade or professional association was unable to carry through completely on its own” (Friedman 1965: 503).

In the early 1930s, Arkansan barbers suffered from the economic ramifications of other states passing barber license laws. Barbers in Hope, Arkansas, complained that “there are twice as many barber-shops and nearly twice as many individual barbers in Hope today as there were when business was at its peak” (Washburn 1934). By requiring examinations and adding other barriers to entry, other states had effectively cut out labor and forced many barbers to close

up shop and move where they could legally work. In the early 1930s, that was Arkansas. In supporting Act 313 of 1937, Rep. Rupert Condrey of Sebastian, Arkansas, argued that because Arkansas had no barber licensure law, it had become “the dumping ground for all barbers in the middle west who [moved] because they are suffering from disease that would prevent their practicing the trade in other states” (*Camden Times* 1937: 1). Although we have not been able to substantiate the claim that disease led to barber migration, it is likely that some barbers moved to the state in search of a less regulated market. The increase of barbers in Arkansas expanded the number of providers and likely lowered their profits. Arkansas barbers responded by attempting to capture the market as their neighbor states had, driving down the supply of barbers, and increasing their own wages and profits in the process.⁵

When John B. Robinson, vice president of the JBIUA, met with a local union in Arkansas, he informed the barbers that “Arkansas, South Carolina, and Virginia are the only Southern States that have no law requiring barbers to be licensed” (*Hope Star* 1937: 1). Robinson may have used this fact to urge Arkansans to adopt a barber licensure law, although the historical record is not clear. Regardless, Arkansas barbers followed suit shortly after his visit. Historian Scott Hall documented the effect that licensing laws had on the supply of barbers in his report on the international union. Hall (1936: 87) found that “the license laws have also reduced the number of barbers” and that license requirements can “keep a number who would normally complete their training from doing so and eliminate a few already in the trade.” Occupational licensing laws, by creating barriers in the form of lengthy training and expensive fees, disincentivized potential barbers from entering into the profession. Those Arkansans who could afford to complete the training and pay the fees necessary to become barbers were only affected in that it took them longer to become barbers. Those who could not afford to meet the regulatory standards, however, were less likely to enter the profession and were forced to look for employment elsewhere.

⁵Arkansas barbers had good reason to believe that licensing their profession would increase wages. Kleiner (2017), Kleiner and Vorotnikov (2017), Pfeffer (2014), and Timmons and Thornton (2010) confirm that licensing had that effect.

The Sanitation Myth

While many variables were at play in the establishing of the new regulating board, the most notable justification for the new law was the need for better sanitation in barbershops. Barbers claimed that legislation was needed to improve the sanitation practices of their field to ensure public safety. In 1932, just five years before Act 313 came into effect, the *Hope Star*, a local newspaper, noted that “there are three different diseases of the skin which may affect the region of the beard in men, and which are usually picked up in dirty barber shops” (Fishbein 1932: 4). Another article, written in 1937, the same year the law came into effect, suggested that unsanitary barbershops could cause “barber’s itch.” What is most striking about the article is that it ended by claiming that “more and more barbers are dispensing with the unsanitary common shaving brush” (Fishbein 1937: 2). Although sanitation was still a concern in the mid-1930s, barbers were adjusting prior to licensing to make their shops cleaner and more sanitary, likely in response to consumer expectations.

In fact, there is evidence of these efforts well before the 1930s. As far back as 1911, the *Daily Arkansas Gazette* mentioned that “ten years ago it was only the exceptional barber shop that was sanitary. Today it is only the exceptional one that is unsanitary” (*Daily Arkansas Gazette* 1911: 4). Ten years later, another newspaper commented on the improving conditions of barbershops and how they reduced the cases of barber’s itch. In 1921, an article by the *Democrat Gazette* claimed that “barber’s itch is almost as rare nowadays as the proverbial hen’s teeth” and that the reason was that “cleanliness has become a habit” (Copeland 1921: 10). In 1937, those barbers who wanted the establishment of a board pushed the idea that licensure was needed to combat unsanitary barbershops. But the newspapers at the time gave only a minuscule amount of attention to the issue and we have been unable to uncover any cases in which any Arkansan was hurt by an unsanitary barber.

It seems likely that barbers only used the sanitation argument as a means of convincing the public and state government to act. The real goal of barbers was to capture the regulatory apparatus. Barbers wanted licensure, not for the sanitation regulations, but to prohibit laborers from joining the industry, thus keeping the number of barbers down and their wages high.

As is often the case, however, perception was more important than reality. Accordingly, barbers in Arkansas and across the country focused on fostering the perception of a sanitation problem as a means of getting licensing laws on their state books. Indeed, appealing to public health was a longstanding tactic of the JBIUA. In the *Journeyman Barber Journal*, the president of the union, James Shanessy, urged the union to “crystallize public sentiment in favor of the Barber’s License Law” by allowing the delegates to bring the matter before the trades councils and “show them it is a health measure” (Shanessy 1922: 152). To convince state legislatures to acquiesce, barber unions felt that the public had to be convinced that regulation was needed to protect public health.

Although barbers used public safety as the public justification for the law, they had other, private reasons for wanting licensure. One barber, in a letter to Governor Bailey urging him to sign the Barber Bill, praised the governor and the bill his General Assembly was able to draft. While praising the bill, the barber explained how proud the state was that the General Assembly saw “the need of such a protection for the general public as well as the barbers of our proud state” (Owen 1937). The bill was supposed to protect the state, but the barbers were also seeking protections for themselves. Specifically, they wanted to protect their employment and, even more importantly, their wages. Because asking the state legislature for protection of their own industry would look blatantly self-interested, the union and participating barbers focused on the view that unsanitary barber-shops were a threat to the public.

When it came to support for the bill in Arkansas, a majority of the General Assembly was in favor. The bill passed by a vote of 52 to 37 (*Camden Times* 1937: 1). Of the few who opposed the bill, Rep. Sam Cunningham offered a unique argument as to why he was against it. Cunningham contended that the bill’s passage would “prevent farmers from cutting their neighbors hair on Sunday afternoon, a popular past-time in rural districts” (*Camden Times* 1937: 1). While the argument might sound strange, Cunningham touched on a reality about government regulation: requiring a license to do a certain job creates criminals. It might sound irrational to make it illegal for rural farmers to cut their neighbors’ hair, but nonetheless the bill passed.

Although the barbers in Arkansas were overwhelmingly in favor of a license law, not all were on board. One barber, H. C. Beaty, was so opposed to the law that he filed a lawsuit in his local court (Pulaski’s

chancery court) that eventually made its way to Arkansas's supreme court. Beaty found fault with many of the specifics outlined by Act 313. Beaty contended that the law would result in a "confiscation of private property without due process of law; and a duplication of State Agencies, having power to prescribe sanitary regulations for the operation of Barber Shops." Beaty believed that requiring a license to become a barber was unconstitutional in that it could result in confiscation of private property and that it "is a burden on the right of a free man, to the pursuit [of] a vocation." More practically, Beaty pointed out that the barber profession in Arkansas was already regulated for health and safety (*Beaty v. Humphrey et al.* 1937).⁶

In 1933 the State Board of Health had reemphasized its regulations of the barber profession, in which sanitation had always been central. The provisions provided by the State Board of Health stated that "combs and brushes shall be thoroughly sterilized after each separate use" and that "every barber shop shall be kept well ventilated and provided with hot and cold water." In fact, one of the only differences between the regulations on the books and those put forward by Act 313 in 1937 was the new requirement to obtain a license in order to be a barber. While Beaty focused on the 1933 codes, in reality, the codes were simply a reaffirmation of the same codes introduced in 1913 by the State Board of Health (Act No. 96 1913: 42). During the trial, Beaty's counsel questioned the president of the Board of Barber Examiners about the additional costs barbers would have to bear in meeting the new standards set by Act 313. Unfortunately, Beaty's counsel did not question the need for the act despite there already being sanitation regulations for the barber industry. Accordingly, the board did not have to explain the need for a duplication of regulations, the answer to which seemingly should have been critical to determining the necessity of the new license law (*Beaty v. Humphrey et al.* 1937).

Beaty's case was dismissed by Pulaski's chancery court but was then granted an appeal to the Supreme Court of Arkansas. When the case came before the Supreme Court, the Arkansas justices came to their conclusion mostly through precedent. The court argued that Arkansas's Barber License Law was "almost an exact copy of the

⁶For an analysis of the legal justifications for occupational licensing and the criticisms, consult Larkin (2016).

Texas Barber Law.” Furthermore, the court argued that 46 states had adopted a barber law and that “without exception wherever the constitutionality of such laws have been brought in question, they have been sustained” (*Beaty v. Humphrey* 1938).

Beaty ultimately lost his case, but he emphasized that if the concern was sanitation, then Act 313 was not necessary as sanitation regulations were already in place. Despite Act 313’s claim that “the barbering profession in this state is utterly without regulation,” sanitation regulations had been on the books in Arkansas since 1913 (Act 313, Section 25). In truth, the new provisions in Act 313 did little to extend those measures. If there was a public health crisis in 1937 despite the existence of regulations, why did Act 313 not place greater emphasis on sanitation? Instead of presenting new measures to ensure safe practices, the 1937 legislation mandated that only licensed barbers could practice. Perhaps the reason barbers thought Act 313 would be more effective is that they believed that only those within the profession (and especially union barbers) had the intimate knowledge necessary to regulate it. Or perhaps they simply wanted to restrict the number of people entering their profession. While those who were currently barbers were grandfathered in, Act 313 required new entrants to the profession to attend a certified barber school for 1,000 hours, among other things, to acquire a license (Act 313 1937). It is hard to understand how such a grandfather clause makes sense if the real reason for the regulation was public health. Beyond putting the barbers in charge of regulating sanitation in their own shops, the 1937 act gave barbers control over who could become a barber, thus granting the union barbers the victory they had long desired.

Board of Barber Examiners

The Barber License Law established a new government board to oversee and regulate the industry. The board was tasked with organizing barbers and enforcing licensure. Beyond simply growing the government in size and scope, the law created the opportunity for corruption and nepotism. When it came time to appoint members of the new board, countless barbers from all across Arkansas sent in their applications. The legislation outlined that the board was supposed to consist only of Arkansas barbers who had worked at least five years in the state. This wisdom of appointing only members from

the industry being regulated later came under criticism and reforms were made.⁷ After the establishment of the board, one barber with the hope of being appointed inspector was told that “there is [sic] others in your section of the state using quite a bit of political influence to get a place as inspector” (Callahan 1937). The result of this was countless appeals for political favors in the form of a job on the Board of Barber Examiners.

Other barbers pointed to their previous political support as reason for their appointment to one of the newly created positions. In a letter to Governor Bailey, Harvey Booth offered loyalty to the cause of licensing as an explanation of why his friend Claud Marsh should be an inspector for the board. Booth contended that “Claud Marsh did a good piece of work for us last summer in White County and deserves a lot of consideration for it. I know that he was more than willing to do anything I suggested to him” (Booth 1937). It is unclear what became of Booth’s recommendation, but the door was open for potential abuse and nepotism. Indeed, members were appointed by the governor himself, and it appears evident that to get a seat, one had to have some strong connections (Act 313, Section 15). Rent-seeking barbers attempted to use their political clout to get a seat on the board and left out any claims to expertise on public health in the barber profession. If the board was created with the sole purpose of protecting public health, the appointing process should have sought out those who understood public health in barbering rather than those with political power.

The level of nepotism in Arkansas is hard to pinpoint, but evidence from other states implies that it played some role. Scott Hall (1936: 85) found that “it has been said the board members are incompetent political henchmen, that they have failed to draw up adequate sanitary and other rules.” He concluded, however, that “it appears that on the whole the board have enforced the acts, although frequently not to the degree the union desires” (Hall 1936: 85). Arkansas’s Board of Barber Examiners and other states’ barber boards erected barriers to the barber profession but usually not to the extent that union barbers wanted.

⁷There is growing recognition that public officials not aligned with the industry being regulated should have a seat on these licensing boards (Lauterbach 2020). Arkansas is somewhat peculiar because it requires that a senior citizen (someone over the age of 65) sit on the board.

Act 313 made barber licensing the law, but the legislation was not executed seamlessly. On August 14, 1939, just two weeks before the deadline, the secretary of the board warned the barbers that they must renew their license by September 1 and that “only a few had been paid to date” (*Courier News* 1939: 1). Enforcing such a new law could not have been easy, and it appears barbers were often confused about the contents of the law itself. One barber, while recommending his friend for the position of board member, admitted to the governor that he had “not had the opportunity of reading the act and am not familiar with its contents” (Walls 1937). In this way, the new regulation would have made the practice of barbering and the cost of a shave even more expensive. Rather than focusing on improving their services, barbers had to dedicate time and effort to understanding Act 313 and how to comply with the new licensing regulations.

Even worse, the rules governing the profession frequently changed. Barbers not only had to familiarize themselves with Act 313 but were also forced to keep up with the changes in the law. It is likely that Arkansas’s General Assembly did not consider how much further the Board of Barber Examiners would go beyond the initial scope of the law. The minimum requirements established with the original law of 1937 might not have caused much contention, but boards tended to grow in requirements and regulations. In Hall’s report on the JBIUA, he found that “an effort is also made to secure amendments to strengthen the law” (Hall 1936: 88). In Arkansas, this again proved true. In an article written in 1944, just seven years after the passing of the original bill, barbers argued that the fees were too low. The board adopted a resolution to double the fees so they could put two full-time inspectors on the road (*Courier News* 1944: 5).

To this day, the board continues to grow in scope and requirements. When Act 313 passed in the General Assembly, it stipulated that those seeking a license must log 1,000 hours at a certified barber school and also be of “good moral character and temperate habits.” Of course, what is considered “good moral character” was subjective, which certainly created the opportunity for abuse of power. Over 80 years later, the board and licensure law are practically the same but the required hours from a barber school has increased to 1,500 hours.

After 1937, the members of the board controlled the industry, but barbers were in some cases willing to work the system. In a report to the *Journeyman Barber Journal*, one journeyman complained that

“some of these birds have secured their license by false affidavit” and that barbers would pay a “stiff” to take their examination for them (Ackerman 1922). Apparently, some entrepreneurial barbers spent time and money to find ways around the existing regulatory apparatus.

It is also important to note that the barbers who were already well established were exempted from examinations by the 1937 legislation. Act 313 stipulated that any barber having worked in the barbering profession in Arkansas for the past six months would be grandfathered in and given a license without examination (Act 313, Section 8). In this way, Act 313 immediately put those already barbering in Arkansas at an advantage over those looking to enter the profession.

Union Influence

The Journeyman Barber Union deserves much of the credit—or blame—for getting the 1937 Barber License Law enacted in Arkansas. The international union and the local union in Arkansas were the main advocates for the law and had a vested interest in seeing it pass. While unions were always trying to get benefits for their workers, the journeymen saw a unique opportunity in occupational licensure. The JBIUA had experienced minute victories and losses, but licensure was seen as an end to all troubles. Unions quickly discovered that a license law could effectively put into practice what they had wanted all along: higher wages, less competition, and control of the profession. The union realized that if it could harness the power of the state, it could obtain new regulations and potentially get control over the regulatory process.

In New York, one editor for the *Journeyman Barber Journal* gloated at the newly enacted barber license laws: Leon Worthall, a barber in one of the first states to create a barber law, loved the idea that if a shop broke any regulations, the Health Department could “revoke the permit and close the place” (Worthall 1922: 55–56). In the past, unions had tried to use the power of organization to effect change, but strikes did not grant them the power that government had to enforce regulations. The key was ensuring that the barbers themselves were the ones regulating their industry. Barber H. E. Brush demonstrated that he understood this well when he argued that barbers should “wield an influence (if they stood solidly together) equal to a political party. Our State Association should be in power in the state instead of being unknown” (Brush 1922: 167).

Another barber, James Smith, in a report to the *Journeyman Barber Journal*, exclaimed it was “now to make the Barber Union 100 per cent” (Smith 1922: 163). Both barbers understood that their union might have had some incremental success, but to achieve all of the union’s goals, the state was needed.

Although barbers were asking for help from legislatures, they did not trust just anyone to do the regulating. Barbers in Arkansas and across the country believed it was only distinguished barbers who could oversee the profession. One barber, M. H. Whitaker, warned the international union against placing enforcement of the law under the purview of the State Board of Health. Whitaker argued that physicians “only see the need of sanitation to protect the public, and do not take as deep an interest in the protection of the barbers as barbers would” (Whitaker 1922: 310–11). Accordingly, the union worked to convince state legislatures that only barbers had the intimate knowledge to regulate their industry. Like other states, Arkansas decided to give the Board of Barber Examiners the power to enforce the Barber License Law and also appointed barbers to each position of the board. Other states went as far as to require a member of their Journeyman Barber Union to sit on the board.⁸

The local Journeyman Barber Union and JBIUA understood how license laws could benefit them and prioritized such laws, but barbers were not the first profession to realize this. Another writer to the *Journeyman Barber Journal* commented on the new license laws being enacted and pleaded that “other crafts are realizing the importance of legislation of this kind for their protection” (Miller 1922). Members of the international union knew what effect license laws could have on a given occupation, and their number one priority became securing legislation for their own profession. Importantly, the Journeyman Barber Unions’ work did not end after license laws were secured. The main goal for the unions was to enact a license law in each state, but once the law was enacted, efforts to secure amendments and strengthen the law were made (Hall 1936: 88). This held true in Arkansas as the Board of Barber Examiners fought to increase fees and strengthen the 1937 law.

⁸ Kentucky still has this requirement (Timmons and Thornton 2010).

Progressivism and Professionalism

Underlying all of the forces that brought about the barber law and Board of Barber Examiners was a Progressive Era attempt to professionalize the trade of barbers. The Progressive Era was marked by the government's attempt to combat the problems that came with industrialization and urbanization, mostly through public health and workplace regulations. The professionalization of the trade of barbering began during the Progressive Era and continued into the 1930s. In another article in the *Journeyman Barber Journal*, a barber argued that barbers should benefit like some of the other "properly regulated callings and professions—for instance, the lawyers, the doctors, the dentist, and many others" (Shanessy 1923: 454). All of the occupations considered very professional enjoyed higher wages and profited from license laws, and the barbers wanted some of the same benefits. The JBIUA made license laws its number one priority but "saw regulation as a central component to their larger vision of professionalizing the trade" (Mills 2013: 127). Beyond just increasing wages, the attempt to professionalize the trade of barbering was also an attempt to control the public sphere and grant barbers the respect they thought they deserved.

By regulating the barbers in Arkansas, the Board of Barber Examiners could exclude laborers it deemed unprofessional, thus making the entire profession respectable. One barber, in a report to the *Journeyman Barber Journal*, argued for license laws so they could "get rid of some scabs" (Ambrey 1922: 20). The same barber reported that a man was using his old barn in Virginia as a barber-shop, where he cut hair for low prices and "work[ed] all day and night and Sundays." What might be considered today as a noble effort by a poor entrepreneurial worker enraged the complaining barber, who finished by insisting that he "would like to see the key turned on them" (Ambrey 1922: 19–20).

Unprofessional and unsanitary barbershops were also seen as a blemish on the noble profession. Members of the JBIUA wanted to "unite in an effort to place the occupation of a barber upon the high plane it justly belongs" (Miller 1922: 100). Barbers in Arkansas, whether they liked it or not, were part of this movement by the international union to define what was a professional barber and keep those unworthy of such a status from sullyng their image. In this way, the barbers had not only the goal of enhancing their wages, but also the goal of elevating their status in society.

In conjunction with the barbers' efforts to professionalize the industry was the nefarious motivation to exclude racial minorities. Scholars have detailed the racial element of professionalization in the Progressive Era, and the national barber profession has been accused of attempting to keep blacks out of barbering (Bernstein 1994; Klein, Powell, and Vorotnikov 2012). In *Cutting Along the Color-Line: Black Barbers and Barber Shops in America*, Quincy Mills documents the history of black barbershops and analyzes how the industry changed during the Progressive Era. Mills lays out the position that "the entire industry during the Progressive Era assumed a new shape." While the new unions and technological innovations had changed the industry, it was the "emphasis on state regulation and professionalization" that defined who would be a barber. Mills (2013: 109) argues that it was through these efforts that "it became clear that blacks were positioned outside the boundaries" of what white barbers envisioned.

As to what white barbers had in mind, Mills explains that white barbers "envisioned a modern shop with sanitary standards" where "skilled craftsmen performed their artistry on white men in urban America." While various things could threaten a barber's employment and wages, "color-line barbers were targeted as direct competition." As far as how unions went about getting legislation that would exclude black barbers, Mills, like this paper, argued that public health was the main focus. Progressive Era measures to safeguard public health were, in Mills's opinion, "means of establishing order in the public sphere," where "the public" meant "the white public." For Mills, references to public health were the union's way of recruiting Progressive Era reformers, and black barbers believed the entire situation to be "a façade for driving them out of the field" (Mills 2013: 126–27).⁹

While there is little evidence that Arkansas barbers had racial motivations, it seems likely given Arkansas's racial history that some black Arkansans were targeted by the new law. Given that the law was enacted during the height of Jim Crow, the law created opportunities

⁹ Mills's analysis runs counter to the findings of Blair and Chung (2019), who conclude that when a profession is licensed, characteristics such as race and gender have less influence on wages and result in smaller wage gaps than in unlicensed areas. It is possible that this discrepancy is a product of the historical context that Mills is writing about (it was the height of racial discrimination: the Jim Crow Era). For more on black barbers, consult Bristol (2015).

for racial discrimination even if that was not the intention of the drafters of the law. One indication that blacks may have been discriminated against can be found in the board's minutes. The mark "(C)" next to certain barbers' names who had taken an examination indicated that the barbers were "colored."¹⁰ Of course it is also possible that this denotation was simply the product of a segregated society in which public officials had an interest in separating whites and blacks in all things, including professions.

Conclusion

While scholars who research occupational licensing are currently focused on the (often negative) effects that licensing laws have on workers, firms, and consumers, they should also be interested in the origins of these measures. In Arkansas, the Barber License Law was created to respond to what the barbers' international union and the local Journeyman Barber Union called a public health crisis: the law was needed to ensure that barbershops were sanitary. But our story of the origins of the Arkansas Board of Barber Examiners undercuts the current explanation of public health as the primary reason for its existence. The lack of evidence that customers were in any way affected by unsanitary barbers leads us to the conclusion that barbers really sought licensing and the creation of the board as means to seek rents and control the profession. Barbers wanted, and were successful in creating, barriers to entry. The JBIUA and barbers in Arkansas had fought for better wages and more secure employment, but unionization could only do so much. Union barbers across the country and in Arkansas found the solution in their state legislatures. Once the Board of Barber Examiners was created, barbers sought even more rents in the form of jobs on the board (only barbers sat on the board) and as inspectors.

Today, numerous policymakers and analysts have documented the negative effects that occupational licensing has on consumers by restricting competition in the industry. There is also a lively discussion about the potentially negative effects of licensing on minorities in the United States. After all, why should you have to be licensed to braid someone's hair? In fields such as barbers, licens-

¹⁰ Minutes provided by the Arkansas Board of Barber Examiners indicate that this practice continued until 1967.

ing regulations require more training to cut hair than to be a police officer.

While other scholars have investigated the consequences of such legislation, we believe more attention should be paid to the origins of occupational licensing laws. Our case study demonstrates that barbers in Arkansas captured the regulatory apparatus and used it to decrease competition in their profession and raise their wages. They also grandfathered themselves in to avoid the newly established requirements to receive a license. Finally, once the Barber Board was created, only barbers sat on the board and many of them engaged in rent seeking to secure positions both on the board and as inspectors. The creation of barber licensure in Arkansas was not established to protect public health but for private gain. Accordingly, we should be skeptical of occupations today that seek licensing for the “public good.”

References

- Ackerman, F. A. (1922) “Correspondence.” *Journeyman Barber Journal* 18 (3).
- Adams, F. A.; Jackson, J.; and Ekelund, R. (2002) “Occupational Licensing in a ‘Competitive’ Labor Market: The Case of Cosmetology.” *Journal of Labor Research* 23 (2): 261–78.
- Ambrey, G. W. (1922) “Correspondence.” *Journeyman Barber Journal* 18 (1).
- Arkansas Board of Barber Examiners (2019) “Minutes of the Arkansas Board of Barber Examiners Regular Meeting Held at Arkansas College of Barbering.” Little Rock, Arkansas (March 11). Available at www.arbarber.com/pdf/March112019-Minutes.pdf.
- Beaty v. Humphrey* (1938) 115 S.W.2d 559, State Auditor (Arkansas).
- Beaty v. Humphrey et al.* (1937) No. 56, 234 (Pulaski County).
- Bernstein, D. (1994) “Licensing Laws: A Historical Example of the Use of Government Regulatory Power against African Americans.” *San Diego Law Review* 31: 89–104.
- Blair, P. Q., and Chung, B. W. (2019) “How Much of Barrier to Entry Is Occupational Licensing?” *British Journal of Industrial Relations* 57 (4): 919–43.
- Bledstein, B. J. (1978) *The Culture of Professionalism: The Middle Class and the Development of Higher Education in America*. New York: W. W. Norton.

- Bluestone, D. M. (1991) "The Pushcart Evil: Peddlers, Merchants, and New York City's Streets, 1890–1940." *Journal of Urban History* 18 (1): 68–92.
- Booth, H. (1937) "Letter to Carl Bailey," March 22. Carl E. Bailey Papers, Barbers. Part 2, Box 1, Folder 21. Little Rock: Arkansas State Archives.
- Briggs, Z. (2019) "Bill Would Abolish Arkansas Barber Board, No License or Education Required to Cut Hair." *KATV News* (March 5).
- Bristol, D. W. (2015) *Knights of the Razor: Black Barbers in Slavery and Freedom*. Baltimore: Johns Hopkins University Press.
- Brush, H. E. (1922) "Correspondence." *Journeyman Barber Journal* 18 (4): 167.
- Burrow, J. G. (1977) *Organized Medicine in the Progressive Era: The Move toward Monopoly*. Baltimore: Johns Hopkins University Press.
- Callahan, M. (1937) "Letter to F. G. Martin," April 25. Carl E. Bailey Papers, Barbers. Part 3, Box 1, Folder 22. Little Rock: Arkansas State Archives.
- Camden Times* (1937) "Barber License Bill Is Passed: House Approved State Examination for Trade." *Camden Times* (February 25): 1.
- Carroll, S., and Gaston, R. (1983) "Occupational Licensing and the Quality of Service: An Overview." *Law and Human Behavior* 7 (2–3): 139–46.
- Chandler, A. D. Jr. (1977) *The Visible Hand: The Managerial Revolution in American Business*. New York: Belknap Press.
- Copeland, R. S. (1921) "Your Health." *Arkansas Democrat* (November 25): 10.
- Courier News* (1939) "Barbers Must Renew Licenses by Sept. 1." *Courier News* (August 14): 1.
- _____ (1944) "Barbers Claim Fees Too Cheap: Legislature Urged by Arkansas Group to Double Licenses." *Courier News* (May 24): 5.
- Daily Arkansas Gazette* (1911) "The Sanitary Barber Shop." *Daily Arkansas Gazette* (November 2): 4.
- Deyo, D. (2017) "Licensing and Service Quality: Evidence Using Yelp Consumer Reviews." Paper presented at San Jose University, San Jose, Calif.
- Downs, A. (1957) *An Economic Theory of Democracy*. New York: Harper & Brothers.

- Fishbein, M. (1932) "Dirty Barber Shops Often Spread Infection on Skin." *Courier News* (September 20): 4.
- _____ (1937) "Family Doctor: Cleanliness in Shaving Will Prevent Barber's Itch, Serious Skin Disease." *Hope Star* (October 27): 2.
- Friedman, L. (1965) "Freedom of Contract and Occupational Licensing, 1890–1910: A Legal and Social Study." *California Law Review* 53: 487–534.
- Furner, M. O. (1975) *Advocacy and Objectivity: A Crisis in the Professionalization of America Social Science, 1865–1905*. Lexington: University Press of Kentucky.
- Geison, G. L. (1983) *Professions and Professional Ideologies in America*. Chapel Hill: University of North Carolina Press.
- Hall, J., and Pokharell, S. (2016) "Barber Licensure and the Supply of Barber Shops: Evidence from U.S. States." *Cato Journal* 36 (3): 647–57.
- Hall, S. (1936) *The Journeyman Barbers' International Union of America*. Baltimore: Johns Hopkins University Press.
- Haskell, T. L. (2000) *The Emergence of Professional Social Science: The American Social Science Association and the Nineteenth-Century Crisis of Authority*. Baltimore: Johns Hopkins University Press.
- High, J., and Coppin, C. A. (1988) "Wiley and the Whiskey Industry: Strategic Behavior in the Passage of the Pure Food Act." *Business History Review* 62 (2): 286–309.
- Hilton, G. (1972) "The Basic Behavior of Regulatory Commissions." *American Economic Review* 62: 47–54.
- Hofstadter, R. (1955) *The Age of Reform: From Bryan to FDR*. New York: Vintage Books.
- _____ (1937) "Barber License Law Sought Here." *Hope Star* (October 13): 1.
- Klein, D. B.; Powell, B.; and Vorotnikov, E. S. (2012) "Was Occupational Licensing Good for Minorities? A Critique of Marc Law and Mindy Marks." *Econ Watch Journal* 9 (3): 210–33.
- Kleiner, M. (2000) "Occupational Licensing." *Journal of Economic Perspectives* 14 (4): 189–202.
- _____ (2017) "The Influence of Occupational Licensing and Regulation." *IZA World of Labor Journal*, Institute of Labor Economics (IZA): 392.

- Kleiner, M., and Vorotnikov, E. (2017) "Analyzing Occupational Licensing among the States." *Journal of Regulatory Economics* 52 (2): 132–58.
- Kolko, G. (1963) *The Triumph of Conservatism: A Reinterpretation of American History, 1900–1916*. New York: Free Press.
- Krueger, A. O. (1974) "The Political Economy of a Rent-Seeking Society." *American Economic Review* 64: 291–303.
- Larkin Jr., P. (2016) "Public Choice Theory and Occupational Licensing." *Harvard Journal of Law and Public Policy* 39 (1): 209–331.
- Lauterbach, C. (2020) "Ducey Signs Occupational Licensing Board Reform into Law." *Center Square* (June 8).
- Leaver, C. (2009) "Bureaucratic Minimal Squawk Behavior: Theory and Evidence from Regulatory Agencies." *American Economic Review* 99 (3): 572–607.
- Leonard, T. C. (2016) *Illiberal Reformers: Race, Eugenics, and American Economics in the Progressive Era*. Princeton, N.J.: Princeton University Press.
- Mahony, P. G. (2001) "The Political Economy of the Securities Act of 1933." *Journal of Law and Economics* 19 (2): 211–40.
- McChesney, F. (1987) "Rent Extraction and Rent Creation in the Economic Theory of Regulation." *Journal of Legal Studies* 16 (1): 101–18.
- Miller, J. A. (1922) "Barber License Law Held to Be Constitutional by the Supreme Court of Georgia." *Journeyman Barber Journal* 18 (3): 100.
- _____ (1922) "Correspondence." *Journeyman Barber Journal* 18 (3).
- Mills, Q. T. (2013) *Cutting Along the Color Line: Black Barbers and Barber Shops in America*. Philadelphia: University of Pennsylvania Press.
- Northwest Arkansas Times* (1938) "Barbers Discuss Forming Union." *Northwest Arkansas Times* (December 15): 1.
- Olson, M. (1965) *The Logic of Collection Action: Public Goods and the Theory of Groups*. Cambridge, Mass.: Harvard University Press.
- Owen, R. E. (1937) "Letter to Carl Bailey," March 14. Carl E. Bailey Papers, Barbers. Part 1, Box 1, Folder 20. Little Rock: Arkansas State Archives.

- Peltzman, S. (1976) "Toward a More General Theory of Regulation." *Journal of Law and Economics* 19 (2): 211–40.
- Pfeffer, J. (2014) "Administrative Regulation and Licensing: Social Problem or Solution?" *Social Problems* 21 (4): 468–79.
- Plemmons, A. (2019) "Occupational Licensing Effects on Firm Entry and Employment." Center for Growth and Opportunity at Utah State University, Working Paper No. 2019.008.
- Posner, R. (1974) "Theories of Economic Regulation." *Bell Journal of Economics and Management Science* 5 (2): 335–58.
- Shanessy, J. C. (1922) "President Page." *Journeyman Barber Journal* 17 (4): 152.
- _____ (1923) "President Page." *Journeyman Barber Journal* 18 (12): 454.
- Smith, J. B. (1922) "Correspondence." *Journeyman Barber Journal* 18 (4): 163.
- Stigler, G. (1971) "Theory of Economic Regulation." *Bell Journal of Economics and Management Science* 2 (April): 3–21.
- Thornton, R., and Timmons, E. (2015) "The De-licensing of Occupations in the United States." *Monthly Law Review* (May).
- Timmons, E. J., and Thornton, R. J. (2010) "The Licensing of Barbers in the USA." *British Journal of Industrial Relations* 48: 740–57.
- _____ (2018) "There and Back Again: The De-Licensing and Re-Licensing of Barbers in Alabama." *British Journal of Industrial Relations* 57: 764–90.
- Timmons, E., and Konieczny C. (2018) "Untangling Hair Braider Deregulation in Virginia." *Cato Journal* 38 (3): 679–99.
- Tullock, G. (1967) "The Welfare Costs of Tariffs, Monopolies and Theft." *Western Economic Journal* 5: 224–32.
- _____ (1993) *Rent Seeking*. Brookfield, Vt.: Edward Elgar.
- Tullock, G.; Seldon, A.; and Brady, G. (2002) *Government Failure: A Primer in Public Choice*. Washington: Cato Institute.
- Walls, C. A. (1937) "Letter to Carl Bailey," March 24. Carl E. Bailey Papers, Barbers. Part 2, Box 1, Folder 21. Little Rock: Arkansas State Archives.
- Washburn, A. H. (1934) "Here and There." *Hope Star* (May 10): 1.
- Whitaker, M. H. (1922) "Correspondence." *Journeyman Barber Journal* 18 (8): 310–11.
- Wiebe, R. (1967) *The Search for Order, 1877–1920*. New York: Hill and Wang.

- Worthall, L. (1922) “Organizer Dept.” *Journeyman Barber Journal* 18 (2): 55–56.
- Zapletal, M. (2019) “The Effects of Occupational Licensing: Evidence from Business Level Data.” *British Journal of Industrial Relations* 57 (4): 894–918.

HOW MISALIGNED INCENTIVES HINDER FOSTER CARE ADOPTION

Isabella M. Pesavento

Adoption, particularly adoption out of foster care, has not been well studied within the field of economics. Researchers may avoid this topic because the adoption market greatly deviates from a typical market, and the system and data collection are highly fragmented, with relatively little federal coordination. Rubin et al. (2007) and Thornberry et al. (1999) show that instability in foster care placements produces negative welfare outcomes, and Hansen (2006), Barth et al. (2006), and Zill (2011) demonstrate that adoption out of foster care is socially and financially beneficial. Yet, children waiting to be adopted out of foster care are in excess supply, which has been exacerbated in recent years. I hypothesize that this is, in part, due to misaligned incentives of government officials and the contracted foster care agencies. I show that earnings are prioritized over ensuring permanent child placement, which hinders the potential for adoption, and government oversight fails to correct such iniquities because of career interests.

Landes and Posner (1978) are the first to reference adoption agencies' misaligned incentives, though only briefly. Gronbjerg, Chen, and Stagner (1995) and Zullo (2002, 2008a, 2008b) discuss private agencies' use of leverage to win contracts and the prioritization of earnings over permanency outcomes. This article provides an updated evaluation of the role of incentives, and, drawing on

Cato Journal, Vol. 41, No. 1 (Winter 2021). Copyright © Cato Institute. All rights reserved. DOI:10.36009/CJ.41.1.7.

Isabella M. Pesavento is a Business Analyst with McKinsey & Company. She thanks John Welborn for comments on earlier versions of this article.

references to rent-seeking behavior and public choice theory, discusses how the interplay between contracted agencies and corresponding government officials' incentives hinders adoption out of foster care. First, I briefly characterize the market setting. Next, I analyze each party's incentives, and, finally, I conclude with a discussion of limitations and policy implications.

Characterizing the Market

While adoption is not frequently characterized in the context of a market, in its most basic form, adoption constitutes a transaction, with the “good”—the child—being transferred from the “supplier”—the foster care agency—to the “demander”—the parents. Government is a strong intermediary to help ensure the protection of child welfare (Moriguchi 2012).

The Supply Side

Children are placed in the foster care system either voluntarily, when parents who are unable to care for their child surrender their parental rights, or involuntarily, by court order in the case of abuse or neglect. For this reason, the children in foster care come disproportionately from troubled families. Although the total number of people under the age of 19 has not changed significantly in the last five years, the number of children in the foster care system has increased (Miller 2020). Neglect is the most common reason for the child's removal (62 percent), but the opioid epidemic has become an increasingly important factor. Drug abuse accounts for 36 percent of the removals (USCB 2019a), though it is higher in some states, including Ohio, where it is estimated to be 50 percent (Reynolds 2017). If a child remains in foster care for 15 of the most recent 22 months, or if the state deems the parents unfit for guardianship (Children's Bureau 2017), then the agency no longer aims to reunify the child with previous guardians; instead, parental rights are permanently terminated and the child becomes classified as “waiting to be adopted” (Bernal et al. 2007). The children will wait, on average, four years for adoption (USCB 2019a).

Currently, over 125,000 children in foster care await adoption (USCB 2019a). This “excess supply” has increased 25 percent between 2012 and 2018 (see Table 1), and for the past decade only about 50 percent of the number of children waiting for adoption

TABLE 1
FOSTER CARE NUMBERS AT A GLANCE

	2012	2014	2016	2018
In foster care on Sept. 30 of the FY	397,122	414,259	434,168	437,283
Waiting to be adopted	101,666	109,951	116,654	125,422
Adopted	52,039	53,555	57,238	63,123

SOURCE: Adoption and Foster Care Analysis and Reporting System.

actually do get adopted each year (USCB 2019b). Legislation has tried to promote reinstatement of birth parents' rights as a way to address this glut, but a steady increase in legal orphans aging out of the system has persisted (Taylor Adams 2014).

The Demand Side

Parents who adopt out of foster care are frequently characterized as having “a big heart and limited resources” (Bernal et al. 2007). Among these parents, 86 percent were found to be motivated by altruism (i.e., to provide a permanent home for a child) and 39 percent by infertility (multiple answers allowed in the survey). Parents often select foster care over other adoption venues because it is less expensive; concordantly, foster care adoptive parents typically come from lower income backgrounds (Bernal et al. 2007; Bethmann and Kvasnicka 2012).

Government Intermediary

Each state runs its own foster care system independently, though many rely significantly on federal funding through block grants, particularly Title IV-E of the Social Security Act (Taylor Adams 2014). Additionally, state governments have increasingly turned to private, often nonprofit organizations to be able to provide the full array of services needed, and thus the state-governmental role is primarily to set policy and provide oversight of private agencies (Krauskopf and Chen 2013). The contracts with private agencies vary widely, though they frequently stipulate some sort of fixed

amount of reimbursement per day or month per child in care (USDHHS 2008). Most states give priority to relatives and current foster parents who are looking to adopt (Ledesma n.d.), because states increasingly view foster care as a steppingstone to adoption (Adoption Exchange Association n.d.).

Value of Adoption Out of Foster Care

It is well documented that adoption is an extremely valuable outcome for children in foster care. Placement instability among foster care homes has been shown to have a significant negative effect on a child's well-being and future success (Rubin et al. 2007; Thornberry et al. 1999). Further, there is mounting evidence that quality parenting is extremely important for success at each stage of life (Reeves and Howard 2013; Kalil 2014), and that parenting need not be provided by a biological parent to achieve the same outcomes (Lamb 2012). Finally, Hansen (2006), Barth et al. (2006), and Zill (2011) have shown that adoption from foster care produces better future welfare and financial outcomes for both the child and society; compared to children who remain in foster care, children who are adopted out achieve better outcomes with regard to education, employment, criminal and disciplinary records, and social skills, among other categories. Adoption provides the government a net savings of \$143,000 per child, and each dollar spent on the adoption yields \$2.45 to \$3.26 in benefits to society (Hansen 2006).

Evaluating Incentives

Given that the choice to adopt a foster child ultimately reduces the government's fiscal burden and improves child and community outcomes, Hansen (2006) regards foster care adoption as a "positive externality." Accordingly, it seems logical that parties involved in the market should promote this transaction. It is in the child's best interest to be placed in a "foster-to-adopt" home as early as possible during their time in the system (Ledesma n.d.). Further, given that foster care parents themselves comprise 78 percent of nonrelative foster care child adoptions (USCB 2019a), placing a child in a safe and caring foster care family is of the utmost importance to promote ultimate adoption. However, it is clear that the current system does not optimally support adoption, as only half the number of children awaiting adoption are actually adopted

(USCB 2019b) and children are increasingly aging out of the system (Taylor Adams 2014).

Contracted Agencies: Earnings Focused

Many agencies' incentives conflict with child welfare and placement permanency goals. Zullo (2008a) suggests that privately contracted foster care agencies make decisions based on financial interests rather than child welfare. These agencies, whether for-profit or nonprofit, are typically paid per child in their care (USDHHS 2008). Therefore, each foster parent represents potential revenue (Zullo 2002) and many agencies provide bonuses to incentivize their workers to bring in more parents (CFUSS 2017). This means there is little incentive to reject inappropriate foster families, investigate concerns, or do anything that might cause the child to leave their program (Zullo 2002; Blackstone and Hakim 2003). In fact, on the contrary, Hatcher (2019) cites contract documents between the state and private foster care agencies that illustrate leadership officials sorting children not based on their needs but rather on how much revenue they can generate. As Zullo states, "What happens is the lives of these children become commodities" (Joseph 2015). As a result, private providers do not spend adequate time and resources on efforts to achieve permanent placement for children (Zullo 2008b).

In 2014, the Mentor Network was a leading provider of human and foster care services, operating in 36 states (CFUSS 2017). In 2015, *BuzzFeed News* and *Mother Jones* released a series of reports that suggested that Mentor placed children with neglectful and physically/sexually abusive foster care families as a way to boost profits. Former Mentor staffers stated, "The success of the program is defined by how many heads are in bed at midnight" and, "The bottom line is a dollar, not a child's well-being" (Joseph 2015). Another former employee admitted, "I became a machine that cared about profits. I didn't care about kids" (Roston and Singer-Vine 2015). Mentor, like half of surveyed agencies, receives almost 100 percent of its revenue from the government (CFUSS 2017: 7), yet it maintains profit margins in some states (Alabama, Ohio) as high as 31 percent and 44 percent (Roston and Singer-Vine 2015). Mentor's profit-oriented practice is not an isolated incident; in fact, a Florida agency publicly reports its techniques to "score" foster children to maximize revenue. And in a Maryland assessment report,

MAXIMUS, Inc., which has operated in 25 states, refers to foster children as a “revenue generating mechanism” (Hatcher 2019).

Nevertheless, Mentor’s scathing media coverage prompted the Senate Finance Committee to investigate privatization within foster care, using Mentor as a case study for the broader industry. It found that Mentor falsely reported that its death rate of children in its foster care system was in line with the national rate, when, in fact, it was 42 percent higher (CFUSS 2017: 24). The U.S. Senate Finance Committee found this report to be “inaccurate and misleading” (CFUSS 2017: 23) and found other reports to be ripe with inconsistencies, missing or inaccurate information, and “diagnostically implausible conditions” (CFUSS 2017: 22). Further, private for-profit agencies “too often failed to provide even the most basic protections” or to recognize and prevent dangerous conditions *ex ante* and *ex post* (CFUSS 2017: 2). With regard to both nonprofit and for-profit foster care agencies, the Committee ultimately concluded that “profits are prioritized over children’s well-being” (CFUSS 2017: 1).

Contracted Agencies: Rent Seeking

Mentor and other private agencies participate in rent-seeking behavior, which is defined as efforts to change laws and regulations in order to privilege one group over another (Dudley and Brito 2012). This practice enables the agencies to acquire and maintain their foster care contracts, which extends their hold over the industry. According to the U.S. Department of Health and Human Services, analysis of the private agencies’ contracts shows they are often “extremely lengthy, unduly complicated, and overly focused on details that [bear] little relationship to the critical issues that [need] to be addressed” (USDHSS 2008). Regulations with a high level of complexity have been found to reflect a high level of rent-seeking activity (Bessen 2016), so it stands to reason that the same may be the case here.

Lobbying and developing close relationships with government leadership is significant to many companies’ success in this industry. In the decade prior to the Senate Finance Committee investigation, Mentor spent \$1.6 million on lobbyists. Moreover, given that the two highest ranking officials on the Committee were Orrin Hatch (chairman) and Ron Wyden (ranking member), it is no coincidence that Mentor hired lobbyists with close relations to the two officials: Makan Delrahin, Hatch’s former adviser, and Josh Kardon, Wyden’s

former chief of staff, who was previously described as the senator's "alter ego" (Roston 2016).

MAXIMUS, Inc., has a similar history of lobbying for contracts in Los Angeles. The company's relationship with the county began in the mid-1980s, yet it has frequently fallen short of its contractual obligations. In 2007, an assessment of MAXIMUS's performance identified failures in five of eight performance categories mandated in their contract. When its contract was at risk of expiring without renewal in 2008, it donated the maximum amount allowed to reelect the county supervisors (who are in charge of renewing foster care contracts, including MAXIMUS's). This tactic had been successful in 2000, when it donated \$25,000 (circumventing campaign finance limitation rules) to a supervisor's political cause; consequently, the supervisor broke from his traditional voting record and supported MAXIMUS's contract. In 2008, the company spent \$124,000 on lobbying, which was the second highest amount any company spent on lobbying in that area. The lobbyists it hired had close ties to the supervisors: one was a supervisor's son and another was his former political consultant (Therolf 2008). MAXIMUS still works for Los Angeles County, with its most recent contract beginning in 2017 (County of Los Angeles: Department of Public Social Services 2016), despite the tragic and highly publicized death of Gabriel Fernandez in 2013, which exposed serious system flaws (County of Los Angeles, Department of Children and Family Services [DCFS] 2020).

In another example, a lobbying incident in North Carolina also jeopardized child welfare. There is nearly universal agreement that group home utilization should be limited, as it is more costly and produces worse welfare outcomes (Wulczyn et al. 2015; Shatzkin 2015; Barth 2002; Dozier et. al 2014). Connecticut and Rhode Island have reduced their usage of group care by as much as 20 percent, but North Carolina remains unchanged, which indicates to Richard Wexler, executive director of the National Coalition for Child Protection Reform, "that their group home industry is too powerful" (Wiltz 2019). In 2016, Congress considered a provision within a larger bill that would limit funding to group homes. In response, Baptist Children's Homes of North Carolina (BCH), one of North Carolina's largest contracted providers of group homes, which receives nearly \$5,000 per month per teenager it houses (Wiltz 2019), called on North Carolina legislators, particularly North Carolina Senator Richard Burr, to strip the provision (Baptist

Children's Homes 2016). As a result of BCH's outcry, Burr blocked the provision, and thus the entire foster care reform act was stripped from the bill (Wiltz 2019).

Finally, spending time, effort, and money to develop relationships with the relevant government agencies, as typical of rent-seeking behavior, is critical to the success of these private agencies. Gronbjerg et al. (1995) suggest that to win contracts, providers are incentivized to use connections and influence to form a relationship with the government, and this leverage may ultimately overshadow market forces. Mentor itself recognizes this behavior, saying they "rely in part on establishing and maintaining relationships with officials of various government agencies, primarily at the state and local level but also including federal agencies" (Roston 2016). Mentor, based in Boston, is infiltrated with Massachusetts ex-government officials who may be able to forge these connections: their chief marketing officer held many positions in the state government, most recently as treasurer (Roca n.d.), and the chair of the board was on the Senate Ways and Means Committee (Mass Inc. n.d.). The company spokesperson worked for the governor (Roston 2015), and the chief human resources officer is married to Senator Thomas McGee, head of the state Democratic party (Mentor Network n.d.).

Sequel, a private for-profit foster care business contracted by Oregon, also relies heavily on relationships. Marketing executives created a strong relationship with Glenda Marshall at Oregon's Child Welfare: they would invite her staff for dinner and she would send them cookies, among other practices (Dake 2019). Despite rising concerns and indiscretions with Sequel's child safety voiced nationwide and directly from Oregon state senators Kim Thatcher and Sara Gelser, Child Welfare leaders still signed a \$14.2 million two-year contract extension, and Marshall assisted Sequel to expand its facilities further into Oregon (*ibid.*).

Nonprofits

Critics may suggest that these aforementioned issues largely arise with for-profit businesses like Mentor and Sequel, and instead the government should only contract with nonprofits. A nonprofit business model theoretically could mitigate the profits-before-welfare mindset exhibited by for-profit businesses. However, given the recent indiscretions of some nonprofit foster care agencies, it is not clear that these businesses actually better utilize funds to promote

welfare than do the for-profits. For example, Oregon previously contracted with nonprofit Give Us This Day to support foster care services. Between 2009 and 2015, the organization founder, Mary Holden Ayla, stole approximately \$1 million of the federal funds the organization received. According to the Federal Bureau of Investigation report, funds chronically ran short from this organization; employees would race to the bank on payday only to have their paychecks still bounce, and when faced with a lack of food in the homes the organization supported, Ayla simply told employees to utilize the food offered at food banks. Ayla did not even have her own bank account, but rather used the Give Us This Day account to pay for her personal expenses, including payments to luxury retailers, spas, her home mortgage, or vacation travels. All of these indiscretions occurred without any state oversight (Federal Bureau of Investigation 2019).

In a similar way, numerous other nonprofits have misappropriated funds. For example, Children's Trust Fund, a nonprofit contracted by Los Angeles, requested government funds for children it did not even have in its care, and it funneled financial and nonmonetary donations to a personal account rather than to the DCFS as instructed (County of Los Angeles: Department of Auditor-Controller 2017). Child Link, a nonprofit contracted by Chicago, misused government funds for charges such as staff meals, traffic tickets, membership to a private social club, a \$1,000 outing to a White Sox game, and a \$6,000 Christmas party. These charges were made despite explicit rules dictating that state dollars could not be used on these types of expenses. Child Link was investigated at the urging of the state inspector general, and it marks the first time in six years that the Chicago DCFS actually conducted an extensive audit of one of its private partners (Gutoswki 2019).

Finally, in several states that do not allow or do not prefer to contract with for-profit agencies, a nonprofit agency may win the public contract and then subcontract out the services to for-profit agencies. This practice occurred in Illinois, which allowed Mentor (a for-profit agency) to effectively run state foster care services through a nonprofit front called Alliance Human Services, Inc. In this instance, the state inspector general investigation found that rather than working for the best interests of the children and families, the agency staff "cultivated a culture of incompetence and lack of forthrightness," and ultimately that "the absence of good faith demonstrated by the

private agency undermined any faith the department or the public would be able to place in the organization” (Kane 2015: 124). Despite these findings, among others, as well as a strong recommendation for the Illinois DCFS to terminate contracts with this agency, the DCFS still indicated a preference to work with this provider moving forward (Roston 2015).

Government Inaction

Bureaucratic leadership is complicit in allowing these behaviors to occur, which, in part, may serve to encourage these agencies’ indiscretions further. These officials may fail to provide adequate oversight and discipline out of interest in attaining acclaim, money, or career success. As a result, children can be placed with unsuitable families, which prevents the opportunity for them to form a good relationship with their foster parents that could ultimately lead to adoption. This line of reasoning is concordant with public choice theory, which recognizes that government officials are driven by self-interest, and thus they work to maximize their own private interests rather than public interest (Dudley and Brito 2012).

There is widespread evidence that state oversight agencies are aware of neglectful or failing contracted agencies, but government leadership takes little action to end the contract, impose consequences, or improve the situation. For example, in a 2018 New York Department of Investigation report, auditors found that 21 out of 22 contracted foster care providers fell short on the federal maltreatment guidelines, and 19 out of 22 fell short by more than double; the lowest three providers scored 2, 19, and 10 out of a possible 100 for maltreatment, and 47, 50, and 55 (out of 100) for safety. The Administration for Children’s Services (ACS) recognized that these were “very bad scores,” but did not regard any scores as “failing,” and typically took no action to address performance issues (Peters 2018: 8). Actually, the ACS renewed contracts with all foster care providers without instituting any additional performance or safety standards (Peters 2018).

In Oregon, a 2018 audit found that the Department of Human Services (DHS) compliance staff noted concerns and recommended not to renew the license of the contracted agency Give Us This Day in 2005, 2009, and 2014. Instead, DHS leadership continued to extend the contract until 2015, when the Department of Justice forced the provider to cease operations. The auditors noted that

Oregon DHS and Child Welfare suffer from “chronic and systemic management shortcomings that have a detrimental effect on the agency’s ability to protect child safety,” with managers completely unwilling to take responsibility for shortcomings, even within the program they directly oversee (Richardson and Memmott 2018).

In Oklahoma, the reviewers found that while the Oklahoma DHS itself had admitted serious accuracy issues with their reporting, staff had only taken “the most limited steps” to address them (Miller 2011: 2). There was a “lack of leadership, accountability, and there [was] no clear vision of the agency’s priorities” (ibid.). While in some ways this lack of oversight may be confounded or exacerbated by a lack of resources, in many cases, as in New York (Stein 2016) and Texas (United States District Court: Southern District of Texas 2015), investigative superiors have explicitly stated that failed oversight was not due to budget issues.

Revolving Door

This inaction reflects weak prioritization of child welfare. In fact, when the Senate Finance Committee requested information from all 50 states for their 2017 investigation, 17 states never responded to the request; the lead senator on the committee reacted by saying, “It’s kind of like some of these managers in states just consider protecting the children an afterthought, not a priority” (Grim and Chavez 2017). Instead, leadership’s priority may be personal career interests, as evidenced by the history of a “revolving door” between government and the private sector in this industry, which has been documented in other areas (Tabakovic and Wollmann 2018; Blanes-i-Vidal, Draca, and Fons-Rosen 2012; Cohen 1986). In order to encourage a future job possibility at a private agency, it is in the best interest of government leaders to award contracts and minimize discipline of private foster care agencies.

Many directors’ tenure forcibly ends due to ethical violations or failed oversight, but in the case of those who voluntarily resign, many leave to work at a contracted foster care agency. For example, when Mentor was founded, its chief customer was Massachusetts Division of Youth Services (DYS), headed by Edward Murphy. Murphy soon abdicated that position to become chief executive officer (CEO) of Mentor (Roston and Singer-Vine 2015). Bruce Naradella took over for Murphy at Massachusetts DHS, before then also taking over for Murphy as CEO of Mentor (United States Securities and Exchange

Commission 2018). In Los Angeles, David Sanders and Jackie Contreras, two DCFS directors within the past decade, both resigned to work at Casey Family Services, a private foster care provider (Casey Family Programs n.d.; Council of State Governments 2020). In fact, the current Casey CEO used to be commissioner of the New York ACS, and the executive vice president of Child and Family Services was executive director of the Colorado DHS (Casey Family Programs n.d.). In Georgia, over the last decade, two of the DCFS directors have left to work at private foster care agencies: Ron Scroggy at Together Georgia (Together Georgia 2017), and Virginia Pryor also at Casey (Los Angeles County Department of Children and Family Services n.d.).

Deflecting Criticism

Even without the suggestion of a “revolving door,” it seems clear that there is a desire of government leadership to deflect criticism, which at the very least suggests that officials are more concerned about their reputation than addressing welfare issues. In the Oklahoma audit, the reviewers stated, “there is evidence that senior managers have massaged these reports to make the numbers look better” (Miller 2011: 26), and that their quality control standards are monitored as to whether the reports “look like they will work” (Obradovic 2011: 3). In Los Angeles, an independent review reasoned that the DCFS’s extensively inconsistent classification of child deaths may be precipitated by incentives to deflect criticism (Office of Independent Review: Los Angeles County 2010). While the Los Angeles DCFS director said the inconsistencies were honest mistakes, a superior stated that “there are some reasons to believe that this is not just an accidental disconnect” (Zev Yaroslavsky: Los Angeles County Supervisor 2010). In 2015, a Texas judge presiding over the state foster care system ruled that the system was unconstitutional because it violated the 14th amendment right to be free from harm caused by the state (USDC: SDT 2015). Recently, the judge called the state’s inaction to improve the unconstitutional system “shameful” (Morris 2019). Further, with regard to its communication with her since 2015, she stated, “I cannot find DFPS [Department of Family and Protective Services] to be credible at any level” (ibid.). The judge is now choosing to rely only on court-appointed monitors to convey progress updates. Finally, former Georgia senator Nancy Schaefer has published public remarks stating: “I believe Child

Protective Services nationwide has become corrupt and that the entire system is broken almost beyond repair. I am convinced parents and families should be warned of the dangers” (Schaefer 2007). She describes Georgia’s DCFS as a “protected empire” and refers to them as “the ‘Gestapo’ at work” (ibid.).

Limitations and Policy Implications

In sum, private foster care agencies contracted by the state are financially incentivized to keep children in their care, no matter if conditions are unsafe, rather than prepare the child for adoption. They exhibit rent-seeking behavior to win and maintain contracts, which extends their hold on the industry. Government agencies provide little oversight, which may encourage this behavior further. As predicted by public choice theory, officials are more strongly motivated to protect their reputations and develop their careers than to discipline these agencies. As a result, foster care adoption has suffered; the number of children waiting for adoption has increased by 25 percent since 2012, and only about half of the number of children waiting for adoption actually do get adopted each year (USCB 2019b).

It is important to recognize limitations of this work. First, a few other documented factors also contribute to why adoption has not been optimally supported. Demand from parents to adopt from foster care is somewhat limited by a mismatch of parental preferences and child characteristics (Baccara et al. 2010). Other negative factors include the increasing availability of Assistive Reproductive Technologies (Gumus and Lee 2010) and the stigma attached to adoption (Small 2013). Second, some flaws in the foster care adoption system are already well recognized, including the system fragmentation and the scarcity of resources devoted to providing adoption services and social work (Hansen and Hansen 2006). There is also mounting evidence that the subsidy rate is too low and increasing it would improve the adoption rate out of foster care (Hansen 2007; Duncan and Argys 2007; Argys and Duncan 2013). All of these factors are significant, and their roles should not be understated. To better guide the focus of child welfare reform, future work should be done to quantify the relevant significance of misaligned incentives as presented in this article compared to these other factors.

Third, while I have provided evidence from many different systems to indicate that these problems are ubiquitous, this is not to say that the identified problems occur everywhere. Some contracts may be more clearly stipulated or better enforced. Many private agencies may not exploit the system in the ways I have mentioned, and many children do end up in good homes with loving foster parents; my motivation is rather to show that in many cases the current system largely provides such *opportunity* for exploitation, which can and does occur. My hope is that this initial thesis will lead to more targeted future work, including exploration of these results within each state, so that needs specific to each state can be identified and addressed precisely. Regardless, it seems clear that governments should rebid contracts with clearly specified quality expectations (Blackstone and Hakim 2003), and payment structures should better emphasize performance quality and permanency outcomes, which has been successful in some regions (Rubin et al. 2007). Creating greater competition within the market, with the government submitting bids to compete with private providers (managed competition), may help promote better performance as well (Blackstone and Hakin 2003). Finally, revolving door restrictions need to be in place for chief officials at the government offices.

References

- Adoption and Foster Care Analysis and Reporting System (AFCARS) (2012, 2014, 2016, 2018) Available at www.acf.hhs.gov/cb/research-data-technology/statistics-research/afcars.
- Adoption Exchange Association (n.d.) "About Adoption from Foster Care." *Adopt US Kids*. Available at www.adoptuskids.org/adoption-and-foster-care/overview/adoption-from-foster-care.
- Argys, L., and Duncan, B. (2013) "Economic Incentives and Foster Child Adoption." *Demography* 50 (3): 933–54.
- Baccara, M.; Collard-Wexler, A.; Felli, L.; and Yariv, L. (2010) "Gender and Racial Biases: Evidence from Child Adoption." Centre for Economic Policy Research Working Paper No. 2921.
- Baptist Children's Homes (2016) "Family First Update: Please Call Washington and Pray." Available at www.bchblog.org/post/2016/11/29/family-first-update-please-call-washington-and-pray.

- Barth, R. P. (2002) "Institutions vs. Foster Homes: The Empirical Base for a Century of Action." Chapel Hill, N.C.: Jordan Institute for Families, School of Social Work.
- Barth, R. P.; Lee, C. L.; Wildfire, J.; and Guo, S. (2006) "A Comparison of the Governmental Costs of Long-Term Foster Care and Adoption." *Social Services Review* 80 (1): 127–58.
- Bernal, R.; Hu, L.; Moriguchi, C.; and Nagypal, E. (2007) "Child Adoption in the United States: Historical Trends and the Determinants of Adoption Demand and Supply, 1951–2002." Working Paper. Available at <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.700.14&rep=rep1&type=pdf>.
- Bessen, J. (2016) "Accounting for Rising Corporate Profits: Intangibles or Regulatory Rents?" Boston University School of Law, Law & Economics Working Paper No. 16–18.
- Bethmann, D., and Kvasnicka, M. (2012) "A Theory of Child Adoption." Ruhr Economic Paper No. 342. Available at <http://ssrn.com/abstract=2084186>.
- Blackstone, E. A., and Hakim, S. (2003) "A Market Alternative to Child Adoption and Foster Care." *Cato Journal* 22 (3): 485–94.
- Blanes-i-Vidal, J.; Draca, M.; and Fons-Rosen, C. (2012) "Revolving Door Lobbyists." *American Economic Review* 102 (7): 3731–48.
- Casey Family Programs (n.d.) "Executive Team." *Casey Family Programs*. Available at www.casey.org/who-we-are/executive-team.
- CFUSS (Committee on Finance: United States Senate) (2017) *An Examination of Foster Care in the United States and the Use of Privatization*. Washington: Senate Finance Committee (October).
- Children's Bureau (2017) "Grounds for Involuntary Termination of Parental Rights." Available at www.childwelfare.gov/pubPDFs/groundtermin.pdf.
- Cohen, J.E. (1986) "The Dynamics of the 'Revolving Door' on the FCC." *American Journal of Political Science* 30 (4): 689–708.
- Council of State Governments (2020) "Strategy 2020: Safely Reducing Foster Care by Half." *Council of State Governments*. Available at www.csg.org/events/annualmeeting/policy_sessions_am09/policy2020_session.aspx.
- County of Los Angeles, Department of Auditor-Controller (2017) "Department of Children and Family Services: Children's Trust Fund Unit Review." Los Angeles, Calif.: Department of Auditor-Controller (May).

- County of Los Angeles, Department of Children and Family Services [DCFS] (2020) "Statement from the Department of Children and Family Services on 'the Trials of Gabriel Fernandez' Netflix Documentary Series." Los Angeles, Calif.: Department of Children and Family Services (March).
- County of Los Angeles, Department of Public Social Services (2016) "Department of Public Social Services: GAIN Case Management Services." Los Angeles, Calif.: Department of Public Social Services (March).
- Dake, L. (2019) "Out of State, Out of Mind." *Oregon Public Broadcasting* (December 11).
- Dozier, M.; Kaufman, J.; Kobak, R.; O'Connor, T. G.; Sagi-Schwartz, A.; Scott, S.; Shauffer, C.; Smetana, J.; van IJzendoorn, M. H.; and Zeanah, C. H. (2014) "Consensus Statement on Group Care for Children and Adolescents: A Statement of Policy of the American Orthopsychiatric Association." *American Journal of Orthopsychiatry* 84 (3): 219–25.
- Dudley, S. E., and Brito, J. (2012) *Regulation: A Primer*. Arlington, Va.: Mercatus Center.
- Duncan, B., and Argys, L. (2007) "Economic Incentives and Foster Care Placement." *Southern Economic Journal* 74 (1): 114–42.
- Federal Bureau of Investigation (2019) "Foster Care Agency Shut Down." *Federal Bureau of Investigation* (September 24).
- Grim, R., and Chavez, A. (2017) "Children Are Dying at Alarming Rates in Foster Care, and Nobody Is Bothering to Investigate." *The Intercept* (October 18).
- Gronbjerg, K. A.; Chen, T. H.; and Stagner, M. W. (1995) "Child Welfare Contracting: Market Forces and Leverage." *Social Services Review* 69 (4): 583–613.
- Gumus, G., and Lee, J. (2010) "The ART of Life: IVF or Child Adoption?" IZA Discussion Paper No. 4761. Bonn: IZA.
- Gutowski, C. (2019) "DCFS Says Nonprofit Misused Taxpayer Dollars, Demands Repayment of \$100K." *Chicago Tribune* (June 5).
- Hansen, M. E. (2006) "The Value of Adoption." Working Paper No. 2006–15. Washington: American University, Department of Economics.
- _____. (2007) "Using Subsidies to Promote the Adoption of Children from Foster Care." *Journal of Family Economic Issues* 28 (3): 377–93.

- Hansen, M. E., and Hansen, B. A. (2006) "The Economics of Adoption of Children from Foster Care." *Child Welfare* 85 (3): 559–83.
- Hatcher, D. L. (2019) "Stop Foster Care Agencies from Taking Children's Resources." *Florida Law Review Forum* 71 (1): 104–12.
- Joseph, B. (2015) "The Brief Life and Private Death of Alexandria Hill." *Mother Jones* (February 26).
- Kalil, A. (2014) "Proposal 2: Addressing the Parenting Divide to Promote Early Childhood Development for Disadvantaged Children." Hamilton Project. Washington: Brookings Institution.
- Kane, D. (2015) *Report to the Governor and the General Assembly*. Illinois. Office of the Inspector General: Illinois Department of Children and Family Services (January).
- Krauskopf, J., and Chen, B. (2013) "Administering Services and Managing Contracts: The Dual Role of Government Human Services Officials." *Journal of Policy Analysis and Management* 29 (3): 625–28.
- Lamb, M. E. (2012) "Mothers, Fathers, and Circumstances: Factors Affecting Children's Adjustment." *Applied Developmental Science* 16 (2): 98–111.
- Landes E. M., and Posner A. R. (1978) "The Economics of the Baby Shortage." *Journal of Legal Studies* 7 (2): 323–48.
- Ledesma, K. (n.d.) "Adopting from Foster Care." *Adoptive Families*. Available at www.adoptivefamilies.com/how-to-adopt/foster-care-adoption/adopting-from-foster-care.
- Los Angeles County Department of Children and Family Services (n.d.). "Virginia Pryor." Available at <https://dcfs.lacounty.gov/about/who-we-are/leadership/virginia-pryor>.
- Mass Inc. (n.d.) "Greg Torres." Available at <https://massinc.org/author/gregtorres>.
- Mentor Network (n.d.). "Executive Team: Maria McGee." *The Mentor Network*. Available at www.thementornetwork.com/team-member/maria-mcgee.
- Miller, A. (2020) "Adoption & Child Welfare Services in the US." Industry Report 62411. IBISWorld. Available at www.ibisworld.com/united-states/market-research-reports/adoption-child-welfare-services-industry.
- Miller, V. P. (2011) "A Review of the Oklahoma Department of Human Services' Child Welfare Practices from a Management Perspective." Independent review prepared for the plaintiffs in

- D.G. v. Henry*. Available at www.childrensrights.org/wp-content/uploads/2011/03/2011-03-15_ok_miller_management_review_final.pdf.
- Moriguchi, C. (2012) “The Evolution of Child Adoption in the United States, 1950–2010: Economic Analysis of Historical Trends.” Institute of Economic Research, Hitotsubashi University, Discussion Paper Series A No. 572. Tokyo: Hitotsubashi University Repository.
- Morris, A. (2019) “Judge Fines Texas \$50K a Day over ‘Broken’ Foster Care System.” *San Antonio Express News* (November 5).
- Obradovic, Z. (2011) “Report on the KIDS System: Review and Analysis.” Analysis prepared for the plaintiffs in *D.G. v. Henry*. Available at https://www.childrensrights.org/wp-content/uploads/2011/03/2011-03-15_ok_obradovic_kids_system_review_final.pdf.
- Office of Independent Review: Los Angeles County (2010) *Re: Status of Implementation of SB39: Current Challenges*. Los Angeles, Calif. Department of Independent Review (August). Available at <http://zevyaroslavsky.org/wp-content/uploads/dcfcs-08-30-10-BOS-Ltr-DCFS-SB39-Report.pdf>.
- Peters, M. G. (2018) “DOI Report Finds the City Administration for Children’s Services Did Not Sufficiently Hold Foster Care Providers Accountable for Safety of Children in Their Care.” Report on the Administration for Children’s Services’ from the City of New York’s Department of Investigation. Available at www1.nyc.gov/assets/doi/reports/pdf/2018/Oct/ACS_Rpt_Release_Final_10122018.pdf.
- Reeves, R. V., and Howard, K. (2013) “The Parenting Gap.” Brookings Institution Report, Center on Children and Families. Available at www.brookings.edu/wp-content/uploads/2016/06/09-parenting-gap-social-mobility-wellbeing-reeves.pdf.
- Reynolds, D. (2017) “Ohio Foster Care System Flooded with Children Amid Opioid Epidemic.” *CBS News* (August 7).
- Richardson, D., and Memmott, K. (2018) “Foster Care in Oregon: Chronic Management Failures and High Caseloads Jeopardize the Safety of Some of the State’s Most Vulnerable Children.” Oregon Audits Division, Report No. 2018–05. Available at <https://sos.oregon.gov/audits/Documents/2018-05.pdf>.
- Roca (n.d.). “Board of Directors: Dwight Robson.” *Roca*. Available at <https://rocainc.org/team/dwight-robson>.

- Roston, A. (2015) “‘Culture of Incompetence’: For-Profit Foster-Care Giant Is Leaving Illinois.” *BuzzFeed News* (April 17).
- _____. (2016) “Foster Care Company’s ‘Abnormal Level of Lobbying.’” *BuzzFeed News* (January 12).
- Roston, A., and Singer-Vine, J. (2015) “Fostering Profits: Abuse and Neglect at America’s Biggest For-Profit Foster Care Company.” *BuzzFeed News* (February 20).
- Rubin, D. M.; O’Reilly, A. L. R.; Luan, X.; and Localio, A. R. (2007) “The Impact of Placement Stability on Behavioral Well-Being for Children in Foster Care.” *Pediatrics* 119: 336–44.
- Schaefer, N. (2007) “The Corrupt Business of Child Protective Services.” Available at <http://fightcps.com/pdf/thecorruptbusinessofchildprotectiveservices.pdf>.
- Shatzkin, K. (2015) “Every Kid Needs a Family: Giving Children in the Child Welfare System the Best Chance for Success.” Annie E. Casey Foundation, Policy Brief.
- Small, J. (2013) “Adopted in America: A Study of Stigma.” Available at <https://ssrn.com/abstract=2280517>.
- Stein, J. (2016) “Three Arrested as City DOI Releases Scathing Report on Close to Home Oversight.” *NYNMEDIA* (April 14).
- Tabakovic, H., and Wollmann, T. G. (2018) “From Revolving Doors to Regulatory Capture? Evidence from Patent Examiners.” NBER Working Paper No. 24638. Cambridge, Mass.: National Bureau of Economic Research.
- Taylor Adams, L. (2014) “Legal Orphans Need Attorneys to Achieve Permanency.” *American Bar Association* (December 1).
- Therolf, G. (2008) “Low-Rated Firm Fights to Keep Rich County Work.” *Los Angeles Times* (October 30).
- Thornberry, T. P.; Smith, C. A.; Rivera, C.; Huizinga, D.; and Stouthamer-Loeber, M. (1999) “Family Disruption and Delinquency.” Available at www.ncjrs.gov/pdffiles1/ojdp/178285.pdf.
- Together Georgia (2017) “Former TG Director, Ron Scroggy, Receives Big Voice Award.” *Together Georgia* (September 6).
- United States District Court: Southern District of Texas (USDC: SDT) (2015) Memorandum Opinion and Verdict of the Court in “*M.D. et al. v. Abbott et al.*,” entered on December 17. Available at <https://cases.justia.com/federal/district-courts/texas/txsdce/2:2011cv00084/876792/368/0.pdf?ts=1450456518>.

- United States Securities and Exchange Commission (2018) "Civitas Solutions, Inc." Proxy Statement Pursuant to Section 14 (a) of the Securities and Exchange Act of 1934. Available at www.sec.gov/Archives/edgar/data/1608638/000160863818000003/fyl7proxy.htm.
- USCB (U.S. Children's Bureau) (2019a) "Adoption and Foster Care Reporting System (AFCARS) Report No. 26." Washington: U.S. Department of Health and Human Services, Administration for Children and Families.
- _____ (2019b) "Trends in Foster Care and Adoption: FY 2009–FY 2018." Washington: U.S. Department of Health and Human Services, Administration for Children and Families.
- USDHHS (U.S. Department of Health and Human Services) (2008) Office of the Assistant Secretary for Planning and Evaluation. "Preparing Effective Contracts in Privatized Child Welfare Systems." Washington: U.S. Department of Health and Human Services (August 30).
- Wiltz, T. (2019) "The Feds Are Cutting Back on Group Homes. Some Say That's a Big Mistake." *Pew Charitable Trusts* (July 8). Available at www.pewtrusts.org/en/research-and-analysis/blogs/stateline/2019/07/08/the-feds-are-cutting-back-on-foster-group-homes-north-carolina-says-thats-a-big-mistake.
- Wulczyn, F.; Alpert, L.; Martinez, Z.; and Weiss, A. (2015) "Within and Between State Variation in the Use of Congregate Care." Report for the Center for State Child Welfare Data, University of Chicago (June).
- Zev Yaroslavsky: Los Angeles County Supervisor (2010) "Secret Child-Death Records to Be Revealed." *Zev Yaroslavsky* (August).
- Zill, N. (2011) "Adoption from Foster Care: Aiding Children While Saving Public Money." Brookings Institution, Center for Children and Families Brief No. 43 (May).
- Zullo, R. (2002) "Private Contracting of Out-of-Home Placements and Child Protection Case Management Outcomes." *Child and Youth Services Review* 24 (8): 583–600.
- _____ (2008a) "Child Welfare Privatization and Child Welfare: Can the Two Be Efficiently Reconciled?" Ann Arbor, Mich.: Labor Studies Center, University of Michigan.
- _____ (2008b) "Is Social Service Contracting Coercive, Competitive, or Collaborative?" *Administration in Social Work* 30 (3): 25–42.

THE ECONOMIC IMPACT OF TAX CHANGES, 1920–1939

Alan Reynolds

Estimates of the elasticity of taxable income (ETI) investigate how high-income taxpayers faced with changes in marginal tax rates respond in ways that reduce expected revenue from higher tax rates, or raise more than expected from lower tax rates. Diamond and Saez (2011) pioneered the use of a statistical formula, which Saez developed, to convert an ETI estimate into a revenue-maximizing (“socially optimal”) top tax rate. For the United States, they found that the optimal top rate was about 73 percent when combining the marginal tax rates on income, payrolls, and sales at the federal, state, and local levels. A related paper by Piketty, Saez, and Stantcheva (2014) concluded that, at the highest income levels, the ETI was so small that comparable top tax rates as high as 83 percent could maximize short-term revenues, supposedly without suppressing long-term economic growth. Such studies could be viewed as part of a larger effort to minimize any efficiency costs of distortive taxation while maximizing assumed revenue gains and redistributive benefits.

A previous article in this journal, “Optimal Tax Rates: A Review and Critique” (Reynolds 2019), analyzed such U.S. postwar ETI estimates that were being misconstrued as recommendations for a 73–83 percent optimal top tax rate for the federal income tax alone. I surveyed evidence and arguments suggesting that even if top tax rates designed to maximize short-term revenue might be “socially optimal”

in some sense, such high marginal rates would not prove to be economically optimal in terms of the incentive effects on sources of longer-term expansion of the economy and the tax base. As Goolsbee (1999: 38) rightly emphasized, “The fact that efficiency costs rise with the square of the tax rate is likely to make the optimal rate well below the revenue maximizing rate.”

The early paper by Goolsbee included estimates of the ETI in the 1920s and 1930s. Together with another paper about the ETI during those years, by Romer and Romer (2014), the prewar studies came to nearly the same conclusion as their postwar counterparts did—namely, that hypothetical top tax rates of 74–83 percent could have maximized federal tax revenues during the Great Depression. Unlike Diamond and Saez (2011), however, the prewar studies excluded state and local taxes (which were much larger than federal taxes) and major new federal taxes on payrolls and sales added in 1932–37.

Romer and Romer (2014: 269) use an average of ETI estimates for federal income tax changes from 1918 to 1941 to conclude that “our estimated elasticity of 0.21 implies an optimal top marginal rate of 74 percent” (2014: 269). Yet their estimated elasticity is twice that high for 1932 (0.42) when the top marginal rate was raised from 25 percent to 63 percent. And their elasticity coefficients for major tax changes in 1934–38, they acknowledge, “cannot be estimated with any useful degree of precision” (2014: 266).

Goolsbee (1999: 36) compares 1931 and 1935 to judge how high-income taxpayers responded to much higher tax rates in 1932 (and higher still in 1934). He concludes the ETI at high incomes was so low that “if there were only one rate in the tax code, the revenue maximizing tax rate given the [low] elasticity estimated . . . [would be] 83 percent using the using 1931 to 1935 data.”

When discussing a smaller 1936 rate increase, confined to incomes above \$50,000, Goolsbee concludes: “Technically, the revenue-maximizing [single tax] rate would be at the maximum of 100 percent using 1934–38 data, since the elasticity was negative” (ibid). A study of postwar data by Piketty, Saez, and Stantcheva (2014: 252) likewise theorized that “the optimal top tax rate . . . actually goes to 100 percent if the real supply-side elasticity is very small.” My review of that paper (Reynolds 2019: 250–54) found their estimated ETI and Pareto parameters to be far below consensus estimates for high incomes and inapplicable to untested tax rates of 83–100 percent. This review of similar prewar studies also finds their ETI estimates

implausibly low and the alleged revenue-maximizing tax rates of 74–83 percent too high.

Romer and Romer (2014) are incorrect in claiming that tax responsiveness was low in the 1920s and 1930s, and Goolsbee is incorrect in making that same claim about just the 1930s. Their erroneous low response estimates lead them to conclude that high tax rates are a good thing. This study finds, instead, that high income taxpayers were very responsive to lower marginal tax rates in the 1920s and higher marginal tax rates in the 1930s.

I find that large reductions in marginal tax rates on incomes above \$50,000 in the 1920s were *always* matched by large *increases* in the amount of high income reported and taxed. Large increases in marginal tax rates on incomes above \$50,000 in the 1930s were *almost* always matched by large *reductions* in the amount of high income reported and taxed, with a brief exception connected with the 1937–38 recession, which is investigated in detail.

The Folly of Raising Taxes in a Deep Depression

An earlier generation of economists found that raising tax rates on incomes, profits, and sales in the 1930s was inexcusably destructive. In 1956, MIT economist E. Cary Brown pointed to the “highly deflationary impact” of the Revenue Act of 1932, which

pushed up rates virtually across the board, but notably on the lower-and middle-income groups. The scope of the act was clearly the equivalent of major wartime enactments. Personal income tax exemptions were slashed, the normal-tax as well as surtax rates were sharply raised, and the earned-income credit equal to 25 per cent of taxes on low incomes was repealed [Brown 1956: 868–69].

In 1958, Arthur Burns wrote:

If prosperity is to flourish, people must have confidence in their own economic future and that of their country. This basic truth was temporarily lost sight of during the 1930's . . . In the five years from 1932 to 1936, unemployment . . . at its highest was 13 million or 25 percent of the labor force. The existence of such vast unemployment did not, however, deter the federal government from imposing new tax burdens. . . . The new taxes encroached on the spending power of both consumers

and business firms at a time when production and employment were seriously depressed. Worse still, they spread fear that the tax system was becoming an instrument for redistributing incomes, if not for punishing success [Burns 1958: 27–28].

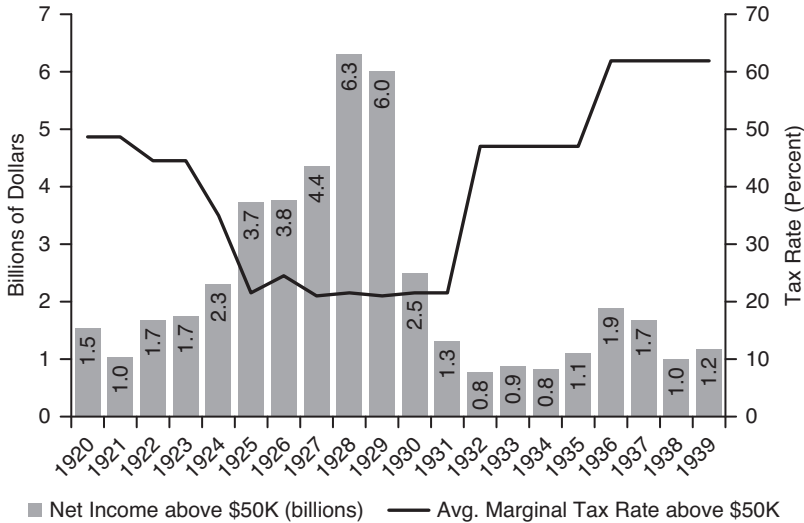
In 1966, Herbert Stein referred to President Hoover's 1932 policies as "the desperate folly of raising taxes in a deep recession" (Stein 1966: 223). In contrast, Romer and Romer (2014) viewed their ETI estimates for 1932–38 as evidence that enormous tax increases in those years had no visible adverse effects on the Depression. To demonstrate the supposedly negligible impact of much higher income and excise taxes in 1932, they enumerate a few upbeat statistics about short-term business conditions. Meanwhile, Cole and Ohanian (1999) and Mulligan (2002) have been even more vocal in asserting that federal income and excise tax increases during 1932–36 share no responsibility for the depressing performance of the economy (and income tax receipts) from 1930 to 1940. The final sections of this article question the "taxes don't matter" arguments and evidence of Romer and Romer, Cole and Ohanian, and Mulligan. Before doing so, however, we must first begin with a scenic detour of some new graphical evidence suggesting that most ETI estimates in Romer and Romer, and Goolsbee, are implausibly low, particularly for higher tax rates in 1932 and 1936.

A Graphical Illustration of Elasticity of Taxable Income, 1920–1939

Figure 1 illustrates yearly connections between (1) changes in the average of all marginal tax rates applied to annual incomes above \$50,000, and (2) the amount of net income, in billions of dollars, reported at incomes above \$50,000. Taxable incomes of high-income taxpayers have grown rapidly when their marginal tax rates were low, and were flat or falling when their marginal tax rates were high. The only apparent exception was 1936–37 when taxable earnings above \$50,000 *briefly* reached the equally unimpressive lows of 1922–23.

Romer and Romer (2014: 248, 252) define "high income" as the top 0.05 percent of income earners, which comprise "about 25,000 returns per year." Taxpayers in that group and the incomes required to be included don't remain constant from year to year. Indeed, "net

FIGURE 1
AVERAGE MARGINAL TAX RATES ON INCOMES ABOVE
\$50,000 AND NET INCOME REPORTED, 1920–1939



SOURCE: Statistics of Income (SOI) Tax Stats Archive.

income cutoffs for being in this group ranged from \$25,400 in 1933 to \$75,100 in 1928” (2014: 248). Consequently, defining high income as a percentage of income makes it a moving target for studying taxpayer response. Romer and Romer (2014) allocate marginal tax rates according to incomes on tax returns. But an income of \$25,400 in 1933 faced only a 21 percent marginal rate in 1933—one-third of the top tax rate of 63 percent that year and even lower than the 23 percent marginal tax on \$75,100 in 1928. Thus, we are unlikely to find a meaningful estimate of how high-income taxpayers react to high marginal tax rates by measuring how they reacted to marginal tax rates as low as 21 percent.

Figure 1 defines high income in a simpler way that is more transparently linked to tax rate schedules—namely, as net (taxable) income above \$50,000—which, in 1935, included 10,680 tax returns and made up the top 0.33 percent of taxpayers (Tax Policy Center 2019b). That threshold combines the two highest of Goolsbee’s three high-income groups. It matches the definition of affluence in FDR’s

1936 tax law, which raised tax rates only on those earning over \$50,000. Net income figures above \$50,000 are added up from tax returns (SOI Tax Stats Archive). The graph is confined to 1920–39 to minimize possible distortions for 1918–19 caused by WWI and for 1940–41 by rearmament, though including those years would not make a great difference except for 1941.

The recession from January 1920 to July 1921 would have reduced high incomes regardless of tax policy. Yet the quest for a strong recovery from that recession was a major reason the average of various marginal tax rates on incomes of \$50,000 or more (in Figure 1) was cut by more than half—from 48.7 percent in 1921 to 44.5 percent in 1922–23, to 35 percent in 1924, to 21.5 percent from 1925 to 1931.

Tax rates on net income below \$10,000 were also reduced from 4–11 percent in 1921 to 1.5–5 percent from 1925 to 1931, and the personal exemption for couples was raised from \$2,000 in 1920 to \$3,500 from 1925 to 1931. Taxes paid by lower-income groups did fall. However, total individual tax receipts rose quickly—from \$704 million in 1924 to nearly \$1.2 billion in 1928. That is because the share of individual income tax paid by the over-\$50,000 group nearly doubled in seven years—from 44.2 percent in 1921 to 51.8 percent in 1922, 69 percent in 1925, and 78.4 percent in 1928 (Frenze 1982). Moreover, the amount of revenue collected from high incomes above \$50,000 nearly tripled at that time—rising from \$318 million in 1921 to \$507 million in 1925, and \$909 million in 1928.

In Figure 1, the average marginal tax rate is an *unweighted* average of *statutory* tax brackets applying to all income groups reporting more than \$50,000 of income. After President Hoover's June 1932 tax increase (retroactive to January) the number of tax brackets above \$50,000 quadrupled from 8 to 32, ranging from 31 percent to 63 percent. The average of many marginal tax rates facing incomes higher than \$50,000 increased from 21.5 percent in 1931 to 47 percent in 1932, and 61.9 percent in 1936. One of the most striking facts in Figure 1 is that the amount of reported income above \$50,000 was almost cut in half in a single year—from \$1.31 billion in 1931 to \$776.7 million in 1932. Of course, we expect to see more high incomes on tax returns during years when the economy was growing rapidly such as 1925 to 1929. Yet real GDP also grew by 10.9 percent a year from 1934 to 1936, and high incomes still remained as depressed as they were in 1922–23, the last time marginal tax rates had been so high with the economy barely out of recession.

In the eight years from 1932 to 1939, the economy was in cyclical contraction for only 28 months. Even in 1940, after two huge increases in income tax rates, individual income tax receipts remained lower (\$1,014 million) than they had been in the 1930 slump (\$1,045 million) when the top tax rate was 25 percent rather than 79 percent. Eight years of prolonged weakness in high incomes and personal tax revenue after tax rates were hugely increased in 1932 cannot be easily brushed away as merely cyclical, rather than a behavioral response to much higher tax rates on additional (marginal) income.

Just as income (and tax revenue) from high-income taxpayers rose spectacularly after top tax rates fell from 1921 to 1928, high incomes and revenue fell just as spectacularly in 1932 when top tax rates rose. These inverse swings in top tax rates and revenues from 1918 to 1939 would be inexplicable if the ETI had been nearly as low as estimated by Goolsbee or Romer and Romer.

What Figure 1 shows is that whenever the higher tax rates went up, the amount of income subjected to those rates always went down, and when those tax rates went down, the amount of higher incomes (and taxes collected from them) always went up. ETI at high incomes was pronounced and evident in every year from 1918 to 1939, with the partial and brief exception of 1936–37 (to be discussed later), which ended as badly as the ill-fated “tax increase” of 1932.

Figure 1 reveals that the amount of high income reported on individual tax returns (including business income) was *persistently* high when top marginal rates were low (1924–30), and reported high income was likewise *persistently* low when top marginal rates were high (1920–23 and 1932–39). This evidence is consistent with persistently high elasticity of taxable income at high incomes.

The following sections discuss the most significant changes in federal tax law in the 1920s and 1930s and explain why my interpretation differs from that of Romer and Romer or Goolsbee, and why they differ from each other.

1922–1929: Top Tax Rate Falls from 73 Percent to 25 Percent and Revenues Soar

Romer and Romer’s ultra-low 0.21 elasticity calculation for the entire 1918 to 1941 period invites the impression that not only were large increases in all tax rates harmless to the economy during the Great Depression, but also that large reductions in all tax rates in

1922–25 supposedly get no credit for the booming economy in 1922–29. To investigate such questions, however, requires looking at what happened when tax rates were changed, rather than looking at a pooled average over many years.

After the First World War ended on November 11, 1918, there was deflationary recession from January 1920 to July 1921. To facilitate recovery, the top U.S. marginal income tax rate was reduced from 73 percent in 1921 to 58 percent in 1922, and later to 46 percent in 1924 and 25 percent in 1925. At an income just above \$50,000, marginal additions to income were subject to a 31 percent tax in 1923, 24 percent in 1924, and 18 percent in 1925 (Tax-Brackets.org). All tax rates at lower incomes were also greatly reduced, and the personal exemption for married couples was increased from \$2,500 to \$3,500. Real GDP grew by 4.7 percent a year from 1922 to 1929, or 3.2 percent on a per capita basis (Johnston and Williamson 2019).

The effect of lower marginal tax rates in lifting higher incomes after 1925 is apparent in Figure 1, and in the research of Goolsbee (1999). Romer and Romer (2014), on the other hand, present their elasticity estimates as averages over dissimilar multiyear periods rather than for the specific years in which tax rates were actually changed. They offer a choice of two long-term averages: one for 24 years (1918–41), and the other for 10 years (1923–32).

Rather than simply excluding the war year of 1918 and rearmament years of 1940–41, Romer and Romer (2014: 269) say, “restricting the analysis to . . . 1923–1932, a period well away from both world wars . . . increases the estimated elasticity” from 0.21 to 0.38, which “implies an optimal top tax rate of 61 percent.” One problem with that 1923–32 pooled average is that it cannot explain what happened when the top tax rate fell from 58 percent to 25 percent in 1923–25, because 1923–32 ends with three years of falling GDP and the top tax rate rising from 25 percent to 63 percent. A bigger problem is that 1923–32 excludes the entire Roosevelt presidency.

Romer and Romer’s method of comparing two adjacent years is also complicated for the tax laws of 1924 and 1926 because they reduced tax rates retroactively in 1923 and 1925. Thus, Romer and Romer calculate elasticities for the year of enactment, thereby showing relatively low responsiveness in 1926 to tax rates that were reduced on income earned in 1925 (and reported on tax returns in April 1926). Yet the retroactive 1925 tax cut enacted on February 26, 1926, was surely *anticipated* in 1925. Treasury Secretary Andrew

Mellon and President Coolidge repeatedly argued for lower surtax rates, including the 1924 Treasury Report and Mellon's Ways and Means testimony October 19, 1925 (Romer and Romer 2012: 10).

Goolsbee's comparison of years before and after the retroactive 1923 and 1925 tax rate cuts may be more instructive in this case than ignoring them. He looks at what happened to reported high incomes after the top tax rate fell from 73 percent to 58 percent in 1922, 47 percent in 1923, and 25 percent in 1925. His difference-in-differences elasticities for high incomes in this period are "relatively large. . . . Two of the implied elasticities are around 0.6 to 0.7, and the third is 1.24." A simpler calculation that includes lower incomes (above \$5,000) also yields relatively high elasticities of 0.52 to 0.64 (Goolsbee 1999: 25). Elasticities of 0.60 to 1.24 *at high incomes* are consistent with most postwar estimates (Reynolds 2019).

Goolsbee (1999) found that lowering the top marginal tax rate from 58 percent to 25 percent generated large increases in the amount of high income available to tax, as is undeniably apparent in Figure 1. Estimates that find such high elasticity of taxable income do not by themselves explain whether such responsiveness mainly reflects (1) tax-induced changes in tax avoidance or (2) changes in real activity that contributes to increased real GDP over time.

Aggressive use of tax deductions is a common way to avoid high tax rates, and legal deductions in the 1920s included such potential loopholes as business and partnership losses, interest paid, contributions, and a do-it-yourself line for "other." Goolsbee (1999: 19) "rules out tax induced changes to deductions" by assuming "the ratio of taxable to gross income is constant." This means his finding of high elasticity in 1922–26 does *not* suggest that lower rates merely reduced the incentive to maximize tax deductions. Instead, high ETI in the late 1920s probably demonstrated *real* supply-side effects—such as greater innovation, work effort, entrepreneurial risk taking—and investment in human, physical, and intangible capital. Goolsbee's finding of a high ETI at a time when the top tax rate fell from 58 percent to 25 percent is therefore consistent with the observed rapid growth of the economy and federal tax revenue during the late 1920s (de Rugy 2003).

Confirming Goolsbee's results for this period, Figure 1 shows that when the top rate fell to 56 percent in 1922–23, reported high income jumped from \$1.04 billion in 1921 to \$1.64 billion in 1922. When the top rate fell to 46 percent in 1924, high income rose again

to \$2.3 billion. When the top rate fell to 25 percent in 1925, high income rose once again to \$3.74 billion and then kept climbing to \$6.3 billion by 1928. Lower marginal tax rates on high incomes were clearly followed by a *much* larger amount of taxable income at the top, which defines high ETI.

As the amount of reported high income rose with lower marginal tax rates, the amount of *revenue* collected from high incomes also increased dramatically. Tax revenues from all but the highest-income taxpayers were reduced by larger personal exemptions and lower tax rates, yet individual tax receipts nonetheless nearly doubled in four years—from \$704 million in 1924 to nearly \$1.2 billion in 1928 (de Rugy 2003).

Gwartney (2002) found that real tax revenue, measured in constant 1929 dollars, collected from those earning less than \$50,000 fell by 45 percent, while real revenue collected from those earning more than \$50,000 rose by 63 percent. In nominal terms, taxpayers with incomes above \$100,000 paid \$302 million in federal income tax in 1922, \$373 million in 1926, and \$714 million in 1928 (de Rugy 2003).

The impressive surge in income tax receipts from high incomes after the top tax rate fell from 73 percent in 1921 to 25 percent in 1925–26 is *inconsistent* with the Romer and Romer (2014) theme that high-income taxpayers were unresponsive to marginal tax rates. But such actual revenues statistics cannot be found in Romer and Romer. What are instead shown in their “Table 1: Significant Interwar Tax Legislation” are vintage *ex ante* revenue *estimates* rather than actual *ex post* revenues (2014: 245). As a result, that collection of static revenue estimates mistakenly depicts tax revenue *falling* with every tax rate reduction from 1921 through 1926, adding up to a wrongly estimated \$1.5 billion cumulative revenue decline while revenues were growing very quickly.

1932: Higher Income Tax Rates Cut Revenues and Not Just Cyclically

From 1930 to 1937, unlike 1923–25, virtually all federal and state tax rates on incomes and sales were repeatedly increased, and many new taxes were added, such as the Smoot-Hawley tariffs in 1930, taxes on alcoholic beverages in December 1933, and a Social Security payroll tax in 1937. Annual growth of per capita GDP from 1929 to 1939 was essentially zero (Johnston and Williamson 2019).

ECONOMIC IMPACT OF TAX CHANGES

Table 1 shows that from 1930 to 1940, marginal federal income tax rates were increased most sharply at incomes below \$10,000, rather than at incomes higher than \$50,000. Even aside from the vastly larger and regressive excise and payroll taxes, the income tax increases of two progressive presidents (Hoover and Roosevelt) were not particularly “progressive.” A low income of \$2,000 (after exemptions) was only in a 1.5 percent tax bracket in 1930, 4 percent in 1932, and 13 percent in 1940. For those with net incomes below about \$25,000, marginal tax rates at least quadrupled or quintupled between 1930 and 1940. Tax rates on added income above a million dollars did little more than triple (from 25 percent in 1930–31 to 79 percent in 1936, and 83 percent in 1940).

When personal income tax rates were increased in 1932, 1936, and 1940, the tax base was also increased to raise revenue from middle-income taxpayers. Personal exemptions for married couples were cut from \$3,500 in 1930 to \$2,500 in 1932, \$2,000 in 1936, and \$1,500 in 1940. With rising tax rates at all income levels and repeated cuts in personal exemptions, annual individual income tax receipts were estimated to have increased by hundreds of millions from 1932 to 1940, as

TABLE 1
MARGINAL FEDERAL INCOME TAX RATES
BY SELECT INCOME GROUP

Net Income	1930	1932	1936	1940	Increase 1930–40
\$2,000	1.5%	4.0%	4.0%	13.0%	767.0%
\$4,000	3.0	8.0	8.0	17.0	467.0
\$8,000	5.0	9.0	9.0	25.0	400.0
\$10,000	6.0	10.0	10.0	29.0	383.0
\$20,000	10.0	16.0	16.0	45.0	350.0
\$32,000	14.0	23.0	23.0	54.0	286.0
\$40,000	16.0	26.0	26.0	57.0	256.0
\$52,000	19.0	32.0	35.0	61.0	221.0
\$60,000	21.0	36.0	39.0	63.0	200.0
\$80,000	23.0	46.0	55.0	65.0	183.0
\$100,000	24.0	56.0	62.0	69.0	188.0

SOURCE: Tax-brackets.org.

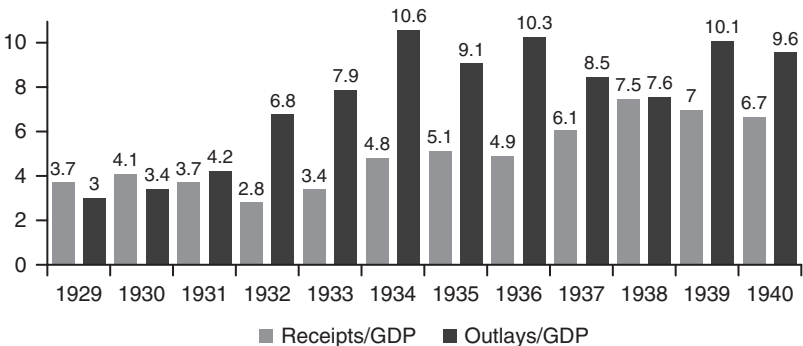
seen in the first table in Romer and Romer (2014: 245). But such rosy revenue estimates were continually disappointed, largely because of the sustained disappearance of high taxable incomes seen in Figure 1.

While tax rates were still low and exemptions high in 1929, the personal income tax brought in \$1.18 billion, and it still raised \$1 billion in 1930 when real GDP fell 8.5 percent (BEA 2020). When tax rates more than doubled in 1932 and exemptions were cut by \$1,000, personal tax revenues dropped 46 percent—from \$532 million in 1931 to \$285 million in 1932 (BEA 2020). Despite a huge array of new and higher excise taxes—nearly tripling excise tax revenues in two years from \$489 million in 1931 to \$633 million in 1932 and \$1,220 million in 1933 (BEA 2020: Table 3.2)—overall federal receipts nonetheless fell even faster than GDP—from 3.7 percent of GDP in 1931 to 2.8 percent in 1932 and 3.4 percent in 1933, as shown in Figure 2.

Higher personal income tax rates remained a symbolic political sideshow in terms of revenue in the Roosevelt years—raising just 1 percent of GDP from 1934 to 1940. Even in 1941, when the class tax began to be transformed into a mass tax with higher tax rates at lower incomes and reduced exemptions, individual income taxes still raised only 1.1 percent of GDP. But regressive excise and payroll taxes in 1941 amounted to 4 percent GNP (OMB 2020: Table 2.3).

Goolsbee (1999) investigates only the impact of higher noncorporate income tax rates on higher incomes in 1932 and 1936. He compares changes in incomes among three high-income groups over four-year

FIGURE 2
FEDERAL OUTLAYS AND RECEIPTS
(PERCENT OF GDP)



SOURCES: OMB (2020: Table 1.2) and St. Louis Fed (for 1929).

spans that neither begin nor end during a year in which tax rates were changed. Like Romer and Romer's long-term averages, this too leaves it unclear about what happened when tax changes took effect.

To estimate the effect of higher tax rates in 1932, for example, Goolsbee compares changes in the mean income of three groups of high-income taxpayers in 1931 with a Pareto-adjusted estimate of such incomes in 1935. There were apt to be changes in names attached to these incomes, however, since GDP fell sharply in 1930–31 and surged sharply in 1934–35. As Goolsbee (1999: 27) explains, "I purposely avoid using the base year as 1929 or 1930, because the output drops were much more dramatic in those years." However, output did not drop in 1929. Moreover, there were also significant changes in 1934 tax rates on income and capital gains, so Goolsbee's 1931–35 interval embraces two tax changes rather than one.

Goolsbee estimated ETI for the top two of the three groups as 0.24 and 0.31, due to negligible differences between those earning \$25,000 to \$50,000 and those earning \$50,000 to \$100,000. From the ETI estimates of 0.24 and 0.31, he deduces that the revenue-maximizing tax rate would be 83 percent if the tax system at that time had only one tax rate. That presumably means an 83 percent rate on all income above \$25,000, rather than on all taxable income, but it clearly implies that Hoover's 1932 tax increase could have maximized revenue if only marginal tax rates on those earning more than \$25,000 had been 83 percent, which would have been almost double their actual (unweighted) average marginal rate of 47 percent.

Using such questionable ETI estimates to suggest the greatly increased 1932 marginal tax rates at high incomes were in any sense revenue maximizing, as both Goolsbee and Romer and Romer do, seems hard to reconcile with the deep, sustained drop in both income tax revenue and reported high incomes from 1932 to 1935.

Figure 2 shows that President Herbert Hoover increased federal outlays from 3 percent of GDP in 1929 to 8 percent in 1933.¹ As a

¹President Hoover's budget for the fiscal year (FY) 1933 was submitted to Congress on December 7, 1931. That fiscal year began June 1, 1932, and ended June 30, 1933. Higher individual tax rates starting with the 1932 calendar tax year also applied to the 1932 fiscal year. Tax liabilities affected by the higher tax rates were first due when tax returns were filed in April 1933, which was within the 1932 fiscal year. The budget for FY 1934 was also President Hoover's plan, submitted December 5, 1932. Figure 3 is based on NIPA data from BEA, which are for calendar (and tax) years rather than fiscal years.

share of GDP, Hoover increased domestic spending more than any president since 1902 except Nixon, according to Steuerle and Mermin (1997), and that was only partly because GDP fell in 1930–33.

President Hoover's soaring spending and falling revenues raised the deficit from 0.5 percent of GDP in 1931 to 4 percent in 1932 and 4.5 percent in 1933. Since Keynesian convention defines "tax increases" by the amount of revenue raised, that approach cannot fathom how the counterproductive 1932 increases in tax rates could possibly have worsened the Depression, since revenues fell. By this standard, however, Hoover's massive income tax rate increases could be reclassified as a *massive tax cut* (and "fiscal stimulus") because income tax revenues *fell*.

In 1936, another increase in the highest tax rates was minor compared with an increase in dividend taxes at all incomes. Together, those tax hikes did appear to raise income tax revenues temporarily to \$732 million in 1936 and briefly to \$1.2–\$1.3 billion in the 1937–38 recession. Several unique reasons for the procyclical tax increase in 1937–38 will be discussed later. After the 1937–38 recession, however, personal income tax receipts dropped back to \$855 million by 1939 and barely reached \$1.0 billion in 1940—slightly less than in 1930 and \$154 million less than in 1929 (BEA 2020).

As a share of GDP, individual income taxes amounted to 1.1 percent of GDP in both 1929 and 1930, when tax rates were 1.1 percent to 25 percent, but then fell to an average of 0.7 percent of GDP from 1932 to 1935, when tax rates were 4 percent to 63 percent. Individual income tax receipts were still no more than 1.0 percent of GDP by 1939–40 when tax rates were 4 percent to 79 percent and personal exemptions had been slashed twice in 1932 and 1936 (BEA 2020).

To summarize: all the repeated increases in tax rates and reductions of exemptions enacted by presidents Hoover and Roosevelt in 1932–36 did not even manage to keep individual income tax collections as high in 1939–40 (in dollars or as a percent of GDP) as they had been in 1929–30. The experience of 1930 to 1940 decisively repudiated any pretense that doubling or tripling marginal tax rates on a much broader base proved to be a revenue-maximizing plan.

The most effective and sustained changes in personal taxes after 1931 were not the symbolic attempts to "soak the rich," but rather the changes deliberately designed to convert the income tax *from a class tax to a mass tax*. The exemption for married couples was

reduced from \$3,500 to \$2,500 in 1932, \$2,000 in 1940, and \$1,500 in 1941. Making more low incomes taxable quadrupled the number of tax returns from 3.7 million in 1930 to 14.7 million in 1940 (Piketty and Saez 2020: Table A0). The lowest tax rate was also raised from 1.1 percent to 4 percent in 1932, 4.4 percent in 1940, and 10 percent in 1941 (Tax Policy Center 2019a).

The percentage of adult households filing income tax returns rose from 6.4 percent in 1931 to 11.2 percent in 1938 and then jumped to 25.7 percent in 1940 and 45.8 percent in 1941 (Piketty and Saez 2020: Table A0). The income tax rate at an income of \$10,000 rose from 11 percent to 14 percent in 1940, and at an income of \$20,000, it rose from 19 percent to 28 percent. In other words, a rising share of middle-class income became taxable at higher tax rates after 1932, 1936, and 1940. But those with incomes above \$50,000 could not or would not pay as much as they had in the 1920s, so the net result was no sustained increase at all in revenues from the individual income tax from 1929–30 to 1939–40.

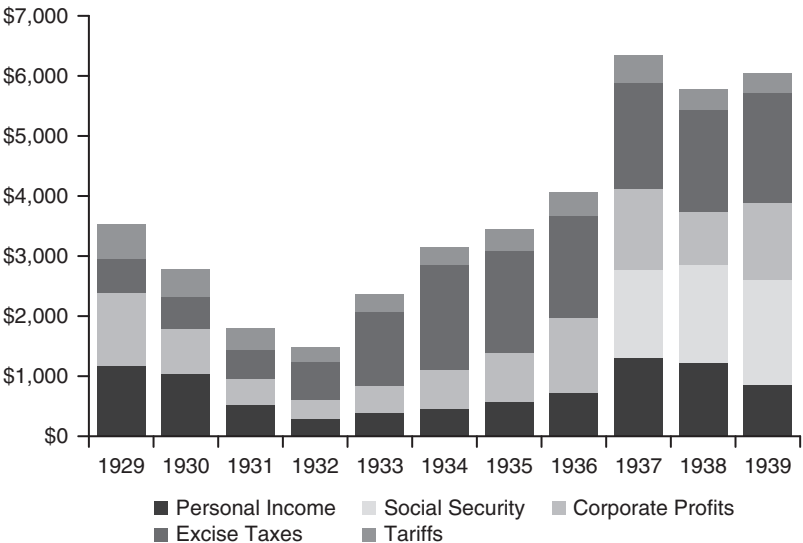
Those earning less than \$50,000 also faced higher tax rates after 1932, but their tax rates were increased only on dividends in 1936 (and another reduction in personal exemptions). Yet any resulting increase in revenues from middle-income taxpayers did not come close to offsetting the loss of income tax revenue from those earning more than \$50,000, with the brief exception of higher taxes due from middle-income taxpayers in April 1937 on 1936 income from veterans' bonuses and from dividends shifted from the (increased) corporate tax by the 1936 undistributed profits tax.

Figure 2 shows that federal receipts from all sources fell from 4 percent of GDP in 1931 to 3.2 percent in 1932, despite much higher excise taxes. By relying on ex ante estimates, however, Romer and Romer (2014: 245) depict revenues increasing \$1,121 million in 1932 by more than 1.9 percent of GDP. If revenue had actually risen by 1.9 percent of GDP, as the Romer and Romer table suggests, then 1932 revenues would have been 5.9 percent of GDP rather than the actual ex post figure of 3.2 percent. In any case, the optimistic revenue estimate is almost irrelevant to Romer and Romer's discussion on personal income tax changes in 1932, because only 16 percent (\$178 million) of the Treasury's additional revenue was estimated to come from higher individual tax rates and lower personal exemptions in 1932 (Thorndike 2003). Most was to come from higher excise taxes, which did rise by \$731 million from 1931 to 1933.

If the ETI at high incomes had been nearly as low after 1932, as Goolsbee and Romer and Romer suggest, the much higher income tax rates at all incomes (and much smaller exemptions) would surely have raised much more individual income tax revenue from all income levels by 1940 than it had in 1929 (\$1.2 billion). In fact, the individual income tax in 1940 raised no more revenue than much lower tax rates and larger exemptions raised during the 1930 slump.

Total federal revenues from all sources did finally rise as a percentage of weakened GDP in the 1937–38 recession, as shown in Figure 2. Yet Figure 3 shows the increases in federal revenues after 1936 were mainly from increased excise taxes and new Social Security payroll taxes. Personal income tax receipts also rose in 1937 because taxes on 1936 income were due when tax returns were filed that April. But, as I will show, the 1936 spurt in taxable incomes was largely due to a \$2 billion payout of taxable benefits to veterans in lower tax brackets and to a one-time payout of dividends made newly taxable at all incomes—rather than to higher marginal tax rates on high incomes.

FIGURE 3
FEDERAL TAX RECEIPTS BY SOURCE
(MILLIONS OF DOLLARS)



SOURCE: BEA (2020: Table 3.2).

1934–1938: Tells Us What Happened in 1934 and 1938, Not 1936

Individual tax rates on dividends and high income were increased in the middle of FY 1936, which began June 1, 1935. Federal receipts only increased from 0.7 percent of GDP in FY 1935 to 0.8 percent in 1936, then 1.0 percent in 1937 and 1.4 percent in 1938, before falling back to 0.9 percent in 1940 (OMB 2020: Table 2.3). The increase in 1938 is somewhat illusory, the result of dividing small gains in revenue by big declines in GDP, because FY 1938 ended June 30, 1938, with four quarters of deep declines in real GDP. Revenue from high-income taxpayers in 1938 was boosted by a much lower 15 percent tax rate on capital gains that year, which encouraged realizations of accumulated capital gains, rather than to higher tax rates on other income. In fact, Figure 1 shows high incomes reported on tax returns falling sharply in 1937 and 1938, ending up lower than in 1935.

What happened when tax rates changed three times from 1934 to 1938 (in 1934, 1936 and 1938) is so complicated that Goolsbee and Romer and Romer do not agree. As Romer and Romer (2014: 266) write, Goolsbee's results are "quite different" from theirs; he "finds an [elasticity] estimate that is large and negative while we obtain one that is large and positive."

Goolsbee uses a relatively rapid increase in high incomes between 1934 and 1938 to approximate their response to higher tax rates in 1936. The trouble is there were three major changes in federal taxes from 1934 to 1938, not just one.

Romer and Romer cannot resolve their 1934–38 differences with Goolsbee, because their elasticities "cannot be estimated with any useful degree of precision (particularly in 1934 and 1938)"—ranging from zero to 3.0 for 1934, zero in 1936, and about minus 2.8 to plus 2.2 in 1938 (Romer and Romer 2014: 266–67). Their problem with 1934 and 1938 is that (unlike Goolsbee) they exclude capital gains from income and thus exclude unprecedented changes in capital gains tax rates in both 1934 and 1938. Goolsbee counts capital gains as income but misinterprets a one-time 1938 surge in high-income gains as a negative taxpayer response to 1936 tax rates on ordinary income rather than a positive response to a new 15 percent tax rate on capital gains.

Until 1933, the capital gains tax was 12.5 percent on assets held more than 2 years. In 1934–37, capital gains on asset sales were

subject to a schedule of steep tax rates, depending on how long an asset was held (Reynolds 2015). The tax rate was reduced from 80 percent to 60 percent of a taxpayer's marginal rate for assets held for 2 to 5 years, then 40 percent after 5 years, and 30 percent after 10. At an income of \$100,000, the marginal tax rate was 56 percent in 1934 and 62 percent in 1936. The capital gains tax on an asset held 2 to 5 years was therefore at least 33.6 percent in 1934 and 37.2 percent in 1936.

Taxpayer decisions to realize capital gains by selling appreciated assets are famously hypersensitive to the largely voluntary tax paid by those who make that decision. In 1933, taxpayers earning more than \$100,000 realized \$97.1 million of long-term capital gains. Statistics of Income (SOI) could not cope with multiple taxes on capital gains from 1934 to 1937, so they lumped short-term and long-term gains together. Total capital gains at incomes above \$100,000 promptly collapsed to \$22.9 million in 1934. After 1936, with even higher capital gains tax rates on high incomes, total short and long-term capital gains fell to \$3.3 million in 1937.

Shortly before June 1938, Congress slashed the capital gains tax on assets sold after two years from 15 percent. In that year, taxpayers with income above \$100,000 realized \$134.8 million in capital gains, up from \$3.3 million the year before. In the year Goolsbee chose to illustrate no response at all of high-income taxpayers to changing marginal tax rates, high-income taxpayers chose to respond quite forcefully to a lower marginal tax rate on capital gains.

The elasticity of capital gains realizations is famously high (about 1.0), because if you rarely sell assets you rarely pay the tax. The collapse of asset sales when the capital gains tax went up in 1934 easily dwarfed later effects of a modest 1936 rise in the slice of ordinary income that exceeded \$50,000. Total IRS income above \$100,000 (including capital gains) rose by \$111 million from 1934 to 1938—from \$419.8 million to \$530.8 million (SOI Stats Archive). However, capital gains alone in that top-income group rose slightly more than total income did—by \$111.9 million—from \$22.9 million in 1934 to \$134.8 million in 1938.

A relatively larger income gain for those reporting incomes of \$100,000 led Goolsbee to claim the ETI was negative in response to higher top tax rates in 1936. That is why he claimed a flat tax of 100 percent in 1936 could have been revenue maximizing. Yet, all of the 1934–38 increase in incomes above \$100,000 happened because

the tax on two-year capital gains was reduced from 38–47 percent in 1937 to 15 percent in 1938—not because of a minor tax rate increase on other sources of income two years earlier.

1936–1938: Big Veterans’ Bonuses, Small Exemptions, and Tax Shifting to Dividends

Despite the 1938 surge of capital gains at incomes above \$100,000, amounts reported by many more returns with taxable income above \$50,000 in Figure 1 only totaled \$1.11 billion in 1935, \$1.89 billion in 1936, \$1.69 billion in 1937, and below \$1 billion in 1938. Taxable high incomes were just not large enough to explain why receipts from the individual income rose to 0.8 percent of GDP in 1936, 1.0 percent in 1937, and 1.4 percent in 1938. Besides, the total of high incomes on tax returns moved in the wrong direction from the ratio of receipts to GDP—falling from 1936 to 1938 rather than rising.

Viewed as a percentage of GDP, federal taxes collected from individual incomes at all incomes appear to rise during the recession of 1937 and 1938 when incomes were falling half the time. This was not because raising the marginal tax rate from 34 percent to 35 percent on \$50,000 incomes in 1936 brought in a ton of revenue. It was because *taxable* income suddenly increased at *all* incomes in 1936, particularly middle incomes. And income earned in 1936 was taxed when returns were filed in April 1937, while most income earned in 1937 was likewise taxed in 1938.

Very little of the added revenue collected in 1937–38 came from those earning more than \$50,000, whose *marginal* tax rates were (slightly) increased. The high-income share of income tax receipts *fell* in 1937 when those receipts peaked.

Aside from some residual monetary stimulus—with nominal GDP rising by 14.3 percent largely because of the 1934 devaluation (Eggertsson 2008; Sumner 2015: chap. 7)—most of the ephemeral spurt in *taxable* income *at all incomes* in 1936–37 was unique and uncomplicated. It happened because of (1) an enormous one-time payment of veterans’ bonuses; (2) shrinking tax exemptions and bracket creep; (3) income shifting from corporate to individual tax forms due to a new corporate tax on earnings not distributed as dividends; and (4) higher tax rates due for the first time from all taxpayers in 1937 on artificially large dividends received in December 1936 from income shifting.

On January 24, 1936, Congress passed, over the president's veto, a World War I veterans' bonus, which paid \$1.76 billion in \$50 bonds to 3.2 million veterans—80 percent of which was redeemed in 1936, particularly from June 15 to July 31 (Hausman 2016: 1105). Veterans' bonuses affected nearly 8 percent of all households and amounted to 2 percent of GDP and at least 30 percent of average income (which was about \$1,500) for those aged 35 to 44 (Tesler 2003). The average bonus was \$581, enough to buy a new car or make a down payment on a house, so sales of consumer durables and homes boomed.

Because the personal exemption was reduced to \$1,000 for singles and \$2,500 for couples in 1936, many veterans' bonuses were *taxable income* for people with incomes (even including the bonus) that were modest. There were 33 tax brackets in 1936, so 3.6 percent inflation that year caused some "bracket creep" by pushing even \$2,000 increases in nominal income into a higher tax bracket.

The Revenue Act of June 22, 1936, also introduced a crucially important graduated surtax of 12–27 percent on undistributed corporate profits. It caused rapid, large "income shifting" from the corporate tax to individual tax by encouraging corporations to increase dividend payouts to escape the tax in 1936–37 (Haas 1937). Income shifting, in turn, contributed to low elasticity estimates for 1936 from Romer and Romer and Goolsbee, because it briefly boosted the share of corporate income reported on high-income shareholder individual tax returns (rather than on corporate returns). Part of what would otherwise have been counted as retained corporate income was briefly reported and taxed as increased dividends at the individual level in 1936 when dividend tax rates at all income levels were substantially increased. Piketty and Saez (2020: Table A7) find that as a percentage of reported taxable income of the top 10 percent (virtually all taxpayers), dividends rose from 12.5 percent in 1934–35 to 15.7 percent in 1937–38, then fell to 11.5 percent in 1938 when tax rates on retained earnings and capital gains fell.

Before 1936, most taxpayers facing low tax rates of 4–8 percent were exempt from any tax on dividends, and others subtracted the 8 percent rate from their rates. The top rate of 63 percent was 55 percent for dividends. The surge of dividends, distributed to individuals in 1936 (to avoid the new corporate tax on retained earnings), were newly subjected to combined basic and surtax rates of up to 79 percent. What is more important for explaining the 1936–37 surge in

taxable income, dividends became widely taxable for the first time even for those in the lowest tax brackets.

Federal revenue from all sources were only 4.6 percent of GDP in 1936 (similar to 1930), but then rose to 5.8 percent in 1937 and 7.7 percent in 1938 (Figure 2). Taxes on incomes earned in 1936 were not due until April 1937. But even in 1937, the peak individual income tax of \$1.3 billion in 1937 was still only 26 percent of federal revenue—smaller than the new Social Security tax (\$1.47 billion) or excise taxes (\$1.77 billion).

An ephemeral 12.8 percent spurt in real GDP in 1936 led by consumer spending (of bonus checks) was promptly followed by deep recession from May 1937 to June 1938, cutting GDP by 10 percent and industrial production by 32 percent, with unemployment rising to 20 percent.

At the time of what critics dubbed the “Roosevelt Recession” economists like Joseph Schumpeter and industrialists like Lamont du Pont “severely criticized . . . the undistributed profits taxes and capital gains taxes” (Roose 1954: 209–16). The Democratic Congress feared voter backlash that fall (when they did lose 72 seats in the House and 7 in the Senate), so on May 28, 1938, Congress greatly reduced or repealed these unpopular taxes and the economy rebounded.

The undistributed profits tax went into force in June 1936, but its effects on corporate profits and liquidity were not apparent, Schumpeter noted, until the first quarter of 1937. Corporations avoided paying the tax by a huge dividend payout in December 1936, which briefly added to personal income in 1936 before adding to taxes due on April 1937. But retaining no earnings left firms precariously short of liquidity and dependent on external funds for investment, which was difficult or impossible for small firms and costly for large ones.

High taxes on capital gains discouraged risky ventures, du Pont argued and Roose agreed. But because he mistakenly thought the capital gains tax rates in 1936–37 were the same as in 1934 (they rose at incomes above \$50,000), Roose concludes that “only the social security and undistributed profits taxes appears to have had a direct influence on the timing and occurrence of the recession” (Roose 1954: 215).

The Revenue Act of May 28, 1938, passed without the president’s signature. That law slashed tax on undistributed profits to 2.5 percent

in 1938 and abolished it in 1939. The top tax on capital gains was cut to only 15 percent on assets held two years. The higher discounted present value of future after-tax earnings was quickly capitalized as the Dow Jones stock index rose from 107.9 the day before the tax cut to 130.5 a month later and 154.8 by the end of the year (Johnston and Williamson 2019).

With confidence and wealth on the rise, the recession ended in June—a few days after the 1938 tax cuts. Yet Romer and Romer (2014), among others, argue that massive increases in numerous income and excise taxes in 1932–36 had no ill effects on the U.S. economy.

Were Higher Taxes Contractionary after 1945, Yet Harmless in 1932–1937?

In a renowned 2010 paper about 1945–2007, Romer and Romer (2010: 763, 781) concluded that “tax increases are highly contractionary,” and “have a very large, sustained, and highly significant negative impact on output.” They also found the prolonged “persistence of the effects [on investment] is suggestive of supply effects” on incentives (*ibid.*: 799).

By contrast, the 2014 Romer and Romer estimates of a low ETI in prewar years, and their related high estimates of revenue-maximizing income tax rates, may appear to demonstrate that higher tax rates were not harmful to the weak and unstable economy of 1932–38. But that would be an unduly broad conclusion to base on such narrow evidence. The Romer and Romer ETI estimates deal exclusively with only one relatively small federal tax (the personal income tax accounted for only 15.3 percent of federal revenue in 1933) and on too small a number of taxpayers (about 25,000) to capture the impact of all the increased taxes on capital, labor, and sales enacted in 1932 and 1936.

Romer and Romer (2014: 278) nonetheless examine connections between increases in top marginal income tax rates and select measures of short-term business activity. They find “no evidence of an important effect of changes in marginal income tax rates on machinery investment and industrial construction.” Importantly, they assume no “effects on long-run economic performance” and, instead, focus on “immediate” or “temporary” effects on two minor indicators of business investment.

In 1932, real investment in business equipment fell 41.1 percent, structures by 38.8 percent, and housing by 43.5 percent (BEA 2020). Growth theory and real business cycle theory might suggest that collapse of fixed investment was at least aggravated by greatly increased 1932 corporate and personal tax rates on expected future returns from investment. Yet Romer and Romer (2014: 278) see “no evidence that the large swings in the after-tax share in the interwar era had a significant impact on investment in new machinery or commercial and industrial construction.” Indeed, they report that, “Machinery production . . . changed little immediately after the very large 1932 tax increase” (ibid.: 273). Also, “commercial and industrial construction contracts . . . changed little following the 1932 tax increase, and rose temporarily after the 1935 tax increase”—which actually took effect in 1936.

Rather than concentrating on how machinery *production* fell (after orders declined), a more forward-looking indicator of a tax shock would be manufacturers’ new *orders* for such durable goods, just as construction contracts are a leading indicator of actual construction (which does not usually stop until projects are completed). The forward-looking NBER index for manufacturers’ new orders for durable goods fell to 34 in 1932 from 60 in the previous year.² And regardless what may have happened *temporarily* to construction contracts, nonresidential construction spending was cut in half in 1932—from over \$1 billion in 1931 to \$502 million in 1932; it fell further to \$406 million in 1934 (BLS 1953: 9).

Romer and Romer (2014: 275) trend-adjust some of their investment proxies “to help address the fact that there were enormous macroeconomic fluctuations in this era.” They make these cyclical adjustments by estimating “how investment behaves in the wake of changes in tax rates given the path of overall economic activity following the changes.” They claim such cyclical adjustment of investment data is “reasonable if the effects of the tax changes on the overall economy are small . . . in light of their small impact on aggregate demand.” That alleged “small impact on aggregate demand” implies that because Hoover’s 1932 tax increase did not raise revenue or reduce the deficit (the opposite in fact), it must have been a fiscal stimulus rather than contractionary within a vintage Keynesian narrative linking budget deficits to nominal GDP growth.

² Available at <https://alfred.stlouisfed.org/series?seid=M0684AUSM343SNBR>.

Even within that narrative, however, Figure 3 shows that increases in total federal tax revenues in the 1930s were dominated by payroll and excise taxes. These broad-based taxes on what people earn and spend cannot easily be dismissed as having had a “small impact on aggregate demand.” Cole and Ohanian (1999: 6) found “consumption fell about 25 percent below trend and remained near that level for the rest of the decade.” This should not be surprising at a time when federal excise taxes alone were at a peacetime record high and many states added or increased sales taxes.

When Romer and Romer (2014) judge the effect of changed income taxes (rather than excise and payroll taxes) by their “fairly small [and sometimes inverted] effects on revenues,” they refer to and depend upon estimated changes in revenues—“statements about the expected revenue effects of a tax change at the time it was scheduled to go into effect.” They inform readers that the 1932 tax law was estimated to raise revenues by 1.9 percent of GDP as though estimated revenue was meaningful, even though actual revenues fell by 0.8 percent of GDP that year.

Romer and Romer speak of “large swings in output” and “enormous macroeconomic fluctuations” in the 1930s as if two depressions in eight years were entirely due to Fed policy or bad luck and not at all to destructive tax, trade, or regulatory policy. “It is difficult to separate the effects of tax changes from the large cyclical movements in investment,” Romer and Romer (2014: 243) acknowledge. Yet to dismiss huge movements in investment as merely cyclical—or to obscure such movements with trend adjustments—tends to minimize possibly damaging effects of tax changes on business and household investment. Moreover, to focus solely on questionable indexes of short-term investment and construction also neglects many other disincentive effects of higher tax rates on productive activity. Akcigit and Stantcheva (2020: 6), for example, find “that both corporate and personal income taxes negatively affect the quantity and quality of innovation at the macro state level and the individual inventor and firm levels.”

Tax Increases? What Tax Increases?

In marked contrast to what Stein, Burns, and Brown once said about Herbert Hoover’s suicidal tax increases in 1932, Cole and Ohanian (1999: 12) recently wrote that “tax rates on both labor and

capital changed very little in 1929–33, which implies that they were not important for the decline.” Yet marginal personal tax rates on both labor and capital (e.g., small business and partnership income, dividends, interest, rent, and capital gains) were greatly increased in 1932 on everyone who was not exempt and exemptions were reduced. Marginal excise tax rates were greatly increased in 1932 and 1936. The effective corporate tax rate was increased from 11 percent in 1929 to 13.75 percent in 1932, 17.6 percent in 1937, and 19 percent in 1938 (Seater 1982: 363).

The 1932 tax law involved raising the top tax rate from 25 percent to 63 percent, quadrupling the lowest rate from 1.1 percent to 4 percent, and increasing corporate and estate tax rates. It is important to realize that federal excise and state sales tax rates also affect marginal rates on factor incomes, and both rose in the 1930s. Hoover’s 1932 tax was mainly about broader and higher excise taxes on goods and services. Those increased excise taxes (as well as the infamous tariffs of 1930) ultimately fell on factor incomes. Taxes on the uses of income, like taxes on the sources of income, reduce after-tax rewards to people who supply the complementary factors of labor and capital.

To focus exclusively on marginal tax rates on only personal income and at only the federal level, clearly understates the increases in marginal tax rates on income and sales at the federal, state, and local levels. In an understatement, Romer and Romer (2012: 15–16) briefly mentioned in their background notes that “the majority of the [estimated] revenue effects were due to a vast increase in excise taxes. The most notable of these was an across-the-board tax of 2¼ percent on all manufactured articles.”

The revenue-maximizing top tax rate formula that Romer and Romer borrow from Diamond and Saez (2011: 168) was never intended to be a top rate for the federal income tax alone as they imply, but to include all taxes on income, payrolls, and sales at the federal, state, and local levels. Even for the narrow purpose of estimating revenue-maximizing tax rates, it is a serious omission for Romer and Romer and Goolsbee to ignore federal and state excise and sales taxes, as well as state income and sales taxes. In fact, to focus only on the economic impact of the federal tax on individual incomes in 1932–40 is to leave out 92.5 percent of all federal, state, and local tax receipts in those years (BEA 2020: Tables 3.4 and 3.5).

In 1932, federal income tax rates were roughly doubled at all income levels, with larger increases at the lowest and highest incomes.

In 1934–36, there were additional increases in income taxes on capital gains and dividends, and marginal tax rates were again increased on net incomes of \$50,000 or more. In 1929, the federal income tax on individuals accounted for 50.9 percent of all federal taxes and 14.1 percent of all federal, state, and local taxes. From 1932 to 1940, by contrast, the federal income tax accounted for an average of 25.8 percent of all federal taxes and only 7.5 percent of total federal, state, and local taxes (BEA 2020: Tables 3.4 and 3.5). If the objective of high marginal tax rates on personal income from labor and capital was to maximize revenue, regardless of any adverse impact on growth of the economy (and therefore growth of taxable income), the Hoover and Roosevelt plans to fix the Depression with higher taxes in 1932–36 were a total failure.

To point out that high and rising federal income tax rates in 1932–36 reduced revenue from \$1,179 million in 1929 to \$855 million in 1939, certainly does not imply that steep marginal tax rates were small in terms of distortions, disincentives, or tax avoidance. On the contrary, the decade-long inverse relationship between tax rates and revenue suggests high tax rates were extremely inefficient, causing large deadweight costs with no lasting revenue gain. The greatly increased tax rates on realized capital gains from 1934 to mid-1938 are a good example. By discouraging asset trading, high tax rates on realized gains (47 percent after two years) reduced realizations and tax receipts. Falling tax receipts, in turn, makes those higher tax rates look unimportant in terms of static bookkeeping and Keynesian macroeconomics. Yet falling revenue from higher tax rates illustrates the high cost of punitive marginal tax rates with no offsetting benefits.

Despite large increases in many such distortive federal and state marginal tax rates on what people earn and on how they spend those earnings, Mulligan (2002) nonetheless relies only the federal tax on individual labor income to agree with Cole and Ohanian that new taxes on labor and capital were trivial and insignificant in 1932–33 (and Mulligan includes 1936–38). He remarks that

Cole and Ohanian (1999) suggest that . . . taxes on factor incomes might help explain some of the Depression economy Of course, taxes on labor income create such a wedge, but Barro and Sahasakul's study suggests that federal taxes on payroll and individual income were trivial, and unchanging. . . . [The] vast majority of the population did not file individual

income tax returns during the 1930s, so that any IRS-induced tax wedge affected very few people (not to mention small effects for the few affected) [Mulligan 2002: 27].

Mulligan emphasizes only the labor tax wedge (without mentioning the 1937 payroll tax) while neglecting marginal tax rates on capital incomes: higher corporate tax rates, surtaxes on undistributed profits, higher individual tax rates on partnerships and small enterprises, higher tax rates on capital gains in 1934–37, on dividends in 1936, and raising the top tax on estates from 20 percent in 1931 to 70 percent 1936.

Wright (1969: 300) estimates the average U.S. marginal tax rate on dividends rose from 11.1 percent in 1928 to 27.1 percent in 1932 and 29.9 percent in 1936, and the marginal tax on interest income rose from 9.2 percent in 1928 to 14.4 percent in 1932 and 17.6 percent in 1936. Federal taxes on gifts and estates rose from \$35 million in 1932 to \$379 million in 1936—nearly half as large as the \$732 million collected that year from individual income taxes (Joulfaian 2007: Table 6).

The only reason both Mulligan and Cole and Ohanian imagine nothing significant happened to tax policy in the 1930s is that both rely entirely on inappropriate or irrelevant estimates of an *income-weighted* “average” of marginal tax rates from the federal individual income tax alone.

To explain why they suppose “tax rates on both labor and capital changed very little 1929–33,” Cole and Ohanian cite a 1981 study by Joines, which shows average marginal rates on labor little changed until 1937 when the Social Security tax was added. Mulligan (2002) cites similar 1983 estimates from Barro and Sahasakul (1983). Yet, Barro and Sahasakul were highly critical of the income tax rate increases of 1932 and 1936, which they explained in detail. They wryly concluded: “Apparently the tax increases between 1932 and 1936 reflect the Hoover-Roosevelt program for fighting the Depression!” (Barro and Sahasakul 1983: 21).

Barro and Sahasakul (*ibid.*: 1) “argue that the explicit rate from the schedule is the right concept for many purposes.” And those rates, they explicitly explained, were greatly increased between 1932 and 1936. Yet the same authors’ income-weighted “average marginal tax rate” *fell* from 4.1 percent in 1928 to 2.9 percent in 1932 and 3.1 percent in 1933. That is why Mulligan (2002: 27) concludes, “Barro and

Sahasakul's study suggests that federal taxes on payroll and individual income were trivial, and unchanging." On the contrary, Barro and Sahasakul (1983: 1) derided higher income tax rates in 1932 and 1936 and apologized for not including payroll taxes (until later). They also apologized for excluding excise taxes. "A full measure of marginal tax rates," they wrote, "would incorporate other levies, some of which are based on property or expenditures" (ibid: 2). Increased federal expenditure (excise) taxes on dozens of good and services directly affected millions more people than the individual income tax, and excise taxes brought in over twice as much revenue (an average of \$1,536 million per year from 1932 to 1939 compared with \$728 million for the income tax).

The seemingly paradoxical decline from 1928 to 1932 in the Barro and Sahasakul "average marginal tax rate" does *not* show that marginal tax rates fell in 1932 or were either trivial or unchanging. What it does show is that the 1928–29 income-weighted average was heavily weighted toward a rising number of *high incomes* above \$50,000 taxed at rates of 18–25 percent, while the 1932–35 average was instead heavily weighted toward many more *low* incomes taxed at 4–10 percent (up from 1.5–5 percent in 1931).

As Seater (1982: 354) explained, an "income-weighted average marginal rate" is computed by "using for weights the fraction of income that fell within each income class." Because very little income still fell into the high-income tax brackets after 1932, the weight of high-tax brackets within the average was diluted. This is an entirely different concept of "average marginal rate" than I used in Figure 1, which is an *unweighted* average of statutory rates.

Income-weighted averages of marginal tax rates in this period are a roundabout way of describing the same phenomenon as Figure 1—namely, that many high incomes dropped out of those averages after 1932. Reported net incomes above \$50,000 collapsed from \$6.3 billion in 1928 to \$800 million in 1932 and never again even hit \$2 billion before WWII. As billions of dollars of high taxable incomes vanished when faced with the high 1932–35 marginal tax rates, all the income-weighted "average marginal tax rates" became dominated by much larger weights then attached to numerous taxpayers with relatively inelastic lower incomes.

Mulligan focuses on the fact that few people paid federal income tax after 1932—especially fewer high-income people—than before. Yet it is also true that relatively few people start new businesses,

employ many people, or supply capital to those who do. As McGrattan (2012: 2) notes, “Although few households paid income taxes, those who did earned almost all of the income distributed by corporations and unincorporated businesses. If the increases in rates were not completely unexpected [Hoover announced them in December 1931], these households would have foreseen large declines in future gross returns on investments . . . even before 1932 when major changes were enacted.”

When it comes to blaming the Depression’s severity and length partly on bad tax policy, McGrattan and others remain somewhat closer to the early postwar consensus of Stein, Burns, and Brown. McGrattan assigns considerable blame for the stubborn severity of the post-1932 depression in general, and 1937–38 recession in particular, on high marginal corporate and individual tax rates on capital. That includes new taxes on retained earnings and higher taxes on dividends in 1936, plus high tax rates on most realized capital gains from 1934 until early 1938.

Calomiris and Hubbard (1995) also find the 1936–37 surtax on undistributed corporate profits was particularly harmful to the working capital and plant and equipment outlays of smaller, rapidly growing firms. Large corporations could avoid the tax by paying more dividends (reported as taxable income on individual income tax returns), but this was not a viable option for smaller growth companies with limited access to external funds and therefore dependent on retained earnings to reinvest for expansion.

Christina Romer (2009) blames the 1937–38 economic collapse partly on the new payroll tax for Social Security, which raised even more revenue that year (\$1,473 million) than the briefly engorged personal income tax (\$1,305 million), although not as much as newly increased excise taxes (\$1,771 million). Cole and Ohanian (1999: 12) estimate that increases in the marginal tax on labor between 1929 and 1939 (notably the payroll tax) reduced the steady-state labor supply by 4 percent.

Mulligan (2002) does acknowledge that excise taxes fall on the income received from supplying labor and capital and could therefore raise marginal tax rates on factor income.³ However, he

³ According to Carl Shoup (1934: 4), the 1932 excise taxes “doubled the [stamp] tax on the issuance of corporation bonds or capital stock . . . and likewise doubled the tax on the transfer of capital stock,” which were clearly taxes on capital.

mistakenly compares today's minor excise taxes (0.4 percent of GDP) with those of 1932–40 (2.1 percent of GDP). He writes, "The federal government did not have a general sales tax, although it does have (and has had) excise taxes on goods such as cigarettes, gasoline, and imports. However, the revenues from these taxes are [today] too few, and not changing enough over time, to drive much a of wedge" (*ibid.*: 25). On the contrary, excise taxes that were added and increased in 1932 and 1936 were neither few in number, small in size, nor modestly changed (see Figure 3).

In 1932, the secretary of the Treasury reported that the Revenue Act of 1932 "effected one of the largest increases in taxes ever imposed by the federal government." The report boasted of "manufacturers' excise taxes on numerous articles" and "other miscellaneous taxes, including new and increased stamp taxes, increased taxes on admissions, and new taxes on telephone, telegraph, cable, and radio messages, checks, leases of safe deposit boxes, transportation of oil by pipe line, and the use of a boat" (U.S. Treasury 1932: 21).

Excise tax receipts in 1933 were \$1.58 billion—up from \$532 million in 1931 and more than four times the \$390 million collected from high personal income taxes. From 1932 to 1939, individual income taxes averaged less than 1 percent of GDP, but excise taxes were 2.1 percent of GDP (OMB 2020: Table 2.3).

Conclusion

In a survey about the theory and evidence of economic growth, Robert Barro and Paul Romer (1990: 3–4) concluded that "all economic improvement can be traced to actions taken by people who respond to economic incentives If government taxes or distortions discourage the activity that generates growth, growth will be slower."

News about abrupt changes in government taxes (or regulatory distortions) may create a policy shock forcing the people to respond to new incentives. Using postwar data, for example, McGrattan (1994: 25) finds "tax rate shocks have a significant effect on the variance of most of the variables in the model" including consumption, output, hours worked, capital stock, and investment. In simulations by Sirbu (2016: 24), changing "news about [future] capital income tax rates . . . generate not only qualitatively but also quantitatively realistic aggregate fluctuations."

Few economic historians who have written about causes of the Great Depression have assigned much importance to tax changes, and many entire books on the subject do not even mention tax policy. In the late 1930s and early postwar decades, however, economists commonly agreed that large increases in personal, corporate, and excise tax rates in 1932 and 1936 contributed to the Depression. This article reviews more recent and very different studies that assert or imply that tax rate shocks in the 1920s and 1930s had no discernible impact on activity that generates growth.

Romer and Romer (2014) and Goolsbee (1999) investigate whether reductions in the highest marginal tax rates during the 1920s encouraged high-income taxpayers to earn and report more income and, conversely, whether higher top tax rates in 1932 and 1934–36 encouraged affluent taxpayers to report less income. Figure 1, by contrast, shows high incomes always moved inversely with marginal tax rates (except in 1936–38) suggesting elasticity of taxable income (ETI) was high.

Romer and Romer conclude otherwise, as does Goolsbee for 1930s. They claim ETI was so low in 1932–38 that top tax rates of 73–84 percent would have been revenue maximizing. Yet raising the top tax rate from 25 percent in 1925–31 to 63–79 percent in later years produced individual income tax revenues no higher in 1940 than they had been in 1930.

Goolsbee and I agree that the ETI was high when such tax rates were greatly reduced in 1923–26. But where I find substantial qualitative response after tax rates rose sharply in 1932, and Romer and Romer find a middling ETI of 0.42, Goolsbee finds a weak taxpayer response (0.27) from 1931 to 1935, which is his proxy for what happened in 1932. He also finds the highest incomes rose 1934 to 1938, which he interprets as zero response to higher tax rates in 1936. But I show that was because capital gains realizations collapsed in 1934 when the capital gains tax was greatly increased then surged in 1938 as that tax was cut to 15 percent.

Romer and Romer's low ETI estimates for tax rate reductions in 1924 and 1926 assume (doubtfully) that taxpayers could not anticipate those preannounced rate cuts would be retroactive. Elasticities for tax changes in 1934–38, they say, "cannot be estimated with any useful degree of precision" (Romer and Romer 2014: 266).

To buttress their ambiguous ETI estimates for 1932–38, Romer and Romer offer a benign interpretation of the 1932 tax law's economic impact, finding few immediate effects on select measures of

business activity and assuming no effect on aggregate demand. That exercise is inherently incomplete because it (1) includes only about 25,000 taxpayers and (2) excludes major changes in excise taxes and in tax rates on dividends in 1936, capital gains in 1934–37, and undistributed corporate profits in 1936–37. I argue that what happened to those widely ignored taxes largely explains the brief anomaly of high incomes and top tax rates moving up together in 1936–38 (e.g., a temporary spike in dividend payouts in 1936 and surge in capital gains in 1938) as well as the simultaneous recession of late 1937 and early 1938 (see McGrattan 1994).

Cole and Ohanian (1999) and Mulligan (2002) do not merely claim higher tax rates were harmless (as Romer and Romer do) but claim there were no significant changes in marginal tax rates on labor or capital in 1932 or 1936. That puzzling claim depends entirely on inapt estimates of income-weighted “average marginal tax rates,” which fell when all statutory tax rates rose. That is because these are averages of tax rates weighted by the number of tax returns in low, middle-, or high-income groups. After top tax rates rose from 25 percent to 63–79 percent in 1932 and 1936, the so-called average marginal tax rates assigned much less weight to the shrinking numbers of taxpayers in higher tax brackets in 1932–39 than they had before. In 1928, 43,184 taxpayers with gross incomes above \$50,000 accounted for 11 percent of all tax returns. In 1935, high-bracket taxpayers accounted for only 10,680 (0.02 percent) of 4.6 million returns (SOI). Far from proving weak response to high tax rates, vanishing high incomes illustrates the opposite (i.e., high ETI).

To focus on the federal tax on individual incomes alone, as all these studies do, misses a much bigger picture. Federal taxes were a small fraction of total government revenues (29.3 percent from 1932 to 1940) and the individual income tax was a small and shrinking fraction of that small federal share. The 1932 and 1936 revival of wartime excise taxes on a huge array of goods and services was enormously damaging to aggregate demand (sales taxes reduce sales) and factor supply (sales taxes are paid from after-tax returns to labor and capital). By 1940, the individual income tax accounted for only 13.6 percent of federal revenue, the corporate tax for 18.3 percent, Social Security tax for 27.3 percent, and excise taxes for 30.2 percent (OMB 2020: Table 2.2).

After presidents Hoover and Roosevelt embraced what Herb Stein dubbed “the desperate folly of raising taxes in a Depression,”

individual income taxes amounted to “1.4 percent of personal income in 1929 and only 1.2 percent in 1939” (Stein 1988: 58). It turned out that, contrary to revenue-maximization simulations of a 74–83 percent top tax rate by Goolsbee or Romer and Romer, a top tax rate of 25 percent on personal income (and 12.5–15 percent on capital gains) proved to be the long-run revenue-maximizing top tax rate of 1918–40.

A previous critical review of estimates of the ETI at high incomes in the postwar years found those estimates far too low and found the authors’ arguments about tax rates being unrelated to economic growth factually flawed (Reynolds 2019). This review of prewar ETI estimates finds them equally understated and also too constricted—dealing with only a tiny fraction of many federal and state taxes on labor and capital. Artificially low ETI estimates for the 1920s and 1930 as well as an income-weighted average of marginal tax rates have both been used to claim that exceptionally large and increasingly regressive tax increases from 1932 to 1937 had nothing to do with two deep recessions in those years. That is too strong a conclusion to be supported by such weak evidence.

References

- Akcigit, U., and Stantcheva, S. (2020) “Taxation and Innovation: What Do We Know?” Becker Friedman Institute, Working Paper No. 2020–70, University of Chicago (May 6).
- Barro, R. J., and Romer, P. M. (1990) “Economic Growth.” *NBER Reporter* (Fall).
- Barro, R. J., and Sahasakul, C. (1983) “Measuring the Average Marginal Tax Rate from the Income Tax.” NBER Working Paper No. 1060 (January).
- BEA, Bureau of Economic Analysis (2020) “GDP and Personal Income” and “Federal Government Current Receipts and Expenditures” Tables 3.2–3.5 (May 28). Interactive data available at <https://apps.bea.gov/itable/index.cfm>.
- BLS, Bureau of Labor Statistics (1953) “Construction during Five Decades, Historical Statistics, 1907–52.” Bulletin No. 1146 (December 15).
- Brown, E. C. (1956) “Fiscal Policy in the ‘Thirties: A Reappraisal.” *American Economic Review* 46 (5): 857–79.
- Burns, A. (1958) *Prosperity without Inflation*. Buffalo, N.Y.: Economica Books.

- Calomiris, C. W., and Hubbard, R. G. (1995) "Internal Finance and Investment: Evidence from the Undistributed Profits Tax of 1936–1937." *Journal of Business* (October): 443–82.
- Cole, H. L., and Ohanian, L. E. (1999) "The Great Depression in the United States from a Neoclassical Perspective." Federal Reserve Bank of Minneapolis *Quarterly Review* 23 (1): 2–24.
- de Rugy, V. (2003) "Tax Rates and Tax Revenue: The Mellon Income Tax Cuts of the 1920s." *Tax & Budget Bulletin* 13 (February). Washington: Cato Institute.
- Diamond, P., and Saez, E. (2011) "The Case for a Progressive Tax: From Basic Research to Policy Recommendations." *Journal of Economic Perspectives* 25 (4): 165–90.
- Eggertsson, G. B. (2008) "Great Expectations and the End of the Depression." *American Economic Review* 98 (4): 1476–516.
- Frenze, C. (1982) "The Mellon and Kennedy Tax Cuts: A Review and Analysis." Joint Economic Committee Staff Study (June 18).
- Goolsbee, A. (1999) "Evidence on the High-Income Laffer Curve from Six Decades of Tax Reform." Brookings Panel on Economic Activity (September 26). Available at <https://faculty.chicagobooth.edu/austan.goolsbee/research/laf.pdf>.
- Gwartney, J. D. (2002) "Supply Side Economics." In D. R. Henderson (ed.), *The Concise Encyclopedia of Economics*. Indianapolis: Liberty Fund.
- Haas, G. (1937) "Rationale of the Undistributed Profits Tax." U.S. Treasury Department Staff Memo (March 17). Available at www.taxhistory.org/Civilization/Documents/UPT/HST8668/hst8668-1.html.
- Hausman J. H. (2016) "Fiscal Policy and Economic Recovery: The Case of the 1936 Veterans' Bonus." *American Economic Review* 106 (4): 1100–43.
- Johnston, L., and Williamson, S. H. (2019) "What Was the U.S. GDP Then?" Available at www.measuringworth.org/usgdp.
- Joulfaian, D. (2007) "The Federal Gift Tax: History, Law, and Economics." U.S. Department of the Treasury, OTA Paper 100 (November).
- McGrattan, E. R. (1994) "The Macroeconomic Effects of Distortionary Taxation." *Journal of Monetary Economics* 33 (3): 573–601.
- (2012) "Capital Taxation during the U.S. Great Depression." *Quarterly Journal of Economics* 127 (3): 1515–50.

- Mulligan, C. B. (2002) “A Dual Method of Empirically Evaluating Dynamic Competitive Equilibrium Models with Distortionary Taxes, Including Applications to the Great Depression and World War II.” NBER Working Paper No. 8775 (February).
- OMB, Office of Management and Budget (2020) “Historical Tables.” *Budget of the United States Government*. Available at www.whitehouse.gov/omb/historical-tables.
- Piketty, T., and Saez, E. (2020) “Income Inequality in the United States, 1913–1998.” Originally published in 2003 in *Quarterly Journal of Economics* 118 (1): 1–39. Tables and Figures Updated for the 2018 tax year in Excel format in 2020 (February). Available at <https://eml.berkeley.edu/~saez/TabFig2018.xls>.
- Piketty, T.; Saez, E.; and Stantcheva, S. (2014) “Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities.” *American Economic Journal: Economic Policy* 6 (1): 230–71.
- Reynolds, A. (2015) “Hillary Parties Like It’s 1938.” *Wall Street Journal* (September 2).
- _____ (2019) “Optimal Top Tax Rates: A Review and Critique.” *Cato Journal* 39 (3): 635–65.
- Romer, C. A. (2009) “The Lessons of 1937.” *The Economist* (June 18).
- Romer, C. A., and Romer, D. H. (2010) “The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks” *American Economic Review* 100 (June): 763–801.
- _____ (2012) “A Narrative Analysis of Interwar Tax Changes,” University of California, Berkeley, Working Paper (February).
- _____ (2014) “The Incentive Effects of Marginal Tax Rates: Evidence from the Interwar Era.” *American Economic Journal: Economic Policy* 6 (3): 242–81.
- Roose, K.D. (1954) *The Economics of Recession and Revival: An Interpretation of 1937–38*. New Haven: Yale University Press.
- Seater, J. (1982) “Marginal Federal Personal and Corporate Income Tax Rates in the U.S., 1909–1975.” *Journal of Monetary Economics* 10 (3): 361–81.
- Shoup, C. (1934) “Excise Taxes” Treasury Department Staff Memo (June 1). Available at www.taxhistory.org/Civilization/Documents/Excise/hst8678.htm.
- Sirbu, A.-I. (2016) “News About Taxes and Expectations-Driven Business Cycles.” *Macroeconomic Dynamics* 23 (4):1–31.
- SOI [Statistics of Income] Tax Stats Archive (various years) Internal Revenue Service. Years 1916–33 available at www.irs.gov

- /statistics/soi-tax-stats-archive-1916-to-1933-statistics-of-income-reports. Years 1934–99 available at www.irs.gov/statistics/soi-tax-stats-archive-1934-to-1953-statistics-of-income-report-part-1.
- Stein, H. (1966) “Pre-Revolutionary Fiscal Policy: The Regime of Herbert Hoover.” *Journal of Law and Economics* 9: 189–223.
- (1988) *Presidential Economics*. Washington: American Enterprise Institute.
- Steuerle, E. C., and Mermin, G. (1997) “The Big Spending Presidents.” *The Future of the Public Sector* No. 11 (April). Washington: The Urban Institute.
- Sumner, S. (2015) *The Midas Paradox*. Oakland, Calif.: Independent Institute.
- Tax Brackets.Org. Tax rates by income can be found by changing the year at the end of this URL: www.tax-brackets.org/federal-taxtable/1923.
- Tax Policy Center (2019a) “Historical Individual Income Tax Parameters” (August 2). Available at www.taxpolicycenter.org/statistics/historical-individual-income-tax-parameters.
- (2019b) Effective Tax Rate by AGI, 1935–2015 (December 24). Available at www.taxpolicycenter.org/statistics/effective-tax-rate-agi-1935-2015.
- Telser, L. G. (2003) “The Veterans’ Bonus of 1936.” *Journal of Post-Keynesian Economics* 26 (2): 227–43.
- Thorndike, J. (2003) “The Republican Roots of New Deal Tax Policy.” *Tax Analysts* (August 23).
- U.S. Treasury (1932) “Revenue Act of 1932.” Available at https://fraser.stlouisfed.org/files/docs/publications/treasar/pages/59359_1930-1934.pdf.
- Wright, C. (1969) “Saving and the Rate of Interest.” In A. C. Harberger and M. J. Bailey (eds.), *The Taxation of Income from Capital*. Washington: Brookings Institution.

BOOK REVIEWS

Why Culture Matters Most

David C. Rose

New York: Oxford University Press, 2019, 190 pp.

What determines whether rich and free societies arise and endure? In *Why Nations Fail*, Daron Acemoglu and James Robinson famously concluded that the quality of a country's political and economic institutions matter most. David Rose disagrees. Successful free-market democracies, he claims, require *high trust* as their essential prerequisite, both between individuals and "with the system." Without high levels of generalized trust, opportunism becomes rife and democracy is overcome with tribalism, undermining the institutions that correlate with prosperity.

To make clear his direction of causation, he concludes: "A high-trust society can easily adopt trust-dependent institutions if there is reason to do so, but the reverse is not true." For example, the rule of law only operates successfully in the long term, he explains, if people trust the court system to be just. In Rose's view then, institutions or policies cannot effectively substitute for trust sustainably, though they can codify or complement it. Collapsing trust, as some fear today, eventually leads to institutional collapse.

If general trustworthiness is essential for prosperity, how is it achieved? Rose's answer is culture. Mass flourishing requires large-group cooperation, meaning overcoming our skepticism of strangers and forgoing opportunities to exploit them. This is achieved by transmitting through generations a moral taste against certain behaviors

that “produce strong and involuntary feelings of guilt upon behaving in an untrustworthy way.” Inculcating children that certain behaviors are wrong, through imitation and teaching, creates an effective constraint against our worst impulses, facilitating a societal environment of general trustworthiness. This lowers the transaction costs of exchange, allowing more trade and good institutions to operate effectively. Hence, *Why Culture Matters Most*.

Yet if this cultural transmission is the foundation for freedom and prosperity, Rose warns it is a fragile one. For it requires each generation to invest in maintaining such moral beliefs. Given the positive externalities, Rose worries we underinvest in them. In fact, he argues the very success of high-trust societies, generating large markets and democracy, can create dynamics that undermine trust. The returns to opportunism, he claims, are higher in a big market where others are trustworthy. Democracy provides opportunities for special interest groups to benefit from regulatory and redistributive favoritism, which can reinvigorate tribalism too. Maintaining a high-trust society therefore becomes ever-more difficult, requiring vigilance and active investment.

In a world in which policy or electing the right leaders is ever debated, Rose’s warning is well taken. When he delineates how political polarization feeds on itself, his framework seems extraordinarily instructive for our current travails. And yet, the book’s highly theoretical approach, and lack of empirical grounding, raises obvious questions. Can the variance in trust or moral tastes taught to children really explain differences in income levels across countries and over time? How does a culture of trust germinate in the first place? And how can one therefore answer the “chicken and egg” question of what comes first, institutions or culture?

Good institutions clearly don’t fall like manna from heaven. Often, they reflect the past embodiment of a polity’s norms, accepted ideals, and customs. That some stand the test of time in certain countries but not other places suggests culture matters. Indeed, among the ex-communist countries, Poland (a relatively high-trust society because of Catholicism, ethnic homogeneity, and a strong national identity) has seemingly performed better than many other states. Mixed records of success in outside imposition of “structural reforms,” constitutions, democratic elections, and more, suggests cultural buy-in matters. In fact, the most persuasive argument against open borders is the unknowable risk of different cultural

norms undermining institutions that currently underpin freedom and prosperity over time.

But Rose doesn't sufficiently explain what caused the cultural norms that developed into general trustworthiness in today's rich countries, such as the United States. He suggests greater farm ownership in the 19th century might explain why America grew richer and more entrepreneurial than Western Europe, for example. Farm ownership, he claims, meant that American parents had stronger economic incentives to teach their worker-children certain ideals—such as “keeping one's word”—which spilt over into a more trustworthy society. Yet if it was farm ownership that sparked high-trust-generating cultural norms, might not other economic policies or institutions create similar incentives today? And might not institutions codifying past norms offer at least a handbrake against rapid cultural change, making them “matter most” today?

Indeed, some evidence suggests that when institutions change rapidly, trust can break down. In a famous 2014 paper, economists tested Berlin residents' willingness to cheat in a simple game for monetary reward. Participants presented passports and ID cards to the researchers, which allowed them to assess their backgrounds. Those from East Germany, which had been under the grip of socialism, were far more likely to cheat than those from the capitalist West. What's more, the “longer individuals were exposed to socialism, the more likely they were to cheat.” Trust can break down quickly under different institutions and policies. Rose's own analysis shows how an expansive state may reduce trust by encouraging free-riding and reducing the return to pro-social behavior. Does a less trustworthy society demand bigger government? Or does bigger government reduce societal trust?

Rose may be correct that free markets work best in generating prosperity when embedded in high-trust environments. But low-trust environments might still be able to benefit and sustain good policies. Indeed, policies that expand markets might positively shape culture. Most economic transactions are akin to a repetitive game, meaning there are costs to deviating and “cheating” on customers or workers. Contra Rose's claim that the incentives to breach trust are higher in a sophisticated economy, sowing the seeds of trust's breakdown, the opposite may be the case: such behavior is very likely to be easier to detect and punish in large, prosperous markets, and the reputational costs more difficult to shake.

Years of engaging in a high-trust, market economy might in fact generate positive spillovers that obviate the need for certain cultural norms. New technologies today are facilitating within-market institutions, such as readily available ratings and review sites, which substitute for trust, or at least reduce the cost of enforcing it. High-trust societies may therefore sometimes be self-reinforcing, not mean reverting (as Rose implies).

A final gripe with Rose's model is that it's difficult to use it to explain differentials in prosperity. Take the drastic rise of incomes from the mid-18th century. Did trustworthiness in England and the Netherlands hit a tipping point that meant larger-scale market transactions became possible? If so, why then? Much more convincing is Deirdre McCloskey's thesis of another cultural phenomenon—that for the first time a widespread acceptance of innovation meant that prosperity-enhancing activity wasn't stamped out by authority. This suggests a certain level of trust may be a necessary but not sufficient condition for mass flourishing on the prosperity frontier. Beyond a certain point, other factors may become more important. Indeed, the United States became vastly more prosperous through the 20th century than many other nations, but it's not clear higher trust can explain the divergence.

Rose deserves credit for rehighlighting the positive dynamics high levels of generalized trust can create, and the dangers of its erosion. Culture is understudied by economists. But his framework clearly presents as many questions as it answers. That means he lands no killer blow to convince that a culture of trustworthiness “matters most” for human flourishing.

Ryan Bourne
Cato Institute

Competition Overdose: How Free Market Mythology Transformed Us from Citizen Kings to Market Servants

Maurice E. Stucke and Ariel Ezrachi

New York: HarperCollins, 2020, 421 pp.

Is competition always the answer to underperforming markets? Maurice E. Stucke, a professor at the University of Tennessee College of Law, and Ariel Ezrachi, the Slaughter and May Professor of Competition Law at the University of Oxford, unequivocally

answer “no.” Introducing competition often unleashes free-market forces to increase consumer options, lower prices, and improve the quality of products and services. However, the authors emphasize a dark side of competition where certain private-market solutions can increase economic inequality, stifle entrepreneurship and innovation, injure consumers both physically and financially, and destroy competitive industry ecosystems.

In Part I (“When Is Competition Toxic?”), Stucke and Ezrachi offer examples of “overdoses” of economic competition and how they are misused, while explaining why they continue to proliferate. For instance, the authors explain how competition harms both university competitors and their intended student and parent beneficiaries. In the case of highly selective/elite universities, such as Harvard, Stanford, and others, competition for students is partially dependent on national rankings found in the annual university and college ranking (based on 15 key measures) of the *U.S. News and World Report*. Such rankings can create a positive feedback loop (i.e., a rise in the annual ranking attracts more “star” students) or a negative feedback loop (i.e., a decline in the annual ranking attracts fewer “star” students).

One metric is the acceptance rate: the lower a university’s acceptance rate, the more selective it is and the better it performs in the rankings. This “toxic competition” has many universities expending financial resources on the *competition* itself rather than on *improving* their educational product or service. But universities can neither deescalate this academic “arms race” (as they fear the consequences of unilaterally opting out and similar competitors filling the gap and increasing their applicant pools), nor can the students (and parents) who are looking for the best “route” to upward social mobility—a prestigious education that will help ensure their future economic survival. Is the latter behavior rational thinking on the part of students and parents? Perhaps, but the authors indicate that many non-Ivy League graduates compare well in median earnings to Ivy and elite schools, as well in the value the institution delivers in improving students’ skills.

Another cited example of a competition overdose is “choice overload.” Many policymakers, economists, and businesspeople believe that the greater the volume of consumer choices, the better. But how much choice do consumers want, and how much should businesses provide? Research studies (including a much-cited 2000 study by

Sheena Iyengar and Mark Lepper) show that when consumers have too many product or service choices, they become overwhelmed with “cognitive costs,” becoming inhibited about making choices. In fact, from the business perspective of generating revenue, this situation can result in the worst-case scenario—consumers not purchasing *any* product. Yet, the authors concede that other research studies have not shown evidence of choice overload or have revealed mixed empirical results.

Why does choice overload occur? One reason is that humans have to think more as the number of options increases. Rational, deliberative thinking is tiring, resulting in brain fatigue. Another reason is that choice overload may lead to choice avoidance, even when any choice is preferable to not choosing. Or, if we do choose, the mental stress involved in considering excessive options and product attributes can reduce the likelihood of our pursuing other, better options. Last, after we finally purchase (after calculating the many tradeoffs posed by the many options), we may experience buyer’s regret because of considering more counterfactuals (“what ifs”) about the options not chosen. Yet choice overload is dependent on so many situational and individual factors that there is no definitive number of options that will prove to be the “trigger” to initiate overload.

So how do businesses react to choice overload? The authors’ answer is threefold. First, they offer decision aids to reduce the mental burden and counterfactuals. An example is free in-flight entertainment guides, which display the offerings under headings, such as Film, Television, and Music, and then under subcategories within each. With the airlines having no incentive to overwhelm us, we expend less energy and our consumer satisfaction in our choice can increase. Second, retailers can offer fewer options per product category, increasing the likelihood of the consumer purchasing a product while increasing purchasing satisfaction. Club stores, like Costco, offer fewer choices in terms of brand, size, and quantity of product, but promise much lower prices. Aldi and Trader Joe’s offer fewer choices too, but primarily their own labels. Third, a retailer can offer the product variety consumers demand, as they cannot unilaterally limit consumer choice without losing profits, so they offer the variety demanded. Amazon is an example of these first and third marketing approaches, attracting consumers with its many choices within its many product categories.

In Part II (“Who Is Pushing the Toxic Competition?”), Stucke and Ezrachi identify four culprits behind competition overdose. In the first instance, *ideologues* have wholeheartedly embraced the mantra that competition is always per se good, fair, or just for the benefit of all in society—in spite of numerous examples where there is suboptimal competition (e.g., “lemon markets” for automobiles), negative externalities (e.g., environmental pollution from manufacturing operations), and public goods (i.e., characterized by the authors as “without an ability to charge for the benefit, market forces may not necessarily provide this public good”). Nevertheless, the antitrust laws, the courts, policymakers, and the business community all demand competition—even if it is just in an abstract sense.

In the second instance, *lobbyists* argue in favor of minimal business regulation and embrace a reductive competition ideology to profit at society’s expense. Stucke and Ezrachi describe this four-step process by which competition ideology is used to deregulate. First, they advocate for trust in the full force of free-market competition. Second, they argue that because the marketplace offers a superior instrument to deliver services, the state’s role in regulating these markets should be limited. Third, they emphatically decry regulation as paternalistic and appeal to our rugged individualism and pride. Fourth, they actively frame the exploitive practice that an industry or business is or plans on embracing as a pro-competitive innovation. As the authors note: “Lobbyists utilize the competition ideology to support crony capitalism, militate against important regulation and safeguards, and ensure that power and money are allowed to subvert democracy.”

In the third instance, *privatization* becomes the elixir for inefficient, low-quality public services. In this case, Stucke and Ezrachi use the example of the proliferation of privately operated prisons in America. According to the U.S. Bureau of Justice Statistics, in 2019, privately operated facilities held 7 percent of state prisoners and 16 percent of federal prisoners. Studies have shown that because these prison enterprises are for-profit ventures, a high occupancy rate is the key to their corporate earnings. Moreover, it is a common practice for private prison corporations to “cream skim” the low-cost, profitable (“healthy”) inmates and offload the more expensive maintenance (“unhealthy”) inmates to state correctional institutions. The authors conclude that there is very little competition in today’s private prison market. According to a 2016 Brookings

Institution study: “Based on available prison facility information, we calculate that the two largest private prison companies account for around 55 percent and 30 percent of all private prison beds, respectively, and the three largest firms provide over 96 percent of the total number of private prison beds.” Moreover, as the U.S. Bureau of Justice Statistics reported in 2016, there is no evidence that they actually save taxpayers any money. The cost savings promised by private prisons have simply not materialized.

In the fourth instance, Stucke and Ezrachi accuse “gamemakers” of being responsible for creating a toxic competitive dynamic that exploits the participants while primarily benefiting the creator. Facebook and Google are not just “providers of helpful and enjoyable products.” Their vast wealth, power, and sophistication elevate them to the level of gamemakers—architects of their own competitive environments. With consumer personal data the bounty for their advertisers (and the source of most of their revenue), the gamemakers, say the authors, are helping app developers on their advertising platforms to create increasingly addictive products for all age ranges, benefiting them with even more financially lucrative personal data provided by consumers. Lastly, for the consumer, it is nearly impossible for a consumer to “opt-out” of data harvesting if they are using the gamemaker products.

In Part III (“What Can We Do about It?”), Stucke and Ezrachi argue that there is no clear definition of “competition.” In antitrust policy, the law generally recognizes the economists’ understanding of the term; yet this definition often differs “from that of lawyers, policymakers, and laypersons.” The authors take aim at Chicago School law-and-economics theorists who characterize “economic competition as relentless zero-sum warfare, where some must lose for others to win.” In contrast, Stucke and Ezrachi propose a non-zero-sum alternative, based on emphasizing trust, collaboration, and fairness, that “can be positive sum—expanding the pie by developing new products, designs, and technologies that will satisfy needs currently unmet by those rivals.”

Stucke and Ezrachi identify four types of competition. First, there is *toxic competition*, which “clearly violates the criteria that require competition to promote overall well-being and abide by ethical standards.” Second is *zero-sum competition*, requiring one’s gain to be offset by another’s loss. While sometimes necessary in practice due to allocation of limited resources, it “may not always serve us or

society, and it often requires us to disregard or outright violate our standards of fairness and morality.” Third is *positive-sum, ethical competition*, “a way of expanding the pie, so that most, if not everyone, benefit.” Moreover, this typology recognizes how ethics and morals can complement, and beneficially inform, the competitive process. Fourth is *noble competition*, “helping your rivals reach their full potential” through a “deep societal and moral awareness.” Noble competition involves each participant in a competition, while seeking to prevail in the marketplace, to be aware of a wider community and recognize how this competitiveness can help its rivals be their best selves. Ultimately, this lifts the level of competition and results in superior benefits for the society as a whole.

The authors take a tripartite approach in promoting aspirational noble competition. First, the state should play two roles in protecting its citizens from toxic competition: initially by employing a variety of legal strategies and then by actively providing what competition cannot deliver. Second, industries need to ensure healthy competition through self-regulation and to actively lobby for rules that will bring companies into compliance with those efforts. Third, consumers can dissent—both at the ballot box to elect candidates who support healthy competition and through consumer purchasing power—and support the type of competition they want to promote. In conclusion, Stucke and Ezrachi argue that “once we view competition as positive sum, we can support markets in which ethical trading and social awareness are celebrated alongside the profit motive.”

While the authors repeatedly discuss the virtues of competition (and their support of it), there is insufficient evidence of why it is their paradigm of choice. In this case, the reigning antitrust paradigm is the Chicago-based, utilitarian approach, which they allege supports (too often) a market “race to the bottom.” While in some cases there is sufficient evidence to support specific antitrust or regulatory intervention, such as the growing list of abusive anticompetitive behaviors associated with Big Tech, their arguments go well beyond “competition” in the narrow sense of antitrust policy.

Stucke and Ezrachi report that support for private prison facilities is on the upswing under the Trump administration. Yet the data show contrary indications. In 2016 (the last year of the Obama administration), *Statista* reports 128,323 prisoners were residing in private prisons in the United States. After the first two years of the Trump administration, the number of prisoners residing in private U.S.

prisons had declined to 118,444, a decline of 7.7 percent nationwide. These most recent private prison capacity utilization numbers certainly do not represent positive news for the corporate earnings of GEO and Core Civic, the two largest for-profit prison companies in the United States, as they are having a negative impact on the private prison industry's need for a high prisoner occupancy rate.

Furthermore, The First Step Act, passed by Congress in December 2018 and strongly supported by the Trump administration, gives federal judges greater leeway when sentencing some drug offenders and boosts prisoner rehabilitation efforts. It also would reduce life sentences for some drug offenders with three convictions, or "three strikes," to 25 years. Another provision would allow about 2,600 federal prisoners sentenced for crack-cocaine offenses before August 2010 the opportunity to petition for a reduced penalty. Yet nowhere in the authors' writings was there any indication of this recent downward trend in privately operated prison occupancy rates, nor the federal prison sentencing reform that would have further potential negative impact on these rates.

A weakness in their approach is that it needs to be offset with the many examples where existing competition is working to benefit consumers and society or how a specific regulatory intervention will improve the outcome for consumers and society. For example, under the Trump administration's deregulatory and tax reduction agenda, the U.S. Census Bureau reported median or average-income family saw a pre-Covid-19 gain of nearly \$5,000 (or 8.2 percent), as median family income in August 2019 was \$65,976, up from about \$61,000 when he entered office in January 2017. This is a notable empirical contrast to the authors' use of data explaining income inequality for middle-income Americans, which reveals that the middle class (40th to 60th percentile) saw their net worth decline 7 percent to \$68,839 between 2000 and 2011.

In addition, the concept of "government failure" did not appear until late in the book. Government failure, in the context of public policy, is an economic inefficiency caused by a government intervention if the inefficiency would not exist in a freely operating market. While the authors acknowledge that government has failed to regulate responsibly in many of their examples (and they offer a couple of specific antitrust exemptions to prevent the race to the bottom, such as when the U.S. government allowed hockey players to agree among themselves to wear helmets), their remedies also represent a

progressive agenda of policy initiatives representing further economic expansion of the social “safety net”—such as legislating an increase in the minimum wage—which may also represent (in certain instances) a continuation of ongoing government failure.

Nevertheless, the authors are correct in their assessment of business, government, and consumers needing to focus on working to realistically attain a positive-sum, ethical form of competition. Their concept of “noble competition” seems a bridge too far and may likely resemble that of unattainable “perfect competition.” Reinforcing *ethical* competition is as much a necessity for a vibrant, innovative capitalist economy (as Milton Friedman recognized) as actively enforcing existing laws and regulations that prevent unfair competition.

Thomas A. Hemphill
University of Michigan-Flint

The Rise of Carry: The Dangerous Consequences of Volatility Suppression and the New Financial Order of Decaying Growth and Recurring Crisis

Tim Lee, Jamie Lee, and Kevin Coldiron

New York: McGraw-Hill Education, 2019, 240 pages

Over the past quarter-century or so, the United States and the world have experienced a relatively new and puzzling, yet old and well-understood, phenomenon. Generally, accommodative monetary policies by major central banks have been punctuated by financial crises large and small rather than by spasms of inflation. The era has been characterized by declining productivity, growth and overall economic vibrancy in spite of an explosion of new ideas and technologies. *The Rise of Carry* looks at the financial mechanisms underpinning these maladies from a new and unusual angle. It identifies at their core an approach to trading heavily reliant on leverage, the use of borrowed funds. Public policy, particularly monetary, encourages leverage and enables it to become excessive, undermining financial stability and ultimately leading to a decline in financial stability and economic dynamism.

The book’s new angle is from the point of view of the carry trade, a trading strategy that serves as a technical description and a metaphor for the incentive system built by the policy prescriptions that keep the financial system leveraged and fragile. The term “carry

trade” was originally applied to a set of foreign exchange transactions in which a sum is obtained, largely by borrowing, in a currency with low money market interest rates, and exchanged for one in which rates are materially higher. If the purchased higher-rate currency appreciates, or at least doesn’t depreciate too much and too fast against the borrowed currency, the trader stands to net some interest income. The difference between the cost of funds and the interest on the purchased currency is called the “carry,” and capturing that gap is the primary aim of the trade. Credit expansion is fostered in the countries issuing both the funding and purchased currencies.

Three features are emblematic of this trading strategy. First, if the carry-trade transactions are not hedging an existing offsetting foreign exchange exposure, then it is an outright bet the purchased currency will retain or even gain in value. Typically, that purchased currency is that of an emerging-markets or export-dependent economy susceptible to infrequent but sharp and sudden depreciations. Second, it can be executed primarily with money borrowed in the low-rate currency from a bank or dealer. At the extreme—to which the reality at times comes remarkably close—in which the entire sum borrowed is owed to a bank or dealer, and the trader has invested no funds of her own, the trade is infinitely leveraged. If it goes well, her rate of return on equity is infinite, and if it’s a loser, she can walk away with at most reputational damage.

Finally, like any other credit transaction, it is in essence also a financial option trading strategy, with several possible framings for analysis. It can be viewed as a long, or purchased, option position, with no inherent limit on the profit of the trader if the trade goes well. Her potential loss is capped at the cost of the option, equal to the amount of her own equity in the trade. The lender, the bank or dealer financing the trade, is in the position of the option seller, with limited profit but exposed to the loss of the entire sum lent.

But the authors set out another, more refined framing of the option-like character of the carry trade. The inherent instability of the carry regime can be most clearly seen through its close kinship with market making in options. As the authors explain, option values can be roughly replicated by “trading with the market,” buying the underlying asset when its price rises and selling when its price falls. Option values are therefore higher when asset prices are moving around a lot and this pattern of trading with the market becomes more costly, that is, when volatility is high.

Selling options and protecting one's position by replicating the options in the underlying asset market is generally profitable, because the demand by many investors for protection from market fluctuations keeps the expected or implied volatility embedded in the options' market prices higher than the actual volatility driving the cost of replication. The market maker in options collects an insurance or risk premium for offering protection through options.

Carry as metaphor unlocks much insight into contemporary finance. Although the assets they invest in are very different, many widely employed trading and investment strategies are similar in structure to those involved more transparently in the original foreign-exchange focused carry trade. For example, like that of the original carry trade, much of the financing of hedge funds and private equity—the key components of the “alternative asset” universe—is provided by banks, securities dealers, and other money market lenders at short term, or in the form of leveraged loans, typically issued with longer terms to maturity but with floating interest rates. The borrowing is collateralized by the assets acquired, like the purchased currency in the original carry trade, and the ultimate creditors also include insurers, pension funds, and mutual funds.

These assets differ vastly from one another, ranging from government bonds to ownership stakes in small, troubled, or undervalued companies, but the type of funding is similar. The “carry regime,” as the authors term it, depends on funding liquidity, that is, the ability to readily finance asset positions predominantly with borrowed rather than own funds. Profitability and even investors' solvency are highly vulnerable to an abrupt retreat from short-term lending or rise in money-market rates, and to a decline in the funded assets' values. The ultimate providers of this liquidity are central banks and regulators. They provide liquidity in two distinct ways, through monetary policy, but also through their explicit and implicit guarantees of the liabilities of the key lenders—the banks, dealers, and institutional investors.

Similarly to option selling, carry trades rely on a persistent gap between the market values and likely realized future values of assets that are risky and distasteful to hold—an emerging-markets currency, a long-term credit-risky bond, a troubled firm's shares—because they are susceptible to sudden large drops in value or tend to drop in value and become hard to sell in troubled times. Just as for option selling, the result is a pattern of small gains punctuated by

sporadic large losses. Carry trades are rewarded for the liquidity and the insurance against adverse changes in price or credit quality they provide to the market, liquidity and insurance that are ultimately provided and backstopped by central banks and governments.

Financial systems in which many participants rely predominantly on leverage and carry rather than the quality of the assets to reach their return goals are unstable since small losses can trigger widespread insolvencies. As the authors write, “[t]he instability of levered trades is the key reason that carry trades eventually crash.” As more funding is drawn into the business of providing liquidity and protection from price fluctuations, the risk premiums with which these services are rewarded decline, liquidity becomes more robust, and volatility is dampened.

Leveraged trades thrive on high volatility and at the same time suppress volatility, generating a self-perpetuating and self-amplifying mechanism, a phenomenon known as the paradox of volatility. As researchers at the Bank for International Settlements have also emphasized, volatility and risk premiums are low, asset prices are highest, and the world appears safest, just when leveraged trading is most prevalent, the degree of leverage is highest, and the trading and investment volumes are greatest.

The Rise of Carry lays out these mechanics clearly and connects them to their primary sources. The mechanism relies on incentives. Both the traders’ and the financiers’ incentives are asymmetrical, lopsided, with very large returns in one direction for asset prices and limited returns in the other. But if the potentially unlimited losses of the financier—bank, dealer, or institutional investor—can also be capped by recourse to the government or a lender of last resort, you get the system we are living in today.

Although it might appear on the surface that this pattern must lead to inflation, as the authors point out, it is in fact deflationary and a drag on real growth. The pattern of expansions characterized by increased leverage is interrupted by regular financial crises large and small. These are countered by further accommodation. Even if central banks tighten credit conditions once leverage has become alarming, they will surely loosen in response to the crackup.

The result is a long-term pattern of misallocation of resources and survival of zombie firms kept alive by low interest rates that could not thrive in a normal interest rate environment. Highly leveraged firms may become reluctant to engage in promising new activity since the

returns would flow to their creditors rather than owners. As the authors state, “central bank interventions together with excessive government involvement in economies . . . have resulted in economies with too little savings, too much debt-financed consumption, and low prospective returns on real investment.” *The Rise of Carry* is a valuable new perspective on this old problem and recommended reading for anyone seeking a deeper understanding and timely reminder of this important and disturbing phenomenon.

Allan M. Malz
Columbia University

Classical Liberalism and the Industrial Working Class: The Economic Thought of Thomas Hodgskin

Alberto Mingardi
Oxford: Routledge, 2021, 160 pp.

The temptation for political movements to claim venerable thinkers from the past among their own is sometimes irresistible. But the truth is often more complicated. That seems to be the case with Thomas Hodgskin (1787–1869), the radical English writer and campaigner of whom Alberto Mingardi has recently written a highly readable intellectual biography. Even though Hodgskin’s Wikipedia entry continues to describe him as a socialist, Mingardi shows that many of his views on the rights of workers and the benefits of commercial society were highly sympathetic to classical liberals.

Hodgskin was an autodidact who grew up poor yet rose to become senior editor of *The Economist*, rubbing shoulders with the most prominent thinkers of his time, such as Jeremy Bentham, Jean-Baptiste Say, and Herbert Spencer. An early stint in the Royal Navy, which he was sent off to join at age 12 by his well-off but spendthrift father, seared him with a lifelong dislike of coercion and arbitrary power. His earliest published work, *An Essay on Naval Discipline*, was a reaction to this experience. In it, he made an uncompromising case against impressment—forced enlistment—which Hodgskin believed unjust, inefficient, and based on the erroneous assumption that Britons would not voluntarily sign up to defend their country.

Mingardi ably situates Hodgskin’s life and writings in their historical context. It was not advocacy for the dictatorship of the proletariat that made Hodgskin a radical, but his opposition to aristocratic

privileges that tipped the scales against workers and the poor. Even though some of his writings had socialist-sounding titles (such as *Labour Defended Against the Claims of Capital* and *Popular Political Economy*), Hodgskin's critique of capitalists focused on their use of the legislative process to alter economic outcomes in their favor, different from what they would have been in a truly free market. Were Hodgskin alive today, it is likely—from Mingardi's telling—that he would rail not against capitalism but against what we now call crony capitalism.

Many of Hodgskin's views aligned with those of modern libertarians. He was a lifelong free trader and an optimist regarding the gradual improvement of man's condition thanks to industrialization. He vehemently opposed Malthusian fatalism, arguing that "the foundation of all national greatness is the increase of the people." Unlike many later socialists, Hodgskin was not enamored of the notion of benevolent government monopolies, believing for instance that the Bank of England's pre-eminent position owing to its unique charter had caused "inconceivable mischief." Nor was he against paper currency, so long as competitive issue could check its inflationary abuse. By illuminating this and other aspects of Hodgskin's prolific writing, Mingardi paints an intellectual portrait that is hard to assign to any conventional political tribe, although it is closer to liberalism than past accounts have recognized.

Being himself a scholar of classical liberal thought—as well as a Cato fellow—Mingardi can spot facets of Hodgskin's oeuvre that more generalist writers might miss, such as his implicit recognition of the special role of entrepreneurship in value creation, separate and distinct from those of capital and labor. So also with Hodgskin's views on the role of the state: while his intense skepticism that politics could ever lead to betterment pushed him toward anarchism, he stopped "a step shy of openly advocating the extinction of the state." He did so, Mingardi writes, out of a belief that letting society develop spontaneously was preferable to "detailed recipes" for change. But, by the same token, Hodgskin mocked the supposedly "scientific government" of Prussia—which he had a chance to study during his travels to Germany—concluding that the "abundance of orders" to which such government gave rise actually impeded the healthy development of society.

Hodgskin's wariness of legislation as a tool for general improvement also caused him to be less optimistic about the expansion of

the franchise. He went so far as to write that “the people, whether middle or lower classes, are not made wise by sharing in the power of Government.” And if Hodgskin welcomed the European revolutions of 1848 as sounding the death knell on government “by the cat-o’-nine-tails and the gallows,” that did not stop him from haranguing France’s revolutionary government for instituting a program of public benefits that lured idle workers into Paris. Far from leveling the playing field between masters and workers, such “class legislation” only flipped the inequality. He likewise ridiculed French leader Lamartine’s vow to transform political economy into “the science of fraternity” as “the most extraordinary instance of ignorance and assumption in a public man we have ever met with.”

The picture that emerges from Mingardi’s account is of an intellectual maverick who had strong views and expressed them clearly. Hodgskin was not a socialist, other than by the loose standards of his time, according to which anyone opposed to the status quo was one. But Mingardi concludes that Hodgskin’s association with the workers’ movement made him less influential with later liberal thinkers than he might—and perhaps should—have been. He did influence Herbert Spencer, a friend and brief *Economist* colleague whose work gained greater notoriety and following than Hodgskin’s, but apparently Spencer did not adequately acknowledge the connection. Yet one can also hear echoes of Hodgskin in F.A. Hayek’s case against central planning, based on the argument that the central authority invariably lacks essential information about the particular circumstances of the agents it is attempting to direct.

Mingardi’s perceptive analysis thus succeeds at revealing a different character from the one that previous biographies have described—no less than “the advocate of an uncompromisingly radical kind of classical liberalism.” For that reason alone, this book deserves to be called “definitive.”

Diego Zuluaga
Cato Institute

Cato Institute

Founded in 1977, the Cato Institute is a public policy research foundation dedicated to broadening the parameters of policy debate to allow consideration of more options that are consistent with the principles of limited government, individual liberty, and peace. To that end, the Institute strives to achieve greater involvement of the intelligent, concerned lay public in questions of policy and the proper role of government.

The Institute is named for Cato's Letters, libertarian pamphlets that were widely read in the American Colonies in the early 18th century and played a major role in laying the philosophical foundation for the American Revolution.

Despite the achievement of the nation's Founders, today virtually no aspect of life is free from government encroachment. A pervasive intolerance for individual rights is shown by government's arbitrary intrusions into private economic transactions and its disregard for civil liberties. And while freedom around the globe has notably increased in the past several decades, many countries have moved in the opposite direction, and most governments still do not respect or safeguard the wide range of civil and economic liberties.

To address those issues, the Cato Institute undertakes an extensive publications program on the complete spectrum of policy issues. Books, monographs, and shorter studies are commissioned to examine the federal budget, Social Security, regulation, military spending, international trade, and myriad other issues. Major policy conferences are held throughout the year, from which papers are published thrice yearly in the Cato Journal. The Institute also publishes the quarterly magazine *Regulation*.

In order to maintain its independence, the Cato Institute accepts no government funding. Contributions are received from foundations, corporations, and individuals, and other revenue is generated from the sale of publications. The Institute is a nonprofit, tax-exempt, educational foundation under Section 501(c)3 of the Internal Revenue Code.

CATO INSTITUTE
1000 Massachusetts Ave., N.W.
Washington, D.C. 20001
www.cato.org