the CATO JOURNAL

AN INTERDISCIPLINARY JOURNAL OF PUBLIC POLICY ANALYSIS

ARTICLES

BOOK REVIEWS

VOLUME 37 NUMBER 1 ● PUBLISHED BY THE CATO INSTITUTE ● WINTER 2017

the CATO JOURNAL

VOLUME 37 NUMBER 1 • PUBLISHED BY THE CATO INSTITUTE • WINTER 2017

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 *Cabell's Directory*, *Journal of Economic Literature*, *Current Contents/Social and Behavioral Sciences*, *America: History and Life*, *Historical Abstracts*, *International Political Science Abstracts*, and the ABI/INFORM database.

Printed in the United States of America.

CATO JOURNAL • VOLUME 37 NUMBER 1 • WINTER 2017

ARTICLES

BOOK REVIEWS

The Microeconomic Perils of Monetary Policy Experiments

Charles W. Calomiris

The Federal Reserve (Fed) tells us that its experiments in quantitative easing (QE) are stimulating the economy, and financial news services obediently echo that message. But there is reason to believe that, under current unusual market conditions—especially near-zero interest rates and tightening prudential regulatory requirements—Fed actions may be having little effect, or even effects opposite to those the Fed intends. Because of changes in how the tools of the Fed work under the current unusual circumstances, raising interest rates and shrinking the Fed's balance sheet have little effect, or even a positive effect, on economic activity. That is especially true when one adopts an appropriate medium-term perspective on monetary policy and takes into account the benefits of avoiding the destabilizing potential consequences for asset markets of the Fed's current mortgage-backed securities (MBS) purchases.

The argument for raising interest rates at this time, however, is not mainly one about incremental accommodating or tightening; rather, it is about restoring predictability to monetary policy by reviving the federal funds market. It is essential to return to a situation where

---

Cato Journal, Vol. 37, No. 1 (Winter 2017).

Charles W. Calomiris is Henry Kaufman Professor of Financial Institutions at Columbia Business School, Director of the Business School's Program for Financial Studies and its Initiative on Finance and Growth in Emerging Markets, and a professor at Columbia's School of International and Public Affairs. This article was first presented at the Manhattan Institute's October 7, 2016, meeting of the Shadow Open Market Committee. The author thanks John Taylor, Allan Meltzer, and Peter Ireland for helpful comments.

(based on decades of empirical evidence) the Fed can use changes in the federal funds rate as its policy tool and confidently project whether its incremental policy actions are stimulative or contractionary. We should not continue under the current unusual circumstances, in which the Fed and the markets cannot tell whether incremental actions by the Fed constitute stepping on the gas pedal or the brake.

I propose a modest 2 percentage point increase in the federal funds rate, along with other measures that on balance would probably have little immediate effect on the economy. Those actions, however, would ensure a speedy return to a normal policy environment—where bank reserves are once again "scarce," where the federal funds rate takes on its traditional usefulness as a gauge of monetary policy, and where Fed actions would have much more predictable consequences.

Background

Monetary policy affects the economy through a variety of "transmission mechanisms." For example, when the central bank expands its balance sheet through securities purchases, there is an increase in "high-powered money" and reserve holdings of banks at the central bank. If commercial banks maintain a constant fraction of their deposits as reserves, the central bank's expansion of securities purchases will create an expansion of bank deposits, which are used to fund bank assets such as loans. This process of deposit and loan expansion that may result from central bank securities purchases operates through what is sometimes called the "loan-supply" transmission mechanism of monetary policy.

Loan-supply changes are only one of the ways monetary policy affects the economy. Purchases of securities affect the economy through channels other than the expansion of the supply of bank loans, such as changes in market interest rates. If the central bank is targeting a particular class of assets in its purchases, the consequent subsidy the Fed is providing to some assets (through its willingness to absorb risks related to the term structure or the mortgage market) may also affect the relative prices of securities. Recent MBS purchases by the Fed, or long-term Treasury purchases, are best seen as a form of fiscal—not monetary—policy, which subsidizes certain risks and thereby favors certain investments. For example, Fed

purchases may have important effects on the MBS spread over Treasuries or on the term structure of Treasuries, which may affect investment in housing.

Because exchange rates reflect the forward-looking value of the dollar relative to other currencies, changes in the current or prospective supply of dollars (controlled via the Federal Reserve's ultimate monopoly over the supply of high-powered money) via securities purchases, or other actions or statements by the Federal Reserve, also can affect exchange rates, which in turn influence the supply and demand for exports, imports, and international capital flows.

Further complicating the analysis of monetary policy effects are the numerous tools the central bank can employ to influence markets through each of these channels. The Federal Reserve's purchases and sales of securities are one tool, but the Fed also sets its discount rate (through which it lends funds to member banks), varies its reserve requirement (which can influence the extent to which an expansion of high-powered money results in changes in deposits and loans), determines the interest on reserves (higher interest reduces bank loan supply by encouraging the accumulation of excess reserves), and promulgates regulations that affect banks and other intermediaries' abilities to supply loans and deposits or engage in repurchases (repos). For example, with respect to regulatory influences, when the Fed recently increased bank capital requirements by setting a "Supplementary Leverage Requirement," which required repos to be included in the definition of bank leverage for some banks, the market supply of repos decreased (Allahrakha, Cetina, and Munyan 2016).

Apart from all of those current actions, the Federal Reserve can also influence markets by issuing "forward guidance" about its future intentions with respect to any of those actions, either through speeches or explicit forecasts of the future path of interest rates and other key variables that it can influence.

During some periods (e.g., from 1983 to 2001), the complexity of the monetary policy transmission mechanism did not pose major problems for predicting the influence of monetary policy. That was the case because (1) the Fed's actions were limited to targeting the federal funds rate, (2) the primary tool was open market operations, (3) the Fed's activities were confined almost exclusively to the purchase and sale of Treasury bills, (4) important regulatory policies (capital and liquidity requirements) were known and not subject to

dramatic change, and (5) the Fed seemed to be following an implicit rule that linked its federal funds target to current levels of inflation and unemployment.

In the current environment, however, it is very hard to know how to gauge the consequences of Fed actions, many of which make use of policy instruments that have not been used in the past. No empirical record exists from past Fed behavior from which to form reliable estimates of the consequences of current Fed behavior. Furthermore, these unprecedented policies are interacting with a unique economic environment (most obviously, one in which nominal interest rates have remained near zero for many years and regulatory policy is subject to constant change). The combination of a unique environment and the use of many new tools (quantitative targeting of the Fed's balance sheet, as in QE1, QE2, and QE3; Fed involvement in the repo market; Fed setting of rates of interest paid on reserves; Fed guidance statements about likely future policy), operating through many potential channels of influence, has made it almost impossible to forecast the influence of Fed policy actions on the real economy. In pointing to policy uncertainty, I am not referring to the well-known problem that the future actions of the Fed are unpredictable because of the absence of a clear policy rule (a problem that many economists have lamented for some time); rather, I am pointing out that it is very hard to know what effects even *monetary policy* actions are having on the economy today, much less what effects *Fed policy* (defined more broadly to include its fiscal policies and regulatory actions) is having on the economy.

This is not just a problem of gauging the precise magnitude of policy effects. It is a more fundamental problem in gauging even the direction of influence. Despite confident Fed pronouncements that it is helping to promote recovery, it is not at all clear that holding interest rates at their current near-zero levels, or maintaining the bloated Fed balance sheet stimulates the U.S. economy. In the medium run, because of concerns about destabilizing bubbles that Fed policy may be creating in agricultural land, housing, commercial real estate, and securities markets, an even stronger case can be made that contracting the Fed's balance sheet and raising interest rates might be expansionary. As John Taylor wrote in 2013: "The Fed's current zero interest-rate policy also creates incentives for otherwise risk-averse investors—retirees, pension funds—to take on questionable investments as they search for higher yields in an

attempt to bolster their minuscule interest income. The low rates also make it possible for banks to roll over rather than write off bad loans, locking up unproductive assets."

My best guess is that Fed monetary policy currently is having a small effect on the economy, which is slightly contractionary (but highly uncertain in the direction of its effect). Meanwhile, Fed fiscal policy is slightly expansionary in the short run (primarily through its implied subsidy to housing finance risk) but likely contractionary in the medium run (by contributing to the possibility of a future asset price crash). Finally, Fed regulatory policy is clearly contractionary in the short and medium runs. My main point, however, is that the unprecedented tools and environment in which the tools are being applied make it hard to know the *direction* of policy effects, much less their magnitude.

My analysis of current Fed policy actions divides into five parts: (1) the consequences for loan supply of prudential regulatory tightening and the reduced market value of bank equity resulting from low interest rates, (2) repo funding cost consequences of monetary policy, (3) exchange rate consequences, (4) corporate balance sheet consequences, and (5) asset price consequences. I conclude that the Fed would likely improve the short- and medium-term growth and stability of the U.S. economy if it *raised* interest rates significantly, restored the functioning of the federal funds market, reduced its balance sheet, and withdrew from the repo market.

Bank Loan Supply, Capital Regulation, and Equity Value Consequences of Monetary Policy

Monetary economics textbooks characterize the loan-supply channel of monetary policy as follows: (1) the central bank buys securities, (2) this results in excess reserves credited to bank balance sheets, (3) banks reduce excess reserves by lending funds and increasing deposits. There is substantial empirical evidence suggesting that in the past, expansionary central bank policy (reductions in the targeted interest rate) are associated with expansions in the supply of lending and a reduction in loan spreads.

But average behavior from the past may not be very useful for predicting loan-supply effects under current circumstances. First, regulatory policy may constrain loan growth. There is ample empirical evidence that regulation has reduced the willingness of banks to

expand deposits and loans. Most obviously, the largest U.S. banks recently have turned down large corporate deposit balances (an unprecedented action) because of the regulatory costs of maintaining those deposits on their balance sheets. More generally, there is a large empirical literature showing that minimum capital ratio requirements can play an important role in constraining bank loan supply.[1] If monetary policy expansion increases banks' excess reserves but capital requirements act as a binding constraint on loan supply, banks will not expand loans and deposits as their excess reserves rise. Recent increases in minimum capital ratios and liquidity coverage imposed by the Fed have limited the ability of banks to transform their excess reserves into loans and deposits (see Allahrakha, Cetina, and Munyan 2016).

Putting aside the effects of regulatory policy (minimum book value ratio requirements and minimum liquidity book value ratio requirements), bank loan supply may be limited by anything that reduces the *market value* of bank equity. The market value of equity matters because banks' ability to operate depends on their maintaining low default risk as perceived by the market. Standard models of finance (e.g., the Black-Scholes-Merton model) characterize default risk as a function of asset risk and the market value of equity relative to assets. Given that banks target low default risk, if a shock reduces banks' market values of equity, banks must respond either by raising capital or by reducing risk (Calomiris and Wilson 2004).[2] Default risk reduction via loan-supply contraction is the most common bank response.

There is a strong correlation between bank equity values and loan-supply growth in recent years (Bird 2016). Furthermore, the 2008 crisis vividly illustrated how reductions in banks' equity values can raise the counterparty risk of banks enough to produce a systemic crisis as troubled banks become unable to roll over their short-term debts (Calomiris and Herring 2013).

[1]See the review of the literature in Aiyar, Calomiris, and Wieladek (2015), and the recent microeconomic evidence for the United Kingdom in Aiyar, Calomiris, and Wieladek (2014a, 2014b, 2016) and Aiyar et al. (2014).

[2]Following the same logic, an increase in uncertainty, for a constant market value of equity ratio, will also contract bank lending. Bordo, Duca, and Koch (2016) show that uncertainty shocks in fact have that effect. Thus, uncertainty about the future path of monetary policy has been another adverse influence on loan supply.

To the extent Fed policy damages bank equity values, loan supply will contract. But how does Fed policy adversely affect the market value of bank equity? Doesn't the Fed's commitment to low interest rates reduce discount rates for corporate earnings, thereby propping up stock prices in the economy, and shouldn't that increase the market value of banks' equity?

Interestingly, in the case of bank equity, monetary policy is having the opposite effect. As Calomiris and Nissim (2014) show, once it became clear that the Fed would be maintaining very low interest rates for a protracted period of time, the market value of "core" deposit relationships changed from a positive to a negative influence on banks' cash flows and the market values of bank equity. Persistent low interest rates meant that banks were stuck with highly negative cash flows from branch leases and employee compensation, but did not reap the benefits of interest cost savings that usually result from having invested in branch networks to serve core deposit customers. Calomiris and Nissim (2014) show that this is a major contributor to the persistently low market-to-book equity ratios of U.S. bank holding companies.

Clearly, both Fed monetary policy and Fed regulatory policy independently are constraining the supply of loans. This explains why the expansion of the Fed's balance sheet has resulted in a huge persistent expansion of excess reserves rather than a substantial increase in bank lending.[3]

Repo Funding Costs and Loan Supply

Over the past several decades, repo has been an important alternative, off-balance sheet source of funding for lending in the U.S. economy, by both regulated banks and nonbank lenders. But, as Gorton and Muir (2016) emphasize, the massive expansion of the Fed's balance sheet over the past decade has withdrawn a large amount of low-risk collateral from the market, thereby making repo funding of loans and other financial transactions harder to arrange.

[3]David Malpass has written numerous op-ed articles in the *Wall Street Journal* (e.g., Malpass 2016) over the past several years arguing that the Fed has been contracting loan supply, and that monetary policy has been a contractionary influence on the economy.

Furthermore, the enactment of the supplementary leverage ratio (SLR) requirement (a policy that includes the quantity of repos in the regulatory measure of leverage, which was announced in 2012, and began to affect bank behavior at that time) has also reduced the supply of repo funding. Allahrakha, Cetina, and Munyan (2016) find that this new requirement increased the cost of repo finance by regulated U.S. institutions.

It is important to recognize that the Fed's new role as a repo counterparty (since 2013) does not offset the collateral drain produced by the Fed's accumulation of repo collateral on its balance sheet. The Fed lends its collateral into the market in exchange for cash. Because the Fed engages in triparty transactions, the repo collateral employed in Fed transactions cannot be re-hypothecated in other transactions. On net, Fed "reverse repos" drain cash from the market and do not provide collateral that can be used by other repo market participants.

The Fed's dual role as a regulator and a repo counterparty also raises important new and disturbing questions about a new conflict of interest. As a repo counterparty, the Fed benefits financially from its imposition of the SLR, which reduces its competitors' abilities to engage in similar transactions. Is it conceivable that the Fed might have taken into account its own financial benefits from being able to engage in reverse repo on more favorable terms when setting regulations for its competitors? Yes, it is. When the Fed began contemplating its reverse repo tool, it was already cognizant that it might want to engage in a large amount of such transactions. The Fed was concerned that if it failed to raise sufficient revenue from lending securities in the repo market, its expected accounting contribution to federal deficits would rise. As many observers noted, the likely financial costs to the Fed from experiencing losses in the future has potentially important adverse political implications for the Fed. Indeed, one of the reasons the Fed planned to engage in massive amounts of repo, rather than selling securities into the market, was political. Employing repos rather than selling securities allows the Fed to avoid the expected accounting consequences of recognizing capital losses from securities sales, which (under current Fed accounting rules) would increase its measured contribution to government deficits. I do not claim to know that the Fed's SLR was motivated in part by a desire to improve its competitive position in the repo market, but the coincidence in timing between the SLR and the Fed's entry into the repo market is disturbing, and there is no question that

the Fed suffers a conflict of interest from being both a repo counterparty and a regulator. That conflict adds to the preexisting list of conflicts that would be resolved by removing the Fed from its role in setting regulatory standards.

In summary, through the combination of the Fed's accumulation of Treasuries and MBS, its decision to use reverse repo rather than sales of those securities in any future tightening of monetary policy, and its SLR regulation, the Fed has been increasing, and will continue to increase, the cost of repo funding, which is another contractionary influence on the supply of lending in the economy.

Exchange Rates

At a recent conference, I reviewed the adverse consequences of Fed policies for bank equity values and repo funding costs and argued that the loan-supply consequences of Fed policy have been contractionary. I was expecting a vigorous disagreement from the representatives of the mainstream point of view, but instead a former high-ranking Federal Reserve Board economist responded that the Fed's internal model of the transmission mechanism of monetary policy agreed with my conclusion. The Federal Reserve Board model, according to him, implies that monetary policy currently is having a contractionary effect on loan supply. But, he said, according to the Fed's model, that contractionary effect is more than outweighed by other effects, especially an expansionary effect *operating through exchange rate depreciation*.

I was stunned. How could the Fed believe that its attempt to depreciate the dollar would work, especially at a time when so many other countries (most obviously, members of the eurozone and Japan) are struggling to restore growth? Indeed, according to IMF research, augmented in Taylor (2016), Fed decisions to lower interest rates have negative effects on economic performance outside the United States—presumably, by disadvantaging foreign exporters. Furthermore, if depreciation is harmful to foreign countries, then wouldn't other central banks respond to Fed actions by intervening to offset the Fed's actions? If so, then the ultimate consequence for exchange rates of Fed policy might be nil. Finally, isn't this a risky strategy for the U.S. central bank? The desirability of avoiding the risks that attend unpredictable depreciation wars among central banks was supposedly a major lesson of the 1930s and a major

reason the IMF was established. How can the Fed defend a policy that will likely be ineffectual (because it will be largely offset by other central banks) and likely will produce a new source of international conflict with the U.S.'s major trading partners?

Even if the Fed were successful in depreciating the dollar, it is not clear that this would boost the U.S. economy. According to the model and evidence in Phelps, Hoon, and Zoega (2005), if monetary policy succeeded in weakening the dollar, that would make U.S. firms respond by increasing their prices, cutting wages, and reducing employment.

I conclude that the Fed's attempt to depreciate the dollar is unwise geopolitically and is unlikely to be an important positive influence on the U.S. economy. The effects of Fed policy on the dollar likely will be offset by other central banks. Moreover, even if the Fed succeeds in depreciating the dollar, U. S. output growth is unlikely to improve and could actually fall (Phelps, Hoon, and Zoega 2005).

Corporate Balance Sheets

According to neoclassical investment theory, lowering interest rates should stimulate investment. However, the link between interest rates and corporate investment has been quite weak in empirical studies of investment. Furthermore, more recent models of investment under asymmetric information have emphasized the importance of healthy balance sheets and the potential constraints on investment that arise when firms suffer adverse shocks to their net worth. For that reason, it is useful to consider balance sheet consequences of the low-interest rate environment.

Keeping interest rates low for a protracted period of time can reduce the net worth of corporations that have large outstanding long-term debts and that operate defined benefit pension plans. According to the *Economist* ("Fade to Grey," September 24, 2016), the consulting firm Mercer estimates that defined benefit plans of large U.S. firms are only funded 77 percent, with a resulting deficit of $570 billion.

Of course, low interest rates increase corporate debt-raising capacity and raise the present value of firms' expected future cash flows from operations, and those influences should boost stock prices and investment for many firms. It is likely that, on net, the effect of low interest rates on stock prices raises investment a bit, as suggested

by Tobin's Q model of investment. However, investment has not responded very favorably to monetary policy thus far in the recovery. In comparison with other business cycle recoveries, net investment as of 2016 is only about half what past patterns suggest that it should be (Gutierrez and Philippon 2016). The growing funding problem for defined benefit plans, along with the negative effects of Fed policy on loan supply discussed above, likely play a part in the explanation of this puzzling investment behavior.

Asset Prices

QE2 and QE3 were justified primarily as means of reducing long-term interest rates. A microeconomic analysis of the mortgage market by DiMaggio, Kermani, and Palmer (2016) finds that the primary positive effects of QE2 and QE3 were the result of the fiscal policy actions of the Fed, not the growth in its balance sheet. In fact, only the Fed's purchases of MBS seem to have mattered for mortgage prices, suggesting that Fed fiscal intervention to subsidize interest rate risk had little effect.

This finding has a disturbing implication. The primary identifiable positive influence of Fed policy actions for the economy seems to be a new housing price rise caused by Fed fiscal subsidies in the MBS market. In the short run, this has been a positive for the economy, as investment in new housing has boosted income and employment, and house price appreciation has helped to fund increased nonhousing consumption. But, as we learned in the recent subprime crisis, there are significant medium-run risks associated with creating asset price bubbles. Using monetary policy to create a short-lived asset price boom is a way to gain short-term growth at the expense of possible medium-term contraction.

A Monetary Policy Reset

In light of the above observations, consider the following policy proposal. The Fed should abandon its recent untested and risky policy experiments and announce that it is returning to a predictable structure for monetary policy, where it will be clear when the Fed is applying the brake versus the throttle. The key is to raise interest rates to make reserves scarce and employ complementary policies that will restore the usefulness of the federal funds market as the

primary instrument of policy. Doing so would allow the Fed to rely on decades of evidence about the connection between raising or lowering the federal funds rates and consequent changes in economic activity. Furthermore, the Fed should undertake and articulate a systematic approach to policy that makes it clear what the use of that tool is intended to accomplish and how policy will evolve in the future.

Specifically, I propose the following five actions:

1. Announce an intent to raise interest rates (initially repo and federal funds rates, later only federal funds rates) by 2 percentage points in 25 basis point increments over the next 18 months. A 2 percent hike would still imply an accommodative level of interest rates (a real interest rate of zero). Incrementally, this might be accommodative or contractionary, but I do not believe it will be a very large net influence in either direction because of offsetting influences. It would have an immediate positive effect on bank equity values and would boost lending, but it would also have a negative effect on real estate borrowing costs. The main point of the rise in interest rates is to restore normalcy to the federal funds market. Given that objective, other complementary policies are also needed.
2. Eliminate the involvement of government-sponsored enterprises in the federal funds market and raise required reserves for large banks (which currently maintain massive amounts of excess reserves on their balance sheets). Those actions would allow the federal funds rate to serve as a monetary policy instrument once again. The higher reserve requirement would have additional medium-term advantages by reducing inflationary expectations that could result from a boom in bank lending and reduce the scale of the open-market sales needed to achieve the Fed's interest rate target.
3. Set an interest rate rule for the interest paid on reserves that fixes that rate to the federal funds rate less 10 basis points. This action would remove another source of uncertainty and make it clear that the Fed is committed to shrinking its balance sheet to a normal size, rather than undertaking an unpredictable manipulation of interest on reserves to influence bank loan supply.
4. Announce a policy rule (e.g., some version of the Taylor Rule, or some other systematic approach) that would substantially reduce uncertainty about Fed actions for the next several years.

This would make whatever changes the Fed undertakes in the federal funds market predictable, which would strengthen its influence and reduce policy uncertainty.

5. End Fed reverse repo transactions to signal the willingness to sell securities in the Fed's portfolio to achieve its federal funds rate targets and make it clear that the Fed will do so without any concern about the accounting consequences regarding the Fed's prospective contribution to the government deficit. Doing so would also avoid future consequences for the repo market of unpredictable Fed securities sales.

Conclusion

Despite the confident pronouncements of central bankers and the compliant echo chamber of the financial press, the combination of new and untested tools being employed by the Fed—and the unusual circumstances under which they are being employed—make the consequences of changes in monetary policy for the economy highly uncertain.

If the Fed abandoned its untested and risky policy experiments and adopted a systematic policy approach, raised interest rates, and restored the usefulness of the federal funds market as the primary instrument of policy, the Fed would once again be in a position where its incremental influence on nominal GDP growth would be predictable, at least directionally.

The goal of this proposed monetary policy reset is to *restore predictable efficacy* to monetary policy. This could be done with little immediate economic consequences. It might be expansionary in the medium run, and possibly also in the short run, but it also might be mildly contractionary. It is hard to gauge the overall incremental consequences for the economy precisely because different aspects of the transmission mechanism imply opposite effects. Loan supply would rise alongside rising bank equity prices, declining repo funding costs, and reduced policy uncertainty, but housing finance costs would rise and house price appreciation would slow to a sustainable long-run path. In any case, policy would remain accommodative at a zero real federal funds rate. Importantly, the Fed would be in a position to tighten or loosen from that point going forward, would be much more able to achieve legitimate stabilization objectives, and both the Fed and the markets would have a clearer sense of the relationship between Fed actions and economic consequences.

References

Aiyar, S.; Calomiris, C. W.; Hooley, J.; Korniyenko, Y.; and Wieladek, T. (2014) "The International Transmission of Bank Capital Requirements: Evidence from the UK." *Journal of Financial Economics* 113 (September): 368–82.

Aiyar, S.; Calomiris, C. W.; and Wieladek, T. (2014a) "Does Macro-Prudential Regulation Leak? Evidence from a U.K. Policy Experiment." *Journal of Money, Credit, and Banking* 46 (1): 181–214.

_____ (2014b) "Identifying Channels of Credit Substitution When Bank Capital Requirements Are Varied." *Economic Policy* 29 (77): 47–71.

_____ (2015) "Bank Capital Regulation: Theory, Empirics, and Policy." *IMF Economic Review* 63: 955–83.

_____ (2016) "How Does Credit Supply Respond to Monetary Policy and Bank Minimum Capital Requirements?" *European Economic Review* 82 (January) 142–65.

Allahrakha, M.; Cetina, J.; and Munyan, B. (2016) "Supplementary Leverage Ratio and Repo Supply." Working Paper, Office of Financial Research, U.S. Treasury.

Bird, M. (2016) "Banks Get Crushed but Still Lend." *Wall Street Journal* (August 26).

Bordo, M. D.; Duca, J. V.; and Koch, C. (2016) "Economic Policy Uncertainty and the Credit Channel: Aggregate and Bank Level U.S. Evidence over Several Decades." NBER Working Paper No. 22021 (February).

Calomiris, C. W., and Herring, R. J. (2013) "How to Design a Contingent Convertible Debt Requirement That Helps Solve Our Too-Big-to-Fail Problem." *Journal of Applied Corporate Finance* 25 (Spring): 66–89.

Calomiris, C. W., and Nissim, D. (2014) "Crisis-Related Shifts in the Market Valuation of Banking Activities." *Journal of Financial Intermediation* 23: 400–35.

Calomiris, C. W., and Wilson, B. (2004) "Bank Capital and Portfolio Management: The 1930s 'Capital Crunch' and Scramble to Shed Risk." *Journal of Business* (July): 421–56.

DiMaggio, M.; Kermani, A.; and Palmer, C. (2016) "How Quantitative Easing Works: Evidence on the Refinancing Channel." Columbia Business School Research Paper No. 16-1. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2700354.

Gorton, G., and Muir, T. (2016) "Mobile Collateral versus Immobile Collateral." BIS Working Paper No. 561.

Gutierrez, G., and Philippon, T. (2016) "Investment-less Growth: An Empirical Investigation." Working Paper, New York University (October).

Malpass, D. (2016) "Pro-Growth Tools for the Frozen Fed." *Wall Street Journal* (October 6).

Phelps, E. S.; Hoon, H. T.; and Zoega, G. (2005) "The Structuralist Perspective on Real Exchange Rate, Share Price Level and Employment Path: What Role Is Left for Money?" In W. Semmler (ed.), *Monetary Policy and Unemployment: The US, Euro Area, and Japan*, 107–32. London: Routledge.

Taylor, J. (2013) "Fed Policy Is a Drag on the Economy." *Wall Street Journal* (February 1).

_____ (2016) "A Rules-Based Cooperatively Managed International Monetary System for the Future." In C. F. Bergsten and R. Green (eds.), 217–36. *International Monetary Cooperation: Lessons Learned from the Plaza Accord after Thirty Years*. Washington: Peterson Institute for International Economics.

The Japanese Experience with QE and QQE

John Greenwood

This article provides an overview of the three episodes of quantitative easing (QE) pursued by the Bank of Japan (BOJ) since 2001. It begins with a brief account of the initial reluctant shift to unorthodox policies under BOJ Governors Hayami and Fukui in 2001–06 (here designated QE1) and then covers the equally reluctant adoption of QE by Governor Shirakawa in 2010–13 (QE2). The article then turns to an account of the attempt since April 2013 by the BOJ under Governor Kuroda, designated "quantitative and qualitative easing" (QQE), to revive the economy and achieve a 2 percent inflation target. None of these attempts at QE has been successful in raising the broad money growth rate for M2 sustainably above the 2–3 percent per annum range where it has languished for the past 25 years. Consequently, Japan's attempts at QE have all failed to raise the equilibrium level of Japanese nominal GDP by any material magnitude, and so far, attainment of the 2 percent inflation target under QQE has remained elusive. At the time of writing (October 2016), the Japanese economy therefore continues to grow at a low rate with periodic lapses into deflation. After discussing the case of Japan, the article compares the experience of the United States in 1929–33, when there was no QE, and the experience of 2008–14, when the

---

Cato Journal, Vol. 37, No. 1 (Winter 2017).

John Greenwood is Chief Economist at Invesco. This article was originally presented at a conference hosted by the Institute of International Monetary Research and the Institute of Economic Affairs, November 3, 2016, titled "Quantitative Easing: Triumph or Folly: Is It Still Relevant?"

Fed conducted QE over three periods. The comparison is deliberately focused on the quantitative aspects of the policy, not its interest rate effects. Finally, the article explains that there are two brands of QE, and that the failure of QE in Japan is fundamentally due to the choice of the wrong brand of QE. Given the type of QE that the Japanese authorities have chosen, the policy cannot be expected to succeed, except under limited conditions.[1] If QE were to be implemented according to a different design, the prospects of success would be much greater. In brief, the primary reason for the failure of BOJ-style QE or QQE derives from the habitual tendency to buy securities from banks instead of from nonbank private-sector entities (such as nonbank financial firms, nonfinancial firms, households, or foreigners). While QE policy in Japan boosts the monetary base, it does not increase broad money. But it is broad money that drives nominal GDP, not the monetary base.

BOJ's QE1: 2001–06 under Governors Hayami and Fukui

The Japanese economy experienced a classic asset bubble during the period 1985–90, featuring steep rises in the prices of equities, real estate, and other assets such as golf club memberships.[2] The stock market peaked in December 1989 and the real estate market peaked in the July–September quarter of 1991. Those peaks were followed by steep declines in asset prices, culminating in contractions of real GDP in several quarters of 1992 and 1993 and deflation as measured by the CPI from July 1994.

[1]The same critique applies almost verbatim to the European Central Bank's QE policy.

[2]The underlying cause of the bubble was a sustained acceleration in the growth of the M2 money supply from 8 percent to 13 percent between 1985 and 1990, along with a parallel surge in bank lending and nonbank credit (from the *Jusen* or mortgage finance companies, and more generally the practice of *Zaitech*—credit creation and financial engineering by nonfinancial corporations). It can be argued that the cause of the acceleration in money and credit was, in turn, Japan's participation in the Plaza Agreement of September 1985 and the Louvre Accord of February 1987. These undertakings derailed domestic Japanese monetary growth, which had been remarkably stable over the preceding decade, by requiring the BOJ first to encourage a depreciation of the U.S. dollar (Plaza), and then to promote dollar appreciation by intervening in the foreign exchange market, buying dollars and creating yen, accelerating M2 growth (Louvre).

During the 1990s, numerous types of countermeasures were adopted by the Japanese authorities to combat the economic downturn, such as a series of fiscal expansion plans (see Wright 2002), a half-hearted attempt to recapitalize the banks and sell off the toxic assets from 1998,[3] cutting the BOJ's overnight policy rate to zero (ZIRP, or zero interest rate policy) by 1999, and intermittently allowing the Japanese yen to depreciate in an effort to promote export-led growth. All of these policies proved ineffective. The reason was that by September 1992, the money supply (M2) had *declined* on a year-on-year basis—unprecedented in postwar Japan—and it continued to grow only at a snail's pace, averaging just 2.5 percent per annum ever since 1992.[4] Without adequate growth of the broad money supply, nominal GDP remained in a prolonged slump, reflected in weak real GDP growth combined with persistent deflation.

After the BOJ adopted ZIRP in the spring of 1999, and after a second round of capital increases for the banks in March 1999, the economy started to perform a little better, with real GDP strengthening into 1999 Q4. Inflation remained negative but moved back toward zero. Influenced by these more favorable developments and by the arguments of crusading BOJ Policy Board member Eiko Shinotsuka that Japanese savers needed higher interest rates from the central bank, the BOJ decided to end ZIRP and increased the (targeted) uncollateralized overnight interest rate to 0.25 percent in August 2000. The rate hike was maintained until March 2001, but by then, stock prices had fallen back and a renewed economic downturn, was evident, accompanied by a recession in the United States following the bursting of the tech bubble. Faced with this further downturn, the BOJ lowered the overnight rate again and finally turned to a radical new proposal: quantitative easing.

The Policy Board statement of March 19, 2001, emphasized the extraordinary nature of its decision: "The Bank has come to a

[3]It was only after the financial crisis of 1997–98 (when one bank, a large brokerage house, and a large insurance company had failed), and almost eight years after the peak in the stock market, that the authorities first began to take serious measures to address the problem of bank capital.

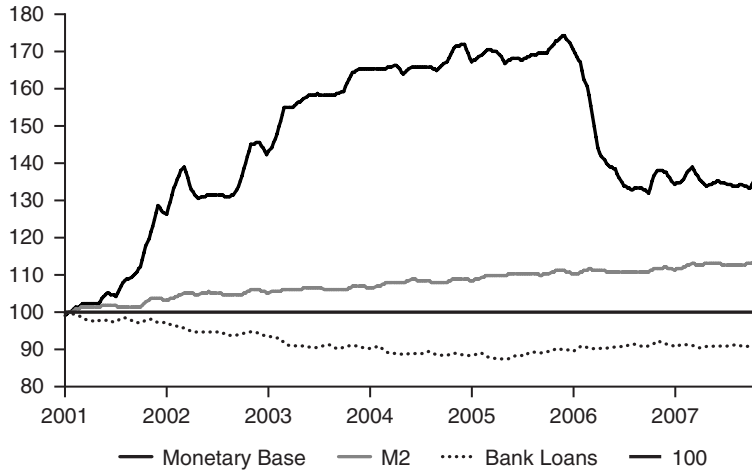
[4]Given the average annual decline of 2.3 percent per annum in M2 velocity since 1980, Japan would have required a minimum of 4 percent per annum M2 growth just to maintain price stability, or 6 percent per annum to hit a 2 percent per annum CPI target.

conclusion that the economic conditions warrant monetary easing as drastic as is unlikely to be taken under ordinary circumstances" (Bank of Japan 2001). In this first episode of QE, which started in March 2001 and ended in March 2006, the BOJ purchased a net 37 trillion yen of securities, expanding its balance sheet from 115.3 trillion yen to 152.3 trillion yen. Its purchases consisted of Japanese government bonds (JGBs) and short-dated financing bills or promissory notes (known as *tegata*).

However, as Figure 1 demonstrates, the increase in Japan's monetary base was not matched by any significant change in trend of the broad money supply, M2. Between 2001 and 2006, the monetary base expanded by 70 percent, but there was virtually no change in the trajectory of broad money (M2), while bank lending declined by close to 10 percent over the period 2001–06.

Even so, the policy started to have some success when stock prices increased by 57 percent between May 2005 and April 2006, although it should also be noted that there was a significant rally on Wall Street at the same time. The economy, too, had started to perform better, recording steady increases in real GDP during the six quarters from 2005 Q1 until 2006 Q2. Moreover, inflation returned to

FIGURE 1
Japanese Monetary Base, M2, and Bank Lending
(March 2001 = 100)



Source: Macrobond.

relative price stability. It can therefore be argued that QE1 was terminated prematurely.

Throughout the period 2001–06, the BOJ Policy Board members, with some exceptions such as the forthright Nobuyuki Nakahara, were far too timid, frequently expressing the desire to return to orthodox policies—by which they meant implementing monetary policy by adjusting short-term interest rates. Fundamentally, they failed to recognize that major balance sheet repair was needed across key sectors—households, and financial and nonfinancial corporations—and they therefore expected the economy to return to normality after only a brief interlude of unorthodox policy (Koo 2003).

In 2003–05, the BOJ had set three conditions for exiting QE, and by late 2005, these conditions were being met. Therefore, on March 9, 2006, the Policy Board decided to terminate QE, deciding to return to ZIRP while the outstanding excess reserves were reduced.

The speed with which the BOJ's balance sheet declined from April 2006 and the minimal impact this had on financial markets or the economy highlight a second aspect of the problem with the BOJ's brand of QE. Instead of buying only long-dated securities such as JGBs that would remain on the BOJ's balance sheet for an extended period, the BOJ purchased large amounts of *tegata* (short-term financing bills with maturities of less than one year), also primarily purchased from banks rather than nonbanks. Consequently, without the BOJ overtly selling any securities into the market, the maturing of these securities reduced the BOJ's balance sheet abruptly from 152 trillion yen in March 2006 to just 114 trillion yen, or by 38 trillion yen by the end of June 2006.

If BOJ-style QE purchases were supposed to have an expansionary effect on asset markets and the economy, a reduction in the BOJ's balance sheet should have led to seriously adverse or contractionary effects on the financial markets and the economy. Yet the drastic decline in the monetary base (or its counterpart, BOJ assets) between April and June 2006, unwinding the entire five-year build-up of QE within three months, had remarkably little impact on either the Japanese stock market or on the Japanese economy. Measured by the Nikkei index, equities continued to rise to a peak of 17,563 on April 7, 2006, and then declined to 14,751 by June 9, a decline of 16 percent. However, by February 23, 2007, almost a year after the end of QE1, the index had reached a new interim high of 18,188, or a rise of 23 percent. Similarly, following the termination of QE1 the

economy continued to grow until 2008 Q1, with only two negative quarters of real GDP growth before the 2008 global recession—in 2006 Q3 (followed by a strong upturn in Q4) and 2007 Q3. Inflation, too, increased to 0.9 percent year-on-year by August 2006 and remained in positive territory until February 2007. In short, the reversal of QE in April–June 2006 had none of the negative consequences that one might expect, suggesting that it had never been as stimulatory as had been intended in the first place. This would be consistent with the view expressed here that the BOJ made a policy design error in purchasing assets from banks rather than from nonbanks.

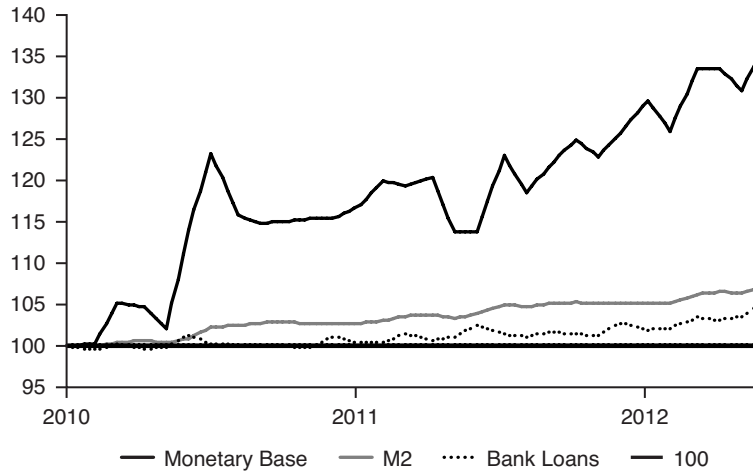
To sum up, there were three major problems with QE1 in 2001–06 under Governors Hayami and Fukui:

- First, the scale of the asset purchases was too small.
- Second, when the BOJ purchased Japanese government bonds, these were predominantly acquired from the commercial banks, not from nonbanks. This had the effect of increasing the monetary base but had a negligible impact on M2 or broad money growth.
- Third, roughly half of the BOJ's purchases were in the form of short-term financing bills held entirely by the banks. This is why, when QE1 was terminated in April 2006, the BOJ's holdings of short-term securities, including *tegata,* fell sharply.

BOJ's QE2: 2010–13 under Governor Shirakawa

Japan's second episode of QE was initiated by Governor Shirakawa and the BOJ's Policy Board in October 2010 (Figure 2). In the July–September quarter of 2010, the headline CPI and the narrow core CPI (excluding fresh food) both declined to −1.1 percent, generating political pressure on the BOJ to take stronger action to eliminate the incipient deflation. In addition, two private-sector economists with known preferences for "easy money" policies were appointed to the Policy Board (Ito 2006). The BOJ responded by setting a target for a CPI increase of 1 percent instead of the previous goal of price stability, which had generally been interpreted as a zero inflation target. Embarking on a new round of asset purchases, the BOJ's balance sheet was expanded from 121 trillion yen in October 2010 to 164 trillion yen (by 43 trillion yen, or by 35.5 percent) by

FIGURE 2
Japanese Monetary Base, M2, and Bank Lending
(October 2010 = 100)

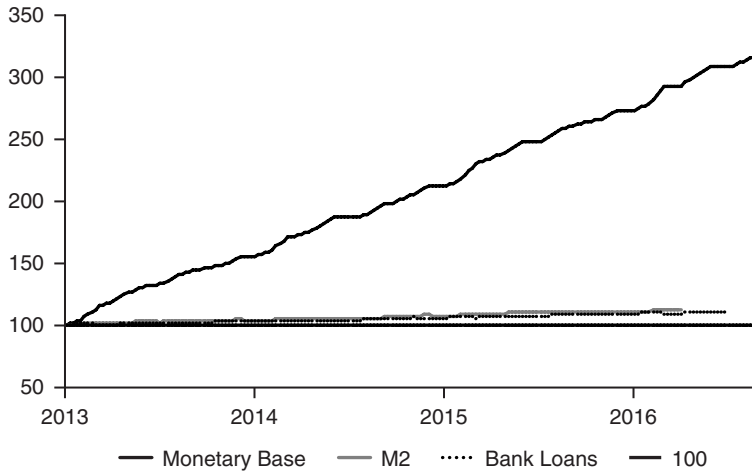


Source: Macrobond.

March 2013. The main assets purchased were once again JGBs along with *tegata*, but Tokyo-listed ETFs and REITs were added to the menu of assets purchased. Although the BOJ's holdings of JGBs and short-term financing bills (*tegata*) expanded rapidly from 77 trillion yen to 121 trillion yen—especially in 2012–13—this expansion was offset by a decline in BOJ holdings of other assets, diluting the effectiveness of the QE program.

Throughout his tenure, Shirakawa was a reluctant expansionist, frequently making speeches to the effect that QE alone would never succeed in reviving the Japanese economy. Originally, he had intended that the QE program would be completed by the end of 2011, but with the CPI still falling in November and December 2011, operations were extended through 2012 and 2013. Compared to QE1 in 2001–06, Shirakawa's QE2 was a much weaker program. In March 2013, when Shirakawa was replaced, the BOJ's balance sheet was only modestly larger than it had been in March 2006 at the end of QE1 (164 trillion yen compared with 152 trillion yen). Not surprisingly, when Shinzo Abe became prime minister in late 2012, Governor Shirakawa was replaced by Governor Kuroda.

FIGURE 3
Japanese Monetary Base, M2, and Bank Lending
(March 2013 = 100)



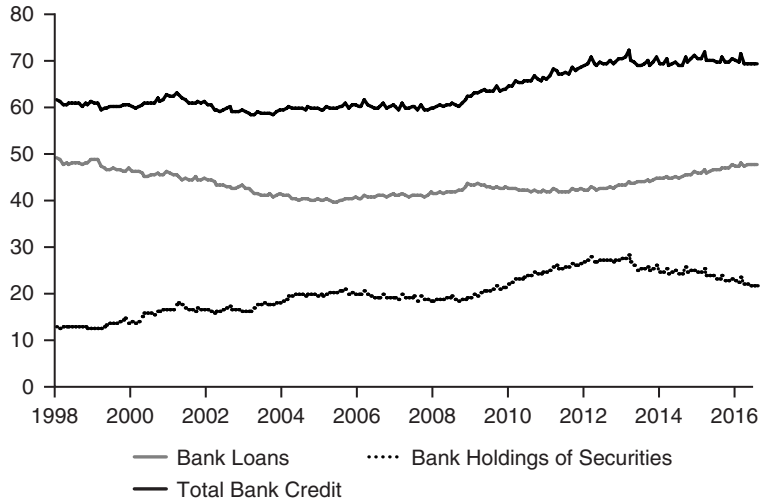
Source: Macrobond.

BOJ's QE3 or QQE: 2013 to Present under Governor Kuroda

With the resignation of Governor Shirakawa and the appointment of Governor Haruhiko Kuroda by Abe in March 2013, the BOJ began a new, more vigorous program of QE, designated "qualitative and quantitative easing." On this occasion, the extravagant promise was made that, within two years, the monetary base would be doubled and a new inflation target of 2 percent would be reached—popularly known as the 2-2-2 plan.[5] The policy has again consisted mainly of JGB purchases, but with a much more rapid expansion of the BOJ balance sheet. However, the program is once again marred by failing to purchase exclusively assets held by nonbanks. In the three and a half years since QQE started, the monetary base has trebled. However, as in QE1 and QE2, there has been almost no perceptible change in the trajectory of the M2 money supply (Figure 3), and the

[5]Under QQE, the BOJ initially set the inflation target at 2 percent for the year-on-year change in the CPI, excluding fresh food but including energy (known as the "core CPI" in Japan), but later changed the relevant index to the CPI, excluding fresh food and energy (known as the "core-core CPI").

FIGURE 4
Japanese Bank Credit (Loans and Securities)
(Yen Trillion)



Source: Macrobond.

inflation target has not been met. Perhaps surprisingly to some observers, and as a direct result of BOJ purchases of securities from banks instead of nonbanks, bank credit has actually *contracted* so far (April 2013–July 2016) as the decline in commercial bank holdings of securities (due to BOJ purchases) has exceeded the increase in their lending (Figure 4). However, as will be shown below in the case of the United States in 2008–11, increased lending is not important to the achievement of broad money growth in the initial stages of QE.

Fundamentally, despite the change of name from QE to QQE, the new version of QE is simply more of the same: larger purchases of securities (again mainly from the banks), and continued purchases of short-term securities, together with much smaller purchases of ETFs and J-REITs. Once again the monetary base is expanding rapidly but with almost no transmission through to the broad money supply. Moreover, the BOJ's underlying philosophy has hardly changed at all. In May 2015, the BOJ published a seven-page "Assessment" of QQE, which concluded that "changes in various economic and financial indicators have been in line with the mechanism anticipated," but that the policy had been derailed by the

decline in oil prices since mid-2014, which in turn had lowered inflation expectations (Bank of Japan 2015).

The BOJ view of the QQE transmission mechanism also ignores the quantitative impact on the money supply, concentrating—as most orthodox studies do—on the interest-rate effects. According to the BOJ, QQE would increase inflation expectations "through a strong and clear commitment to the price stability target of 2 percent and large-scale monetary expansion to underpin the commitment." At the same time, the BOJ's "massive purchases of Japanese government bonds (JGBs)" would exert "downward pressure . . . on the entire yield curve." The result of these actions was expected to decrease real interest rates and stimulate private demand, leading to an upturn in the economy and an improvement in the output gap. Moreover, the BOJ expected higher actual inflation, enhanced inflation expectations, higher asset prices, and portfolio rebalancing effects (Bank of Japan 2015).

The BOJ's judgment was that QQE lowered real interest rates by slightly less than 1 percentage point, and that the actual improvement in economic activity and prices was mostly in line with the mechanism anticipated by QQE. Soon after this "Assessment" was released, however, the year-on-year rate of increase in the CPI slowed abruptly due to the effects of the decline in crude oil prices, adversely affecting inflation expectations, and again delaying the BOJ's attainment of its 2 percent inflation target.

A year later, by mid-summer 2016, the BOJ was still not achieving the 2 percent inflation goal, so a further "Comprehensive Assessment" of QQE was commissioned and published in September 2016. This document makes no serious reexamination of the underlying strategy and even goes so far as to reproduce virtually the same diagram of the supposed QQE mechanism that had been published with the earlier May 2015 Assessment. This time, the Comprehensive Assessment was 65 pages in length, and the blame for not attaining the 2 percent inflation target was spread more widely. First, inflation expectations were reduced by the fall in oil prices; second, demand weakness followed the increase in the consumption tax in April 2014; and third, the slowdown in emerging economies and volatile global financial markets had lowered the observed inflation rate.

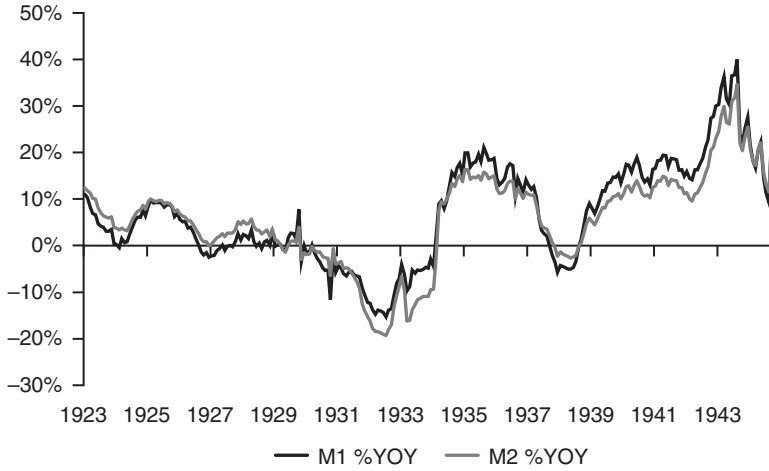
Although the monetary base is mentioned 11 times in the September 2016 Comprehensive Assessment, mainly in relation to

the formation of inflation expectations, M2 or money supply is not mentioned once in the entire document (Bank of Japan 2016). The only mention of money is "monetary expansion," but the authors are referring to the monetary base, which undoubtedly has been expanding, and not broad money, which has barely shifted from its previous trajectory.

A huge problem in this debate on the effectiveness of QE is the way that monetary policy is measured and assessed. In this article, I take the position—as I always have—that interest rates or long-term yields are not a good measure of the stance of monetary policy. Yet most of the academic literature on QE focuses on the interest-rate effects. However, in general, it is always better to assess monetary policy based on the rate of growth of a broad measure of money such as M2 or M3, sometimes including shadow banks, depending on the circumstances. By contrast, assessing monetary policy on the basis of the level of nominal (or even real) interest rates is highly problematic, as pointed out a century ago by Irving Fisher (1911). For example, high nominal rates can be either the consequence of a prolonged period of rapid growth in the quantity of money and, hence, the result of high inflation, or they could indicate the start of a period of monetary tightening. Conversely, low nominal rates can be either the consequence of a prolonged period of slow growth in the quantity of money and, hence, low inflation or even deflation (as in Japan over much of the past two decades), or they could indicate the start of a period of monetary ease.

Yet despite these drawbacks in the usefulness of interest rates as a measure of monetary policy, central banks and economists have mainly designed and assessed QE in terms of its interest-rate effects. The problem here is that central bank purchases of securities may initially lower longer-term interest rates (depending on the securities purchased), but if these purchases do not prompt a faster growth of the quantity of money, then low inflation or deflation may persist, leading to the apparent need to move to even lower or negative interest rates. It is therefore no coincidence that the two main areas experiencing subpar growth, near deflation, and negative interest rates—that is, Japan and the eurozone (plus the three euro-linked economies of Sweden, Denmark, and Switzerland)—are also the economies where the major central banks have implemented flawed versions of QE, relying mainly on the interest-rate effects of their asset purchases, not the quantitative effects.

FIGURE 5
U.S. Money Growth Rates, 1923–44



Source: Friedman and Schwartz (1963).

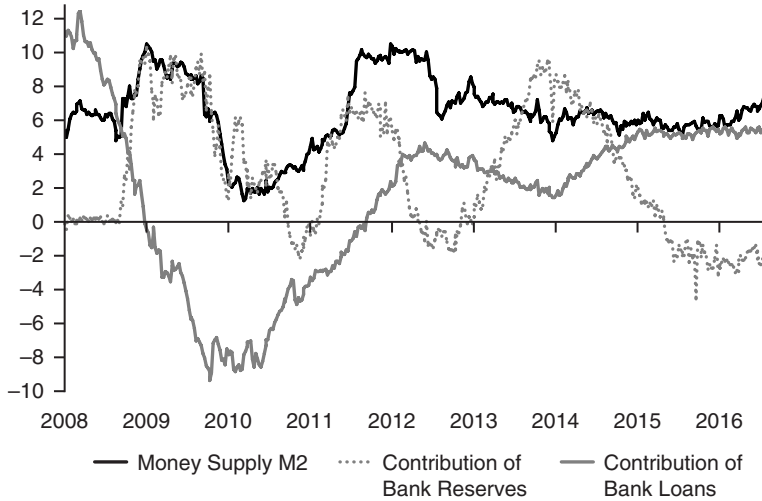
U.S. Experience with QE: 1929–33 and 2008–14

To establish the importance of the quantitative aspects of monetary policy, consider the experience of the United States in 1929–33. As shown in Figure 5, the rates of monetary growth (M1 and M2) slowed in 1929 and moved into negative territory on a year-on-year basis between November 1929 and January/February 1934. Cumulatively, the declines in the money stock amounted to 38 percent in the case of M2 and 32 percent in the case of M1. This was the basis for the Great Contraction, along with the runs on banks, in turn precipitated by concerns about their creditworthiness and the fragility of their loan portfolios (Friedman and Schwartz 1963). The runs on deposits across the country led to widespread conversions of deposits into currency, as reflected in the increases in the monetary base in 1931–32.[6]

For better or worse, the Federal Reserve did not counteract the downturn in money growth in 1931–33, either with security purchases or money creation, and thereby exacerbated the depth of the recession, the level of unemployment, and duration of the deflation.

[6]The monetary base consists of currency held by the public plus reserve deposits of banks including vault cash held by the banks. Base money did not increase due to purchases of assets by the Fed, but mostly due to conversions of deposits into currency by the public and due to decisions by the banks to hold more reserves.

FIGURE 6
U.S. Contributions to Growth of M2 (%yoy)

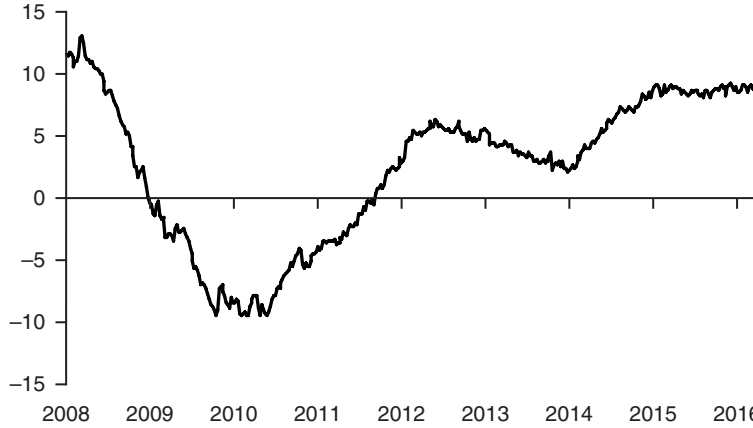


Source: Macrobond.

The key difference in 2008–11 compared with 1929–33 was that the Federal Reserve took countervailing action to buy securities from non-banks and thereby create new deposits in the banking system. The result was that a monetary contraction was avoided. Growth in bank reserves took the place of loan growth (Figure 6), allowing loans to contract while money (or deposits) continued to grow. It is true that the Fed under Ben Bernanke's leadership placed major emphasis on the interest-rate effects of its operations (whereas the Bank of England [BOE] under Mervyn King placed more emphasis on changes in the stock of money), but nonetheless it can be argued that the success of the policy came from avoiding a monetary contraction, not from lowering interest rates.

Despite central bank rates being lowered to 0–0.25 percent in the United States (and 0.5 percent in the United Kingdom in late 2008), there was essentially no appetite in the United States to lend or to borrow even at these interest rates between November 2008 and the final quarter of 2011. This is clear from the collapse in U.S. commercial banks' loans shown in Figure 7, amounting to a cumulative decline of 14 percent in the outstanding stock of bank loans over the period. Without any offsetting action, there would have been a parallel collapse in deposits and, hence, in broad money. The United States would have been on the road to repeating the disaster of 1929–33.

FIGURE 7
U.S. Commercial Bank Loans and Leases, Adjusted (%yoy)



Source: Macrobond.

In quantitative terms, the fundamental problem faced by policymakers in 2008–09 was how to avoid repeating the 1931–33 experience in the United States. In this sense, the role of QE (meaning central bank asset purchases from nonbanks) was to create new deposits in the banking system, in effect taking the place of those that were being lost through loan and securities write-offs, repayments, and deleveraging. Whether the Fed's QE in 2008–14 lowered interest rates along the yield curve or not, it certainly maintained the supply of money at a higher level than it would have been if the Fed had repeated the inaction of 1929–33.

As shown in Figure 6, the annual percentage changes in M2 can be accounted for by the contributions of three items: bank credit, bank reserves, and all other items (net)—not shown. Even after 2011, bank credit continued to grow very slowly until the start of 2014. If the authorities had done nothing, we can reasonably suppose that deposits or money supply on the other side of banks' balance sheets would have also declined by roughly the same amount—reflecting households' and nonbank financial institutions' unwinding of leverage, repayment of loans, and repair of balance sheets. All this would have deepened the recession, raised unemployment further, and intensified the deflation. Instead, M2 growth

averaged 6.8 percent per annum between January 2009 and October 2013, helping to alleviate the contractionary pressures.

The role of QE1–QE3 in the United States was therefore to ensure that the money supply did not contract in line with bank credit—as it had done in 1929–33. In this respect, the policy was successful.

Fortuitously, as the tapering of QE purchases began in December 2013, the rate of growth of bank lending began to accelerate. Consequently, although the contribution of banks' reserves to M2 growth began to decline, by now bank credit was growing more vigorously, meaning that banks were once again creating credit independently of the Fed. In short, the Fed had successfully handed the money creation baton over to the commercial banks.

From the above sketches of Japanese and U.S. experience, we can derive two rules for central banks to follow when designing a QE program:

- First, the central bank should only buy securities from nonbanks. The reason is that the primary purpose of doing QE is—or should be—to expand the money supply. If the central bank buys securities from banks, there can be no assurance that the money supply will increase. Also, if banks create new credit, leverage in the private sector is not reduced, but will increase *pari passu* with the creation of new loans. However, if the central bank buys securities from nonbanks, this guarantees that new deposits will be created, expanding the money supply, without adding to leverage.
- Second, the central bank should buy only long-term securities. This is only partly to bring down yields at the longer end of the curve (flattening the yield curve). More importantly, it means the central bank's portfolio and the stock of new deposits or money in the banking system is not eroded by the maturing or running down of its holdings. As a result, the volume of deposits created or funds injected into the economy can remain stable for a long period of time.[7]

[7]As is well known, the Fed made the mistake of buying short-term securities (Treasury Bills) during QE2, and was therefore obliged to undertake a lengthy maturity extension program between September 2011 and December 2012 amounting to $667 billion. The scheme was popularly known as "Operation Twist" because the Fed was selling short-term securities while buying longer-term securities, "twisting" the yield curve. In the case of Japan, the BOJ has continued to buy short-term securities throughout QQE.

The next section will show that the Bank of Japan has repeatedly broken both these rules. By contrast, when the Bank of England announced its QE program in February 2009, it undertook explicitly to purchase gilts with longer-term maturities (10–15 years), which U.K. banks tend not to hold due to the high capital risk, precisely so that these purchases would be from nonbanks. "The aim of the policy was to inject money into the economy in order to boost nominal spending and thus help achieve the 2 percent inflation target" (Bank of England 2011: 201). By buying securities primarily from nonbanks, the Bank of England guaranteed the success of its program.

Two Brands of QE, and Japan's Choice

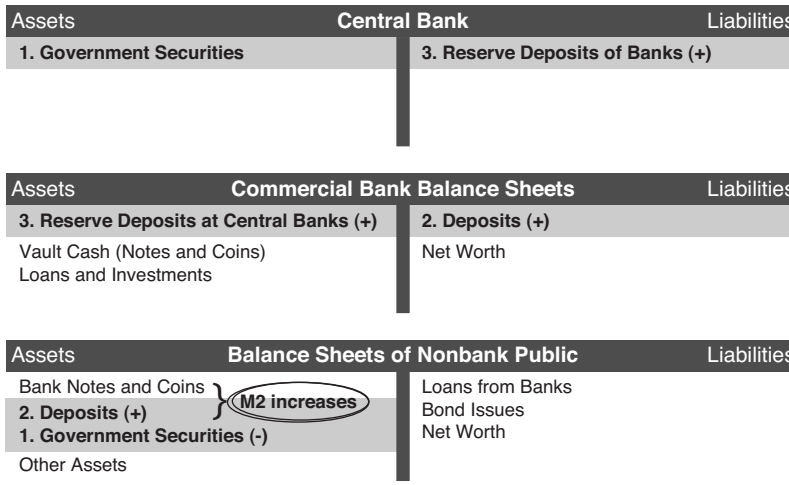
To explain the difference between the Fed or Bank of England operations on the one hand and the BOJ or ECB operations on the other, it is helpful to review the impact of their transactions on the balance sheets of the banks and the nonbank public. Figure 8 sets out a series of paired transactions (1–3), demonstrating the impact of Fed or BOE-style asset purchases:

1. The central bank purchases government securities from nonbank entities. Nonbank entities (e.g., insurance companies, pension funds, asset managers, or foreigners) sell government securities to the central bank.
2. The sellers receive new deposits from the central bank in settlement of their sale, which expands the money supply. The sellers deposit their newly acquired funds in commercial bank deposit accounts.
3. The banks deposit the payment drafts they receive from the sellers of government securities with the central bank. Banks' holdings of deposits or reserves at the central bank are increased by an amount that exactly matches the central bank's initial purchases.

Note that after these transactions, both sides of the central and commercial banks' balance sheets have expanded, with increases in assets matched by increases in liabilities. However, at this stage, the balance sheet of the nonbank public has not increased—it has simply become more liquid, as government securities are replaced with new deposits.

There are two key points about this series of transactions. First, money (M2, M3, or M4 depending on the local definition) in the

FIGURE 8
A Well-Designed Asset Purchase Plan
(Effective QE)



hands of firms and households has now increased, and, given that interest rates are at the zero bound, the holders will almost certainly wish to adjust their asset holdings by buying new investments, kick-starting the portfolio rebalancing effect, or shift their mix of consumption and savings. Second, since deposits have increased without an accompanying increase in bank lending, nonbank private-sector leverage will have declined.

Next, consider the effects of another type of central bank transaction such as those conducted by the BOJ—or the ECB under its QE program.[8] Figure 9 shows the corresponding series of paired transactions (1–2) in the T-form balance sheets:

1. The central bank buys securities from the commercial banks. Bank holdings of JGBs or *tegata* decline; BOJ holdings increase.
2. Commercial banks receive a credit from the BOJ for their sale of their JGBs or *tegata*; reserve deposits of banks at the BOJ increase.

[8]The ECB's LTRO and targeted-LTRO programs also involved transactions only with the banks, not with the nonbank public. The result was that banks accepted the ECB's new, cheaper funding, but did not increase total bank credit or deposits. Consequently, there was no impact on eurozone M3.

FIGURE 9
A Poorly Designed Asset Purchase Plan
(Ineffective QE)

Assets	Central Bank	Liabilities
1. Government Securities		**2. Reserve Deposits of Banks (+)**
Loans to Banks		
Foreign Assets		

Assets	Commercial Bank Balance Sheets	Liabilities
2. Reserve Deposits at Central Bank (+)		Deposits
Vault Cash (Notes and Coins)		Net Worth
1. Reserve Deposits at Central Banks (+)		
Loans and Investments		

Assets	Balance Sheets of Nonbank Public	Liabilities
Bank Notes and Coins	**no change in M2**	Loans from Banks
Deposits		Bonds Issued
Other Assets		Net Worth

Note that after these transactions, both sides of the central bank's balance sheet have expanded, with increases in BOJ assets matched by increases in liabilities. But commercial banks' balance sheets have not expanded; they have merely undertaken an asset swap, holding less JGBs or *tegata* but more deposits at the BOJ. In this version of QE, the balance sheets of the nonbanks are unaffected. The key point is that the money holdings (M2, M3, or M4) in the hands of the nonbank public have not increased.

Moreover, in this BOJ brand of QE, given the starting point of risk aversion by the banks and the reluctance to borrow by bank customers, there can be no assurance that the banks will expand their lending and therefore no assurance that deposits (or money) will expand either. Equally, portfolio rebalancing and new investment or consumption spending is unlikely to follow. Even if banks were to expand their lending, this would be accompanied by a parallel increase in leverage by firms or households—the opposite of the balance sheet repair process that policymakers should be seeking to achieve.

In short, only purchases of securities from nonbanks would be consistent with both balance sheet repair and enhanced liquidity (broad money) in the hands of firms and households.

A secondary problem—which continues to this day—is that the BOJ dilutes the effectiveness of its purchases by buying substantial amounts of short-term securities, also from the banks. Since these securities (MOF financing bills or promissory notes) mature within a few weeks or months, the BOJ needs to replace the maturing securities with further new purchases (which count toward its announced monthly totals) in order to maintain the size of its balance sheet or the monetary base. Inevitably, the maturing of these securities erodes the impact of the BOJ's overall asset purchase operations on M2.[9]

BOJ purchases under the QQE programme can be split into three elements: JGBs, T-Bills and other promissory notes such as commercial paper (or *tegata*), and other securities. The latter, which are collectively small, consist of corporate bonds, equity shares in the form of ETFs, and investment trusts in the form of J-REITs (Bank of Japan 2014).

Initially, from April 2013, the BOJ purchased assets to increase the monetary base at a rate of 60 trillion yen per year, but this pace was accelerated to 80 trillion yen per year from November 2014. The BOJ purchased JGBs initially at the rate of 50 trillion yen per year, but this was similarly accelerated to about 80 trillion yen per year from November 2014, while the pace of ETF and J-REIT purchases was also raised to 3 trillion yen and 90 billion yen, respectively. However, purchases of commercial paper and corporate bonds were not raised as it was decided to maintain the outstanding balances on the BOJ's balance sheet.

In sum, the Bank of Japan has repeatedly broken both the two rules set out earlier for the design of a successful QE program by (1) buying securities mainly from banks, and (2) buying significant quantities of short-term securities that require frequent replacement on the BOJ's balance sheet and therefore erode the effectiveness of its total purchases.

[9]For example, in the fiscal year ended March 2015, while purchases of JGBs amounted to 96.6 trillion yen and largely remained on the BOJ's balance sheet, purchases of short-term securities amounted to 101.8 trillion yen, but only showed up as an outstanding balance of 49.7 trillion yen due to persistent maturities and rollovers.

Conclusion

One way to contrast the two brands of QE is to consider the role of the monetary transmission mechanism in each version. On the one hand, the Fed and the BOE implicitly acknowledged that the traditional transmission mechanism for monetary policy was broken and therefore operated a brand of QE that circumvented the banking system by creating new deposits independently of the banks. By contrast, the BOJ is attempting to implement a monetary expansion through a banking system where the traditional transmission mechanism of monetary policy is not working. Normally, when banks lend to nonbank customers in the private sector, they expand deposits and, hence, the level of broad money in the economy, so that loans and deposits (or money) grow roughly in parallel. However, in the current circumstances of the Japanese economy, banks are reluctant to lend and firms and households are similarly reluctant to borrow from banks.

The result of this standoff is stagnation in the growth of broad money and a parallel weakness in the growth of nominal GDP. The deadlock could easily be broken if the BOJ were to (1) purchase most of its securities from nonbanks, and (2) ensure that its purchases consisted entirely of long-term securities. In so doing, it would directly create new deposits in the hands of nonbanks, thereby overcoming the banking system's reluctance to create new loans and deposits. Moreover, such a revision of QE design would be consistent with a measure of desirable deleveraging in the economy. As long as the nonbank private sector is not actively deleveraging (i.e., repaying loans, thus destroying deposits or money balances), broad money (or M2 plus CDs in Japan) would grow *pari passu* with the scale of the BOJ's purchases of securities.

Analysis by the BOJ (in May 2015 and in September 2016) nevertheless claimed that the program was working out largely as expected. However, the BOJ analysis focuses almost exclusively on the interest-rate effects of QE and the effect of QE on inflation expectations; it largely excludes any consideration of banks' balance sheets or the stock of money in the hands of households or nonbank companies.

It is no coincidence that in both Japan and the euroarea, economic growth is subpar, the economies are close to deflation, and both have

negative interest rates. The common source of these problems is the obsession with using interest rates as the primary tool of monetary policy, while failing to conduct QE in such a way as to create new deposit balances in the hands of firms and households, thereby overcoming the inherent reluctance of banks to lend and firms and consumers to borrow. In both regions, the monetary base is expanding, but without any significant effect on broad money.

As a consequence of the failure of Japan's QQE to gain traction, economic activity in Japan is likely to remain sluggish, while core CPI is unlikely to reach the 2 percent target and may even revert to deflation if energy prices do not rise significantly in the remainder of 2016 or in 2017.

Finally, this critique of QE policies currently implemented in Japan and the eurozone can be taken further. Some economists have claimed that even in the United States, new regulations "are constraining the supply of loans" (Calomiris 2016), but the truth is that U.S. bank loans and leases (in the H8 Release of the Federal Reserve Board) have increased at an average annual rate of 7.5 percent since mid-2014 when QE was terminated—surely an adequate growth rate for an economy with a potential real growth rate of just 2 percent per annum. In addition, a more recent fashion has been to assert that monetary policy is reaching some kind of limit in terms of what it can be expected to achieve. These commentators have therefore argued that governments and central banks should either engage in the provision of "helicopter money," or, alternatively, that since the limits of monetary policy have been reached, now is the time to engage in fiscal expansion. However, the frustration of economists and policymakers with monetary policy is not due to the failure of QE per se, but rather to the particular design of QE selected.

This article has shown that if QE is done correctly—following the brand implemented by the Fed or the Bank of England in 2008–14—then the nonbank private sector will be reliquefied in a manner that is consistent with balance sheet repair, the money supply will resume normal growth rates, and nominal GDP will recover. On the basis that QE in Japan has not been implemented in a way that would have been compatible with these objectives, there is no case for either "helicopter money" or further fiscal expansion. It would be better to redesign QE and implement it according to the U.S.–U.K. template.

References

Bank of England (2011) *Quarterly Bulletin* (2011 Q3).

Bank of Japan (2001) *Monetary Policy Statement* (March 19): www.boj.or.jp/en/announcements/release_2001/k010319a.htm.

_____ (2014) "Expansion of the Quantitative and Qualitative Monetary Easing." *Monetary Policy Statement* (October 31): www.boj.or.jp/en/announcements/release_2014/k141031a.pdf.

_____ (2015) "Quantitative and Qualitative Monetary Easing: Assessment of Its Effects in the Two Years since Its Introduction." *Bank of Japan Review* (May): www.boj.or.jp/en/research/wps_rev/rev_2015/data/rev15e03.pdf.

_____ (2016) "Comprehensive Assessment: Developments in Economic Activity and Prices as Well as Policy Effects since the Introduction of Quantitative and Qualitative Easing (QQE)." Press Release (September 21): www.boj.or.jp/en/announcements/release_2016/k160921b.pdf.

Calomiris, C. (2016) "The Microeconomic Perils of Monetary Policy Experiments." *Cato Journal* 37 (1): 1–15.

Fisher, I. (1911) *The Purchasing Power of Money: Its Determination and Relation to Credit, Interest, and Crises*. New York: Macmillan.

Friedman, M., and Schwartz, A. (1963) *A Monetary History of the United States: 1867–1960.* Princeton: Princeton University Press for NBER.

Ito, T. (2006) "Japanese Monetary Policy: 1998–2005 and Beyond." In *Monetary Policy in Asia: Approaches and Implementation*, 105–32. BIS Paper No. 31. Basel: Bank for International Settlements.

Koo, R. (2003) *Balance Sheet Recession: Japan's Struggle with Uncharted Economics and Its Global Implications*. Singapore: John Wiley.

Wright, M. (2002) *Japan's Fiscal Crisis: the Ministry of Finance and the Politics of Public Spending 1975–2000*. New York: Oxford University Press.

Why Have the Fed's Policies Failed to Stimulate the Economy?

Mickey D. Levy

The Fed's sustained low policy rate, quantitative easing (QE), and forward guidance have stimulated financial markets and boosted asset prices but have failed to stimulate the economy. As planned, the Fed's efforts to lower bond yields and reduce the real cost of capital, encourage risk taking, and lift stock and real estate values have succeeded. But nominal GDP growth has actually decelerated to 2.5 percent in the last year from its subdued 3.9 percent average pace of the prior six years, and real growth has languished.

The most disappointing aspect of the slow economic expansion has been the weak rise in business investment. Consumption and residential investment have grown fairly steadily. But despite lower costs of capital and only modest increases in labor costs, investment has fallen persistently below expectations while employment gains have actually been strong. Labor productivity has risen at a painfully slow 0.5 percent pace in the last six years and has fallen in the last three quarters, a unique trend during modern economic expansions. These trends have far-reaching implications.

In response, estimates of potential growth and the natural rate of interest have been reduced sharply. The Fed has delayed normalizing rates, and bond yields hover near historic lows. Even with

---

Cato Journal, Vol. 37, No. 1 (Winter 2017).

Mickey D. Levy is Chief Economist for the Americas and Asia at Berenberg Capital Markets, LLC. The views expressed in this article are the author's and do not reflect those of Berenberg Capital Markets, LLC. An earlier version of this article was presented at the Shadow Open Market Committee Meeting, October 7, 2016, at the Princeton Club in New York City.

mounting evidence that monetary policy is having little stimulative impact on the economy, a constant Fed theme has been that as long as inflation is below its longer-run 2 percent target and inflationary expectations remain well anchored, sustaining monetary ease is appropriate. This theme presumes that the economy is constrained by insufficient demand that may be remedied by monetary policy. Until recently, very few Fed members have challenged that assessment.

Recent trends make it increasingly clear that economic performance has been constrained by factors that are beyond the scope of monetary policy and that the Fed's policies are contributing to mounting financial distortions with unknown consequences. Such policies are inconsistent with the Fed's macroprudential risk objectives—a point emphasized by Peter Fisher (2016).

Factors Constraining Investment and Growth

Standard explanations of weak investment are that business capital spending has been slowed by the rising share of GDP in less capital-intensive production, particularly in some labor-intensive services, rising investment overseas (that is not measured in GDP), and measurement issues. The largest U.S. companies based on market capitalization are investing less in traditional physical capital than the largest companies in prior decades. Measurement problems center on the challenge of fully capturing information technology, human capital, and intangible capital in the National Income and Product Accounts. These factors likely explain part of the weakness in measured domestic investment and GDP.

Government policies have been a key source of the weak investment and economic growth. The negative impact of the rising public debt overhang and expectations of future tax increases on potential growth has been widely discussed (Reinhart, Reinhart, and Rogoff 2012). In addition, a growing web of government regulations (at the federal, state, and local levels), mandated expenses, and higher tax burdens have weighed on the investment environment and constrained growth. Higher tax burdens have stemmed from administrative rulings as well as rate increases. Considered separately, the vast majority of these policies has had little impact on macroeconomic performance. However, they combine to increase aggregate operating costs, distort production processes and labor inputs, and lower after-tax rates of return on investment. Anecdotal evidence and

business survey results reflect these negative impacts, but their cumulative effects are not captured in standard macro models, including the Fed's FRB-US model. They also influence household spending and financial decisions.

While the Fed's monetary policies have lowered real costs of capital, the governments' economic policies and expectations of higher taxes, more mandated expenses, and additional regulatory burdens in the future lower expected rates of return on investment and add a layer of uncertainty. Businesses considering a five- or seven-year investment project certainly take into account the possibility of higher taxes and regulatory burdens over the duration of the investment project. These expectations raise the "hurdle rates" on business investment decisions and offset the benefits of the lower costs of capital. Consequently, businesses take a more conservative approach to investment spending: replacing aged equipment and software, while trimming some large expansion plans. Businesses also have an incentive to expand overseas and purchase foreign firms for tax reduction purposes ("tax inversions").

In response to some government-mandated expenses and labor laws, businesses are changing labor inputs and relying more on part-time workers. With less investment spending, businesses also invest less in training employees on how to use new capital. This reduction in capital and on-the-job training contributes to weaker labor productivity.

Businesses are taking advantage of the Fed-induced low yields to issue bonds but are using the proceeds to buy back shares to meet the demands of yield-hungry investors. The rising corporate leverage and higher cash distributions to stockholders are efficient from a corporate perspective but result in less investment and lower potential growth.

Household behavior is also affected. Dimmed expectations of future disposable incomes have led to more precautionary saving, and real consumption has not kept pace with real disposable incomes. Households are allocating more out-of-pocket spending to medical care and health insurance, in part due to the Affordable Care Act, so they have less to spend on other goods and services. Tight mortgage credit standards and more onerous administrative costs have constrained mortgage originations.

The negative economic impacts of government policies are structural and beyond the scope of the Fed to remedy through

monetary policy. Potential growth has been constrained. Prior to 2008–09, potential growth was estimated to be approximately 2.6 percent but is now closer to 2 percent—the Fed's latest estimate is 1.8 percent. This is a dramatic shift, with a huge cumulative economic impact.

Throughout most of this expansion, the Fed has argued that the weak economic growth and labor market underperformance have been due to insufficient demand. This position has supported the view that aggressive monetary ease is needed to stimulate the economy. The Fed has frequently argued that had it not pursued aggressive monetary ease, economic performance would have been much worse. That argument may be appropriate in describing the success of the Fed's alternative liquidity facilities and the first round of QE, but it grossly overstates the efficacy of monetary policy in recent years.

In the nearly six years since the Fed initiated QE2 (followed by "calendar-based" forward guidance, Operation Twist, QE3, and the use of various moving targets to signal its wish to sustain the negative real Fed funds rate), the deceleration of GDP growth and subdued business investment highlight the nonmonetary nature of the disappointing economic performance that is beyond the Fed's ability to influence. The Fed's assertion that its monetary policy has generated several million new jobs during this period is implausible.

Noteworthy, former Fed Chair Ben Bernanke stated in a recent blog that there may be supply constraints that are inhibiting economic growth, and if so, the Fed cannot do anything about it (Bernanke 2016). This is an important acknowledgment by the influential former Fed Chair. Leading economic media may also be acknowledging the Fed's limitations; witness a recent *Wall Street Journal* front page headline: "Central Bank Tools Losing Their Edge" (Ip 2016).

Recently, the Fed's view has evolved toward a growing perception that its monetary policy is having a diminishing economic impact. At the same time, some Fed members are expressing concerns about mounting financial distortions. The three official dissents at the September 2016 FOMC meeting and votes of 8 of 12 Federal Reserve District Banks recommending an increase in the discount rate in July reflect the increasing unease at the Fed about its current policy. (The Board of Governors voted to keep the discount rate unchanged.)

A Lower Natural Rate of Interest

The natural rate of interest has fallen as expected rates of return have declined and potential growth has been reduced by weak investment and productivity. These trends have heightened pessimism and lowered expectations about the future. In describing the fundamental linkages between economic performance, the time preferences of households, and the natural rate, Marvin Goodfriend (2016) emphasizes the important role of taxes, regulations, markups, and other distortions underlying the pessimism about future incomes that have driven down the natural rate. Although persistently poor performance and diminished expectations about future incomes have lowered the natural rate of interest, nobody knows with any precision what the natural rate really is. The Fed's projections imply a natural rate of 0.9 percent (the median FOMC member projects a 2.9 percent longer-run Fed funds rate and the Fed's inflation target is 2 percent). Presumably, the Fed's estimate of potential growth of 1.8 percent is consistent with a positive real rate of return on capital and a positive natural rate of interest.

This implies that the Fed's monetary policy is very accommodative, with its current real policy rate of minus 1 percent and its extremely large balance sheet. Such policy is inconsistent with the Fed achieving its dual mandate.

What Should the Fed Do?

The Fed should commence raising rates toward a neutral rate consistent with its estimates of potential growth and its 2 percent inflation target and shift the focus of its communications to emphasize how monetary policy is limited in its ability to achieve the Fed's dual mandate while deemphasizing short-run economic and market concerns. The Fed must cease altering policy in response to global and financial turmoil that does not materially influence the U.S. economy and make clear that volatility is a normal characteristic of financial markets. The Fed's effort to be transparent must involve articulating how economic performance is influenced by other policies and real factors that are beyond the Fed's scope.

Gradually raising rates would leave the Fed's easy monetary policy intact, maintain a negative real Fed funds rate and plentiful excess bank reserves, and would not harm economic performance. History shows clearly that during economic expansions when the

Fed raises rates from an accommodative stance of monetary policy, growth is sustained. A clear Fed explanation of why it is normalizing rates—and why it is no longer delaying raising rates—would boost confidence. The Fed projects that real GDP will grow at a 2 percent pace through 2018, slightly above its estimate of potential growth, even as it raises rates along the Fed's estimated appropriate path (the so-called dots in the Fed's official September projections). The Fed must align its monetary policy with its forecasts.

In addition, the Fed should stop using its bloated balance sheet as a forward guidance signaling device and cease reinvesting the proceeds of maturing assets. Allowing for a very gradual unwinding of excess reserves without any outright sales would have no impact on credit supply. The low bond yields that have resulted from the Fed's forward guidance have not stimulated capital spending or the economy, and the excess reserves that exceed $2.5 trillion only add to financial distortions. Any modest increase in bond yields from current very low levels would have negligible economic impact.

A clearer explanation by the Fed of the nonmonetary policies and factors that have contributed to lower potential growth, weak capital spending and productivity, and structural unemployment would help steer the policy debate toward the issues that really matter for economic performance. The Fed needs to correct the misperceptions that monetary policy is capable of managing every aspect of economic performance and that activist monetary policy is necessary because the government's economic and fiscal policy processes are dysfunctional. Monetary ease cannot offset or cover up for misguided tax, spending, and regulatory policies. The Fed should also spell out clearly how its easy monetary policies influence federal budget and fiscal policies.

Such clarity may not sit well with Congress, which has come to rely excessively on the Fed, but it would reset monetary policy and enhance the Fed's credibility.

References

Bernanke, B. S. (2016) "The Fed's Shifting Perspective on the Economy and Its Implications for Monetary Policy." Brookings Institution (August 8).

Fisher, P. (2016) "What's the Matter with the Fed?" Speech given at the Shadow Open Market Committee meeting (April 29):

http://shadowfed.org/wp-content/uploads/2016/04/FisherSOMC-April2016.pdf.

Goodfriend, M. (2016) "The Case for Unencumbering Interest Rate Policy at the Zero Bound." Paper presented at the Federal Reserve Bank of Kansas City Jackson Hole Policy Symposium (August 26). Available at www.kansascityfed.org.

Ip, G. (2016) "Central Bank Tools Losing Their Edge." *Wall Street Journal* (September 22).

Reinhart, C.; Reinhart, V.; and Rogoff, K. (2012) "Public Debt Overhangs: Advanced-Economy Episodes since 1800." *Journal of Economic Perspectives* 26 (3): 69–86.

The Fiscal Theory of the Price Level: A Comment on Tutino and Zarazaga

James A. Dorn

In their essay, "Inflation Is Not Always and Everywhere a Monetary Phenomenon" (*Economic Letter*, Federal Reserve Bank of Dallas, June 2014), Antonella Tutino and Carlos E. J. M. Zarazaga question Milton Friedman's famous dictum that "inflation is always and everywhere a monetary phenomenon" (Friedman 1970: 11). In doing so, they rely on the strong version of the fiscal theory of the price level (FTPL) as proposed by Christopher Sims (1994), which holds that "fiscal policy affects the price level and the path of inflation *independent* of monetary policy" (Carlstrom and Fuerst 2000: 23; emphasis added).

Tutino and Zarazaga (2014: 3) note that, given the strong assumptions of some FTPL models, hyperinflation can emerge when it is expected *"even if the money supply is kept constant."* That expectation results in an explosive rise in the velocity of money without any change in the money supply (see McCallum and Nelson 2005). The strong version of FTPL contradicts Phillip Cagan's monetary theory of hyperinflation, which holds that "variations in real cash balances mainly depend on variations in the expected rate of change in prices"—which, in turn, depends on "a dynamic process in which

---

Cato Journal, Vol. 37, No. 1 (Winter 2017).

James A. Dorn is Vice President for Monetary Studies and a Senior Fellow with the Center for Monetary and Financial Alternatives (CMFA) at the Cato Institute. He thanks Thomas M. Humphrey and Ari Blask for comments on an earlier version of this article, which appeared on CMFA's blog, *Alt-M* (Dorn 2016).

current price movements reflect past and current changes in the quantity of money" (Cagan 1956: 27).

Current price movements also reflect expected future changes in the quantity of money (rational expectations). If individuals, from past experience or knowledge of past inflations, think monetary authorities will print money in the future to pay for government deficits, and lower the real value of public debt, they will increase their rate of spending today—before the inflation tax decreases the real value of their cash balances. Doing so will increase monetary velocity and cause inflation to exceed the current rate of growth of the money supply. This scenario is fully consistent with the quantity theory of money, even though Tutino and Zarazaga (p. 1) claim that the German hyperinflation of 1921–23, in which the rise in the general level of prices outpaced money growth, suggests that "something is wrong" with Friedman's dictum.

In this article, I examine the strong version of FTPL and contrast it with the weak version, which holds that fiscal policy drives monetary policy, which is assumed to be passive. The fiscal authority's deficit spending, however, cannot by itself cause a sustained rise in the price level unless accompanied by expansionary monetary policy—that is, monetization of the debt.

The key point that Tutino and Zarazaga (hereafter, TZ) make is that "hyperinflation is fiscal in nature because it can only happen if the fiscal authority—the central government—remains on the sidelines" (p. 3). They turn to the German hyperinflation in the 1920s for support of their argument, holding that the government ended runaway inflation by implementing "an active fiscal policy." In particular, TZ argue that it was the backing of the rentenmark by real estate revenues that ended the hyperinflation. The crux of their argument is that it was "the government's ability to raise revenues from the real estate market . . . [that] successfully broke the link between mutually reinforcing lower fiscal revenues—implying higher fiscal deficits—and rising price levels" (pp. 3–4). I examine this argument by taking a close look at the German hyperinflation and stabilization. Evidence does not support the strong version of FTPL: fiscal policy cannot explain either the hyperinflation or the stabilization of the German currency. Passive fiscal policy did not usher in the hyperinflation, and activist fiscal policy did not end it. The article concludes by noting the importance of a proper understanding of monetary history in evaluating macroeconomic models such as FTPL.

Fiscal Theory of the Price Level

As noted, there are two versions of the fiscal theory of the price level: the weak version and the strong version.[1] The weak version holds that if the fiscal authority dominates the policy space, then fiscal deficits could be monetized by the central bank. This version is consistent with the quantity theory of money because inflation is ultimately determined by excess growth in the money supply. The second version of FTPL, the so-called strong version, holds that *even if the money supply is held constant*, inflation can occur if the fiscal authority is passive. All that is needed is for the public to expect prices to rise. People will then spend their given money balances at a faster rate—increasing the velocity of money—and prices will rise until expectations change.

Tutino and Zarazaga (2014: 3) note that the strong models of the FTPL can "give rise to hyperinflation quite easily, *even if [the] money supply is kept constant,*" because "nothing in the internal logic of these models anchors the evolution of inflation." Rather, "the dynamics of inflation are entirely determined by household expectations." Thus, "if households anticipate ever-rising inflation, they will try to get rid of their money balances and exchange them for goods. The resulting increase in demand for goods accelerates inflation even further." The authors conclude: "This hyperinflationary process cannot be categorized as 'monetary' in the usual sense, because that would have required an equally explosive expansion of [the] money supply, which was kept constant."

This analysis ignores the reality that if the quantity of money is constant and people spend more of it on some goods, there will be less of it to spend on other goods, so the overall price level can't spiral upward. However, it is possible that if households expect the fiscal authority to cooperate with the monetary authority, and thus expect future deficits to be funded by printing money, then prices in general could begin to rise *before* the actual increases in future money growth. But as Tutino and Zarazaga (2014: 4, n. 4) point out, the assumption of a constant money stock "rules out the possibility of inflation arising from monetization of fiscal deficits."

[1]For a more formal discussion of the weak and strong versions of the FTPL, see Carlstrom and Fuerst (2000).

The strong version of FTPL also implies that fiscal action—not monetary reform—is the primary tool for ending a hyperinflation. Yet historical evidence shows that the determining factor in generating hyperinflation is explosive growth in the money supply (or the *expectation* that such growth will occur), and that stabilization primarily stems from *credible monetary reform*. One notable example is the German hyperinflation of 1921–23 and the rapid stabilization that ended runaway inflation. We shall see that it was not fiscal policy—but rather monetary policy—that enabled the rapid rise in the price level and abruptly ended it.

The German Hyperinflation

The problem with TZ's argument that the German hyperinflation was ended by fiscal measures (namely, backing the rentenmark by real estate revenues) is that it ignores the fact that the mortgage-backing of the rentenmark was not sufficient to change expectations of further inflation, although it did help the public accept the new currency. Expectations changed because the public knew there was a *legal limit* on the total value of rentenmarks that could be issued by the Rentenbank, which was under the jurisdiction of the Reichsbank (the central bank). The backing of the currency by real estate was not relevant for stabilizing prices. There was no official convertibility between the inflated paper marks and the rentenmark, and the latter was not legal tender. The rentenmark was a parallel currency, added to the circulation of existing paper marks (see Bresciani-Turroni [1931] 1953: 334–37).

It is true that 500 rentenmarks could be converted into a bond with a nominal value of 500 gold marks, "which was guaranteed by a legal mortgage on German property and which yielded a rate of interest at 5 percent *in gold* (actually payable in paper at the exchange rate of the gold mark)," but as Bresciani-Turroni (p. 340) points out, "the stability of the value of the rentenmark could not be due to the possibility of converting the latter into mortgage securities." The reason is simple:

> The market value of the mortgage bonds was lower than the nominal value. The market rate of interest was then much higher than 5 percent. . . . Besides, the increase of the issues of rentenmarks would continually add to the Government's burden on interest on mortgage bonds, for which the public

> would exchange increasing quantities of rentenmarks; and therefore, in a precarious financial position, the uncertainty of the Government being able to continue the payment of interest would increase [ibid.]

Confidence in the rentenmark stemmed, in part, from the fact that it was a new currency and the public "believed in the efficacy of the mortgage guarantee." But, according to Bresciani-Turroni (p. 348), that confidence "would have been quickly dissipated if the public had been led to expect that, despite the obligation imposed on the rentenbank by decree, the Government would exceed the pre-arranged limit to the issues." There was an attempt to circumvent the legal limit on the issuance of rentenmarks in December 1923. However, as Bresciani-Turroni observes, that attempt "was confronted by a determined refusal by the management of the Rentenbank," which "helped to strengthen confidence in the new money. The limitation of the quantity was then of primary and fundamental importance" (ibid.) Thus, in contrast to TZ, Bresciani-Turroni emphasizes the importance of monetary policy—not fiscal policy (the expected revenue from mortgage securities)—in ending the German hyperinflation.

The Dynamic Theory of Money

The assumption in the strong version of FTPL that the money supply is constant, abstracts from the reality of what actually occurred to bring about Germany's hyperinflation (from June 1921 until January 1924) and the rapid stabilization of the currency. In doing so, it also ignores the dynamics of the quantity theory of money.

The dynamic theory of money—also known as "the theory of monetary disequilibrium"—holds that large increases in the quantity of money relative to the trend rate of real output depreciate the value of money and lead to a subsequent rise in the velocity of money, which accentuates the rise in prices, further reducing the real money stock (Warburton 1966: 4–5; also see Dorn 1987 and Yeager 1997). This inflationary spiral will continue until the monetary authority changes expectations by adopting fundamental monetary reform that ends excessive money creation. That is what happened during the German hyperinflation.[2]

[2]See Humphrey (1980: 4) for a summary of the dynamic theory of money as it operated in the Weimar Republic.

It is important to recognize that the rentenmark did not begin to circulate until November 16, 1923, and was added to the existing stock of paper marks, which were still the only legal tender. At the same time, the Reichsbank stopped monetizing government debt by ending the discounting of Treasury bills. Bresciani-Turroni (p. 337) calls that monetary reform "a fact of fundamental importance"—yet it is ignored by TZ.

Even though newly created paper marks could not be used to finance government profligacy, the central bank continued to supply marks for commercial uses. Between November 16, 1923, and November 30, 1923, the amount of paper marks in circulation increased from 93 trillion to more than 400 trillion, and reached 1,211 trillion by July 31, 1924. Meanwhile, the quantity of rentenmarks went from 501 million on November 30, 1923, to 1,803 million on July 31, 1924. Consequently, "the stabilization of the German exchange was not obtained by means of contraction, or even by a stoppage of the expansion of the circulation of *legal* currency" (ibid.).

Most notably, and in contrast to the FTPL as stated by TZ, "The exchange was stabilized before there existed the conditions (above all the equilibrium of the Reich Budget) which alone could assure a *lasting* recovery of the monetary situation" (Bresciani-Turroni, p. 355).

The Weimar Republic faced hyperinflation because it chose to finance postwar reparation payments by money creation, and once the printing presses started rolling, it was hard to stop them. The Reichsbank was under the influence of the real bills doctrine and met all demands for credit with newly minted paper marks, believing that inflation was unlikely if the bank only discounted short-term bills that reflected real output. The problem is that bank credit is expressed in *nominal* terms. Thus, as prices rose because of rapid money growth, the demand for credit increased and businesses repaid debts in depreciated currency. As Ragnar Nurkse (1946: 16–17) stated in a League of Nations report,

> German economic thought failed to apprehend that the expansion in the money supply was at least an essential *condition* without which the general rise in prices could not have gone far. And this intellectual failure accounts in great part for the weakness of the defences which the spring tide of inflation encountered in Germany [cited in Yeager 1976: 314].

That intellectual error is also apparent in the TZ account. If they had read the account of the German hyperinflation by Thomas M. Humphrey (1980), who had a long career as a monetary economist/historian at the Federal Reserve Bank of Richmond, they would have seen that the strong version of the FTPL has "feet of clay," as Willem Buiter (2002) vividly noted. There is no doubt that monetary expansion was the fuel that fed the inflationary fire, although fiscal latitude instigated that expansion.

Major Fallacies That Misguided German Monetary Policy

Humphrey (1980: 3) carefully lays out the major fallacies that misguided monetary policymakers, blindsiding them to the dangers of excess money growth, and points to the significance of monetary reform in quickly stabilizing the value of the currency.

- First, monetary authorities blamed the inflation on external factors, believing that exchange rate depreciation was the culprit rather than domestic monetary policy.
- Second, there was a general acceptance of "a reverse causation theory of the link between money and prices."
- Third, officials falsely thought that a decrease in the real money stock (M/P) was a sign of too little rather than too much money.[3] In other words, they failed to recognize how excessive increases in the *nominal* stock of money affect the velocity of money and the price level; they overlooked the dynamic theory of money.
- Fourth, policymakers and bankers blithely accepted "the real bills doctrine according to which the money supply should accommodate itself to the needs of trade."
- Fifth, officials were misled by thinking that "the central bank can stabilize nominal market interest rates simply by pegging its discount rate at some arbitrary level."

[3]Yeager (1976: 315) points out that "prices had risen so much faster than the money supply [as people desperately tried to get rid of their depreciated paper marks caused by the Reichsbank's irresponsible increases in the quantity of money] that complaints became common of an acute *shortage* of money, despite eventual issue of denominations as high as 100 million marks."

In his discussion of the monetary reforms that led to the stabilization of the currency, Humphrey (1980: 5–6) mentions the introduction of the rentenmark, the limitation on its quantity, and the end of debt monetization, as well as the importance of central bank credibility in changing the public's expectations. He also notes that fiscal policy was implemented to cut the size of government deficits by a combination of spending cuts and tax increases.

What the work of Bresciani-Turroni and Humphrey teaches us is that models like the strong version of FTPL are not sufficient to inform us of the forces that underlie hyperinflation and stabilization. A close study of the policy actions taken in Weimar Germany shows that inflationary expectations are grounded in the credibility of central banks as well as fiscal authorities. The competing theory that "explosive expectations" can generate runaway inflation *without any change in the money supply* cannot be supported by the experience of the German hyperinflation. Likewise, there is no evidence to support TZ's claim that the hyperinflation was ended by fiscal action—that is, backing the rentenmark by real estate revenues. Rather, it was ended by fundamental monetary reform and a credible commitment to return to price stability as well as fiscal fortitude.

Emergency Monies

There is another feature of the German hyperinflation and stabilization that needs more attention. In October 1923, before the introduction of the rentenmark, the public faced a rapidly depreciating legal paper mark and "demanded a means of payment with a *stable value*" (Bresciani-Turroni, p. 343). Consequently, "the Government authorized and even encouraged the issue of 'emergency monies with a constant value'" (ibid.). Those issues were backed by "Gold Loan securities or by a special type of Gold Treasury Bond," but that "guarantee . . . was purely fictitious" (p. 344). Nevertheless, the public preferred to hold the "stable-value" emergency monies and rejected the legal tender money (i.e., the paper mark).[4] That outcome "was evidence of the

[4]During the last stage of the hyperinflation, "legal money was rejected by the public" as people switched to foreign currencies, brought into circulation the old national metallic money, and used new monies supplied by private firms (Bresciani-Turroni 1953: 341).

spontaneous reaction of the economic organism against the depreciation of the legal currency" (p. 345).[5]

The introduction of the rentenmark and its quantity constraint helped relieve some of the "monetary chaos." Moreover, the public's confidence in the new currency was reinforced by the "constant-value clause," which obligated those who took out loans from the Rentenbank to repay their debts in the same quantity of *gold marks* as represented in the original loan. That clause was intended to prevent the speculation that occurred during the hyperinflation when businesses and others took out bank loans in nominal paper marks but repaid them using greatly depreciated marks, thus giving speculators a strong incentive to support runaway inflation (Bresciani-Turroni, p. 353).

The Monetary Law of 1924

German monetary experts, writing in the Dawes Report of 1924, which sought to restore monetary and economic stability in Weimar Germany, viewed the "liquid cover" for the rentenmark as "insufficient to guarantee a permanent [monetary] system." They argued for the removal of the rentenmark and the introduction of a convertible currency. Their proposal was accepted with the passage of the Monetary Law of August 30, 1924, which made the reichsmark the new legal tender (Bresciani-Turroni, p. 338).

When the Monetary Law of August 30, 1924, became effective on October 11, the Reichsbank introduced the new currency, the "reichsmark," and abolished the constant-value clause, which was deemed unnecessary as the paper mark was now convertible into the reichsmark at an exchange rate of 1 reichsmark = 1 billion paper marks (1 billion = $1{,}000{,}000^2$), and the rentenmark was convertible into the new currency at a rate of 1 to 1. On June 5, 1925, the legal tender status of the old paper mark ended and it was taken out of circulation (Bresciani-Turroni, pp. 353–54).[6]

[5]Other types of emergency currencies, most of them illegal, already had appeared before the stable-value currencies. By October 1923, it is estimated that 2,000 types of emergency currencies were circulating in the Weimar Republic (Bresciani-Turroni 1953: 343).

[6]The reichsmark had a fixed gold content but was not convertible into gold until April 1930, at the discretion of the Reichsbank (Bresciani-Turroni 1953: 354).

Misdiagnosis of Fed Policy

In addition to misinterpreting the German hyperinflation and stabilization, TZ also misdiagnose recent Fed policy. They ignore the dynamic theory of money and use the crude quantity theory of money as a straw man to argue that since the U.S. monetary base (currency held by the public plus bank reserves) increased by 32.3 percent between November 2008 and September 2012, the price level should have increased by a similar amount. Yet, despite massive quantitative easing (QE) during that period, inflation remained tame. It did so, according to TZ, because the new base money created by the Fed's large-scale purchase of mortgage-backed securities was "backed by the returns from real estate investments." In short, "as long as the expected primary surpluses backing existing government liabilities haven't changed, there is no reason for the price level to change either" (Tutino and Zarazaga, p. 4).

What TZ ignore is the fact that *money growth*, in contrast to base growth, remained relatively slow because the Fed sterilized most of the new base money by selling Treasury bills, using reverse repos, and paying interest on reserves (IOR)—a policy that began in October 2008.[7] Macroprudential regulation also helped plug up the monetary transmission mechanism and keep the money multiplier historically low (see Dorn 2015).

All in all, one must agree with Buiter when he says the FTPL model, in its strong version, is "made of clay."[8] More telling, by arguing that "inflation is not always and everywhere a monetary phenomenon," Tutino and Zarazaga undermine the responsibility of central banks to maintain the long-run value of fiat money.

[7]For a detailed discussion of the impact of the Fed's IOR policy, see Selgin (2016).

[8]Carlstrom and Fuerst (2000: 31) conclude their analysis of the strong form of FTPL, "in which fiscal policy affects the price level independent of the money supply process," by noting that this version of FTPL "is little more than an intellectual curiosity." Paul Roemer would no doubt agree (see Roemer 2016). In a more recent paper delivered at the 2016 Jackson Hole Conference, Sims takes a more moderate approach to FTPL, arguing that it "does not . . . simply replace the notion that the quantity of money determines the price level with the idea that the quantity of government debt, or the sequence of nominal deficits, determines the price level. It implies that interest rate policy, tax policy, and expenditure policy, both now and as they are expected to evolve in the future, jointly determine the price level" (Sims 2016: 5).

Conclusion

The fiscal theory of the price level (in either its weak or strong form) does not overturn the idea that "inflation is always and everywhere a monetary phenomenon." This is not to say that that fiscal policy cannot influence monetary policy; no central bank has complete independence. History is replete with episodes of high inflation brought about by governments that used central banks to finance deficit spending. The FTPL must be examined in its various forms within the context of actual monetary history. Without a proper understanding of the sequence of events leading to inflation and deflation, theories of price-level determination risk being what Ronald Coase (1992: 714) called "blackboard economics."

Implementing a monetary rule to constrain the power of central banks would help depoliticize monetary policy. However, a more permanent separation of money and politics could occur if discretionary government fiat money were replaced by a free-market monetary system, in which private contracts, competition, and a convertible currency safeguarded the property right individuals have in a sound currency.

References

Bresciani-Turroni, C. ([1931] 1953) *The Economics of Inflation: A Study of Currency Depreciation in Post-War Germany, 1914–1923.* Foreword by L. Robbins; translated by M. E. Savers. London: George Allen and Unwin.

Buiter, W. H. (2002) "The Fiscal Theory of the Price Level: A Critique." *The Economic Journal* 112 (July): 459–80.

Cagan, P. (1956) "The Monetary Dynamics of Hyperinflation." In M. Friedman (ed.), *Studies in the Quantity Theory of Money*, 25–117. Chicago: University of Chicago Press.

Carlstrom, C. T., and Fuerst, T. S. (2000) "The Fiscal Theory of the Price Level." Federal Reserve Bank of Cleveland *Economic Review* 36 (1): 22–32.

Coase, R. (1992) "The Institutional Structure of Production." *American Economic Review* 82 (4): 713–19.

Dorn, J. A. (1987) "The Search for Stable Money: A Historical Perspective." In J. A. Dorn and A. J. Schwartz (eds.), *The Search for Stable Money: Essays on Monetary Reform*, chap. 1. Chicago: University of Chicago Press.

_____ (2015) "Unconventional Fed Policy Hasn't Paid Off." *U.S. News & World Report* (December 17).

_____ (2016) "The Fiscal Theory of the Price Level: True and False." *Alt-M* (August 23): www.alt-m.org.

Friedman, M. (1970) "The Counter-Revolution in Monetary Theory." IEA Occasional Paper No. 33. London: Institute of Economic Affairs.

Humphrey, T. M. (1980) "Eliminating Runaway Inflation: Lessons from the German Hyperinflation." Federal Reserve Bank of Richmond *Economic Review* (July/August): 3–7.

McCallum, B. T., and Nelson, E. (2005) "Monetary and Fiscal Theories of the Price Level: The Irreconcilable Differences." *Oxford Review of Economic Policy* 21 (4): 565–83.

Nurkse, R. (1946) *The Course and Control of Inflation: A Review of Monetary Experience in Europe After World War I*. Geneva: League of Nations. (Nurkse wrote Part 1 of this report.)

Roemer, P. (2016) "The Trouble with Macroeconomics." Working Paper, Stern School of Business, New York University. Forthcoming in *The American Economist*.

Selgin, G. (2016) "Interest on Reserves and the Fed's Balance Sheet." Testimony before the Monetary Policy and Trade Subcommittee of the House Committee on Financial Services (May 17); reprinted at www.alt-m.org/2016/05/17/interest-on-reserves-fed-balance-sheet.

Sims, C. A. (1994) "A Simple Model for Study of the Determination of the Price Level and the Interaction of Monetary and Fiscal Policy." *Economic Theory* 4 (3): 381–99.

_____ (2016) "Fiscal Policy, Monetary Policy and Central Bank Independence." Paper presented at the Kansas City Federal Reserve Bank's Annual Jackson Hole Conference (August 26).

Tutino, A., and Zarazaga, C. (2014) "Inflation Is Not Always and Everywhere a Monetary Phenomenon." Federal Reserve Bank of Dallas *Economic Letter* 9 (6): 1–4.

Warburton, C. (1966) *Depression, Inflation, and Monetary Policy: Selected Papers, 1945–1953*. Baltimore: The Johns Hopkins Press.

Yeager, L. B. (1976) *International Monetary Relations: Theory, History, and Policy*. 2nd ed. New York: Harper and Row.

_____ (1997) *The Fluttering Veil: Essay on Monetary Disequilibrium*. Edited by G. Selgin. Indianapolis: Liberty Fund.

Economic Policy Uncertainty and Small Business Decisions

Robert Krol

The U.S. economy has experienced a slow recovery from the 2007–09 recession. Economic growth remains below its historical average. One possible contributor to the poor economic performance is economic policy uncertainty. For example, the future course of monetary policy has been unclear over the recovery time period. Given the important role of small businesses in job creation, this article examines the impact of economic policy uncertainty on small-business decisions.[1]

A number of economists have examined the impact of general economic uncertainty on business decisions. Bernanke (1983); Dixit and Pindyck (1994); Bloom, Bond, and Reenen (2007); and Bloom (2009, 2014) have shown the adverse impact of general economic uncertainty on business investment decisions. Bloom, Bond, and Reenen (2007) speculate that general economic uncertainty will also adversely impact hiring decisions. Ghosal and Ye (2015) find this to be the case. Lower investment and employment occur because uncertainty makes firms less sure about the returns associated with capital expenditures or hiring. Since there are nonrecoverable costs

---

Cato Journal, Vol. 37, No. 1 (Winter 2017).

Robert Krol is Professor of Economics at California State University, Northridge. He thanks William Dunkelberg of the National Federation of Independent Business for providing the survey data used in this article. Shirley Svorny provided helpful comments.

[1]For a discussion of the important role of small businesses in economic growth, see Decker et al. (2014).

associated with a decision to invest in capital or hire and train workers, uncertainty makes it prudent to delay capital expenditures or hiring. Uncertainty also worsens information asymmetries between lenders and borrowers. With greater uncertainty, the chances of bankruptcy increase. As a result, banks tend to delay lending to firms, slowing business expansion (Greenwald and Stiglitz 1990).

Baker, Bloom, and Davis (2016) extend the notion of uncertainty to include economic policy uncertainty (EPU). They construct an index to measure EPU using a computer-based search that quantifies the frequency of articles dealing with uncertain policy issues in leading U.S. newspapers. Their study provides evidence at the firm and aggregate levels that EPU has a significant impact on economic activity. In sectors of the economy that do a substantial portion of business with the government, Baker, Bloom, and Davis find higher EPU increases stock price volatility and lowers investment and employment at the firm level. An example of this would be firms in the defense industry. At the macroeconomic level, they estimate a vector auto-regression and find a 90 basis point increase in the EPU index decreases aggregate industrial production by 1.2 percent and employment by 0.35 percent.

In this article, I examine the impact of EPU on small business decisions using the National Federation of Independent Business survey on economic trends. This survey asks questions that quantify small business expansion plans and their economic outlook. Schweitzer and Shane (2011) also use this data to examine the impact of EPU on small business decisions, but my analysis differs from Schweitzer and Shane in a number of ways. In assessing small business response to EPU, Schweitzer and Shane focus only on what firms said about their plans for employment and capital investments. I examine a broader set of survey responses, adding plans to increase worker compensation, plans for general expansion, and the degree of business optimism among small business respondents.

Schweitzer and Shane control for general economic and credit market conditions, but not supply shocks and general economic uncertainty. Supply shocks can influence small business expansion plans by changing production costs. I take supply shocks into account. In order to identify the impact of EPU on small business decisions, general economic uncertainty must be controlled. I use the Chicago Board of Options 30-day volatility index for S&P 500 options to measure general economic uncertainty.

Another issue is whether business responses to changes in economic policy persist over time. Schweitzer and Shane do not address that question. I add a lagged dependent variable to the regressions to see if business responses to changes in economic policy persist over time. Finally, the sample period I use is longer than that used by Schweitzer and Shane: it covers an additional four-and-a-half years of data following the 2007–09 recession.

Regression Model and Data

The regression model used to examine the impact of EPU on small business decisions is shown in Equation 1:

$$(1)\quad SBD_{i,t} = \alpha + \beta\ EPU_t + \gamma\ VIX_t + \varphi_j\ CONTROL_{j,t} + \mu_t.$$

The dependent variable, small business decision, $SBD_{i,t}$, captures small business owners' responses to questions in the National Federation of Independent Business monthly survey on economic trends. Survey questions examine small businesses' plans to increase (1) employment, (2) capital expenditures, or (3) compensation and ask about (4) general business expansion plans and (5) business optimism (see National Federation of Independent Business 2016). This provides five alternative measures of the dependent variable in Equation 1. The monthly survey began in 1986 and contains the responses of more than 1,000 businesses each month.

EPU_t is measured using the index constructed by Baker, Bloom, and Davis (2016). This index is constructed using a computer-based search that quantifies the frequency of articles dealing with economic policy uncertainty in 10 leading U.S. newspapers.[2] Articles counted contain triple combinations of words such as "uncertainty or uncertain," "economic or economy," and a policy term like "Congress, deficit, Federal Reserve, legislation, regulation, or White House." The relevant article count is divided by the total number of articles in the newspaper for each month. This calculation is then divided by the standard deviation of the series. An average is calculated for the 10 newspapers and is normalized so the average for the sample

[2]The 10 newspapers are *USA Today, Miami Herald, Chicago Tribune, Washington Post, Los Angeles Times, Boston Globe, San Francisco Chronicle, Dallas Morning News, New York Times,* and *Wall Street Journal.*

period is equal to 100. Higher EPU is expected to have a *negative* impact on small business expansion and worker compensation.

To proxy general economic uncertainty, I follow Baker, Bloom, and Davis (2016), Krol (2014), and Bloom (2009) by using the Chicago Board of Options 30-day volatility index (VIX) for S&P 500 options. The VIX index provides a measure of investor sentiment and implied market volatility. The VIX index uses the Black-Scholes option-pricing model to calculate the expected volatility based on market prices. The index weighs put and call option prices that turn out to be unprofitable on a particular date. Because option prices are positively related to market volatility, a higher VIX index implies greater expected economic uncertainty (see Chicago Board Options Exchange 2014). Higher general economic uncertainty is expected to have a negative impact on small business expansion and worker compensation.

When examining (isolating) the impact of EPU on the economy, it is important to control for general economic uncertainty—hence, the control variable ($CONTROL_{j,t}$) in Equation 1. At issue is whether the two uncertainty measures provide unique information about the economy. While the two kinds of uncertainty are related, Baker, Bloom, and Davis (2016) provide evidence that the economic policy index shows "distinct variation" that corresponds with time periods of high levels of economic policy uncertainty.

Schweitzer and Shane (2011) do not use this index nor do they control for general economic uncertainty. Instead, they use the Michigan Consumer Sentiment Index in some specifications over concerns that the EPU index may be capturing general swings in consumer sentiment.[3] However, they don't address the general level of uncertainty directly.

To isolate the impact of uncertainty, it is also necessary to control for the business cycle, supply shocks, and credit market conditions. The unemployment rate and industrial production index are used to measure current economic conditions. The unemployment rate is expected to have a negative impact on small business expansion and worker compensation, while industrial production is expected to have a positive impact. The price of West Texas Intermediate crude

[3]They do not explicitly report their regression results that use the Michigan Consumer Sentiment Index. They only report that the economic policy uncertainty variable remains statistically significant when it is included in the regression.

oil, deflated by the consumer price index, serves as a proxy for major supply or cost shocks. Higher real crude oil prices should have a negative impact on small business expansion and worker compensation. Following Schweitzer and Shane, the prime rate and Libor interest rates are used to measure current credit market conditions (i.e., the cost of credit). The prime rate and Libor interest rates are expected to have a negative impact on small business expansion and worker compensation. As the cost of credit rises, some firms will find expansion to be uneconomic. The regression also includes a linear time trend to capture long-term economic growth factors that are not related to the cyclical component of small business performance.

To examine the persistence of the influence of economic policy uncertainty over time, each survey response regression is estimated with and without a lagged dependent variable. A significant lagged dependent variable suggests that the survey response persists for more than one month. In other words, last month's survey response is related to the current month's response, suggesting that businesses modify their perceptions of economic uncertainty over time in a way that may be consistent with a dynamic adjustment process.

Results

The regression results are reported in Table 1. The regression is estimated for each of the five different survey responses (plans to increase employment, capital expenditures, compensation, expansion plans, and business optimism) for the period beginning in January 1990 and ending January 2016.[4]

Economic policy uncertainty has a large negative impact on small business expansion decisions. Consistent with the hypothesis that increases in EPU discourage small business expansion, the EPU variable is negative and significant at the 10 percent level or less in 9 of the 10 regressions, and it is significant in all of the regressions with a lagged dependent variable that controls for persistence in the survey responses. The evidence supports the premise that there is persistence in survey responses over time, as the lagged dependent variable, small business decisions, $SBD_{i,t-1}$ is always significant.

The measure of general economic uncertainty, captured by the VIX index, is also negative and significant at the 10 percent level or

[4]The sample starting date is determined by the availability of VIX data.

TABLE 1
Regression Results, 1990:1 to 2016:1

	emp	emp	expand	expand	invest	invest	opt	opt	comp	comp
constant	−73.35	−35.50	−172.17	−47.93	−42.39	−20.05	3.76	0.95	59.61	30.29
	(.008)	(.015)	(.000)	(.002)	(.090)	(.135)	(.000)	(.000)	(9.017)	(.069)
lagged dep		0.51		0.66		0.49		0.75		0.44
		(.000)		(.000)		(.000)		(.000)		(.000)
epu	−3.22	−2.26	−5.57	−2.93	−2.16	−1.65	−0.03	−0.02	−1.41	−1.23
	(.028)	(.022)	(.000)	(.000)	(.090)	(.023)	(.004)	(.000)	(.126)	(.057)
vix	−4.45	−2.41	−3.18	−0.90	−3.30	−1.77	−0.04	−0.01	−0.60	−0.56
	(.002)	(.002)	(.017)	(.145)	(.001)	(.004)	(.004)	(.020)	(.568)	(.359)
un	−1.07	−0.35	0.26	0.28	−0.58	−0.20	−0.01	0.001	−2.29	−1.17
	(.003)	(.175)	(.085)	(.056)	(.063)	(.331)	(.200)	(.689)	(.000)	(.000)
ipi	25.59	12.93	51.32	15.97	22.34	11.62	0.31	0.09	−6.64	−2.53
	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.223)	(.473)
realoil	−3.97	−2.24	−5.01	−1.85	−3.27	−1.75	−0.02	−0.01	−0.38	−0.33
	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.112)	(.010)	(.664)	(.529)
prime	0.26	0.16	−2.03	−0.78	0.08	−0.02	−0.01	−0.004	0.65	0.34
	(.650)	(.614)	(.021)	(.023)	(.890)	(.957)	(.127)	(.106)	(.148)	(.194)
libor	0.07	−0.002	2.18	0.721	0.10	0.07	0.001	0.001	−0.17	−0.08
	(.892)	(.994)	(.000)	(.003)	(.830)	(.806)	(.869)	(.607)	(.712)	(.782)

trend	−0.03	−0.01	−0.08	−0.03	−0.06	−0.03	−0.001	−0.0002	0.01	0.01
	(.050	(.059)	(.000)	(.005)	(.000)	(.000)	(.000)	(.001)	(.329)	(.499)
rbarsq	0.79	0.85	0.78	0.88	0.82	0.86	0.76	0.90	0.83	0.86
see	2.73	2.34	2.86	2.13	2.45	2.15	0.03	0.02	2.00	1.80

NOTES: Table 1 reports ordinary least squares estimates of coefficients and p-values in parentheses for each regression. The standard errors used in calculating the p-values are heteroscedasticity-autocorrelated consistent (Newey and West 1987). The first row reports the dependent variable for the regression results reported in the column: emp = net percentage of firms planning to increase employment; expand = percentage of firms saying it is a good time to expand; invest = net percentage of firms planning to increase capital expenditures; opt = an index of business optimism; and comp = net percentage of firms planning to increase compensation. The first column reports explanatory variables and regression statistics: constant = regression intercept; lagged dep = the lagged dependent variable; epu = economic policy uncertainty index; vix = the volatility index used to measure general economic uncertainty; un = the unemployment rate; ipi = the industrial production index; realoil = the price of oil deflated by the consumer price index; prime = the prime interest rate; libor = the libor interest rate; and trend = a time trend. R-bar-squared (rbarsq) and regression standard error (see) are reported in the bottom two rows, respectively.

less in 7 of the 10 regressions. In 5 of the 7 regressions where both uncertainty variables were significant, the VIX index had a larger impact. The EPU variable had a larger negative impact in the decision to expand a business. These results support the idea that uncertainty, whether general or policy generated, negatively impacts small business decisions to expand.

Schweitzer and Shane (2011) do not report their estimated individual coefficients or p-values so a direct comparison of results is not possible. They only report the EPU variable "has a statistically significant negative effect" on business hiring and capital spending plans (Schweitzer and Shane 2011: 4). The R-bar-squared measures I report in this article for the hiring and capital expenditure regressions are similar to the results found in Schweitzer and Shane. They report that inclusion of the Michigan Consumer Sentiment Index does not affect their results.

Looking at the control variables, a better economy—measured by higher industrial production or a lower unemployment rate—has a large positive impact on small businesses, except when it comes to compensation decisions. Real oil price shocks have a large negative impact on small business expansion, except for compensation decisions. General credit conditions do not appear to be a consistently important factor in small business decisions. One possible reason for this result is that personal financing plays an important role in small business startups and expansions (see Miller, Hoffer, and Wille 2016). Finally, the relatively high R-bar-squared indicates that each regression has considerable explanatory power.

Conclusion

The U.S. economy has experienced a slow recovery from the 2007–09 recession. One possible contributor to the poor economic performance is economic policy uncertainty—that is, uncertainty about what the government may do to try to stabilize the economy and uncertainty about the consequences of those actions. Given the important role of small businesses in job creation, it makes sense to examine the impact of economic policy and general economic uncertainty on small business decisions.

Economic theory suggests greater uncertainty in the economy increases the value of waiting to invest and to hire additional workers.

The evidence presented in this article suggests, like much of the previous work on small businesses, that general economic uncertainty adversely affects small businesses. This article tests the importance of an additional measure of uncertainty: economic policy uncertainty. Looking at small business survey responses, I find that increases in EPU lead respondents to say they are reducing employment, investment, and expansion. Increases in EPU also lead to a decline in general optimism among the survey respondents. Clearly, EPU has a negative impact on small business activity. Furthermore, the adverse impact of EPU tends to persist over time. Controls for the current state of the economy and supply shocks are significant but, more important, they do not eliminate the finding of a negative impact of EPU on small business activity.

There appears to be considerable disagreement among economists and policymakers on what policies are needed to restore economic growth. Policymakers must get policy right, but lack of clarity in the policymaking process is not a good thing, as it reduces small business activity, slowing economic growth and job creation.

References

Baker, S. R.; Bloom, N.; and Davis, S. (2016) "Measuring Economic Policy Uncertainty." *Quarterly Journal of Economics* (131): 1593–1636.

Bernanke, B. S. (1983) "Irreversibility, Uncertainty, and Cyclical Investment." *Quarterly Journal of Economics* (96): 85–106.

Bloom, N. (2009) "The Impact of Uncertainty Shocks." *Econometrica* (77): 623–85.

_____ (2014) "Fluctuations in Uncertainty." *Journal of Economic Perspectives* (28): 153–75.

Bloom, N.; Bond, S.; and Reenen, J. V. (2007) "Uncertainty and Investment Dynamics." *Review of Economic Studies* (74): 391–415.

Chicago Board Options Exchange (2014) "The CBOE Volatility Index-VIX." White Paper: www.cboe.com/micro/vix/vixwhite.pdf.

Decker, R.; Haltiwanger, J.; Jarmin, R.; and Miranda, J. (2014) "The Role of Entrepreneurship in U.S. Job Creation and Economic Dynamism." *Journal of Economic Perspectives* (28): 3–24.

Dixit, A. K., and Pindyck, R. S. (1994) *Investment under Uncertainty*. Princeton: Princeton University Press.

Ghosal, V., and Ye, Y. (2015) "Uncertainty and the Employment Dynamics of Small and Large Businesses." *Small Business Economics* (44): 529–58.

Greenwald, B., and Stiglitz, J. (1990) "Macroeconomic Models with Equity and Credit Rationing." In R. G. Hubbard (ed.), *Asymmetric Information, Corporate Finance, and Investment*, 15–42. Chicago: University of Chicago Press.

Krol, R. (2014) "Economic Policy Uncertainty and Exchange Rate Volatility." *International Finance* (17): 241–56.

Miller, M. M.; Hoffer, A. J.; and Wille, D. (2016) "Small-Business Financing after the Financial Crisis; Lessons from the Literature." Mercatus Working Paper, George Mason University (August).

National Federation of Independent Business (2016) "Small Business Economic Trends." January 2016 Report: www.nfib.com/surveys/small-business-economic-trends.

Newey, W. K., and West, K. D. (1987) "A Simple Positive-Definite Heteroskedasticity and Autocorrelated Covarience Matrix." *Econometrica* (55): 703–08.

Schweitzer, M. E., and Shane, S. (2011) "Economic Policy Uncertainty and Small Business Expansion." Federal Reserve Bank of Cleveland *Economic Commentary*, No. 24.

The Stationary Bandit Model of Intellectual Property

Sinclair Davidson and Jason Potts

We propose a new model of intellectual property that presents a different view than the market failure/monopoly rent model advanced by Arrow (1962), in which governments protect inventors from private theft. Instead, using Olson (1993), we represent a *public theft* model of intellectual property arising when entrepreneurs acting in global markets seek protection from a stationary bandit (their home government) principally against the depredations of other governments (the roving bandits). This model explains why institutional quality matters to the global location of R&D intensive industries, such as biopharma, and why so much intellectual property is located in tax havens.

Government, Citizens, and Intellectual Property

The first duty of government is to protect its citizens. Put the other way, the value to a citizen of a strong government is to protect their life, liberty, and property from the depredations of others. In the political romantic view, this is the protection of the weak by the strong. But the real nature of the bargain is that the weak must then

---

Cato Journal, Vol. 37, No. 1 (Winter 2017).

Sinclair Davidson is Professor of Institutional Economics at RMIT University and a Senior Research Fellow at the Institute of Public Affairs, both in Melbourne, Australia. Jason Potts is Professor of Economics at RMIT University and an Adjunct Fellow at the Institute of Public Affairs. The authors thank participants at the Southern Economic Association Conference (2016) for useful comments on a prior version of this article.

pay tribute to the strong. In his antisocial contract model of the origin of government, Mancur Olson (1993) argued that governments offering protection are essentially "roving bandits" who have settled down to establish a monopoly on theft as "stationary bandits" who protect their tributes—now called tax-paying citizens—from external threats. Stationary bandits still plunder their captives to the maximum extent, but they do so rationally, leaving them sufficient resources and furnishing peaceful order and public goods to maximize the future stream of taxes.

In using the stationary bandit model of government,[1] we argue that new ideas—of the sort that become patents, copyrights, and trademarks—emerge as economic rights,[2] born global into a world of roving bandits. The holders of those rights seek protection from a stationary bandit who extracts tribute in return. The key insight of our model, however, is a sharper distinction of who those bandits are. In the standard model of intellectual property, benevolent national governments grant a temporary monopoly privilege to protect the creative inventor citizen from the unscrupulous depredations of private competitors or even consumers. The argument goes that without a legislative prohibition on copying (institutionally defined as theft), a competitive market will provide only weak incentives to invest in creating new ideas—that is, there will be market failure (Arrow 1962), and society will suffer a suboptimal level of creative-inventive activity (Posner 2005). In the standard model, government protects holders of intellectual property from private theft, making the nation safe for private creativity.

Now private theft is certainly a problem, but we maintain that *public theft* is much worse. The main predators on intellectual property, we argue, are not private competitors (e.g., copying a technology) or individual consumers (e.g., illegal downloads), but other governments through their client services and cronies, who variously engage in outright theft or coercive measures to diminish and

[1]Our approach also builds on Barzel's (2002) Hobbesean model of the state as a violent third-party enforcement mechanism that citizens consent to when safeguards are in place to mitigate abuse.

[2]In the Allen and Alchian (1977: 114) sense of "economic rights" as the expectation of benefiting from the value of the assets created, requiring third-party enforcement. See also Haber (2016).

deplete the value of their intellectual property (Ezell, Nager, and Atkinson 2016). A common example is the treatment of biopharma, routinely subject to outright theft, compulsory licensing, and other practices that diminish its value (Wu and Ezell 2016). In our new model, vulnerable subjects seek protection for their private economic property from the banditry of other governments, by registering their property with a strong government whom they trust to be powerful enough to protect it as they peacefully engage in trade and commerce throughout the world. The origin of intellectual property is when one of these roving bandits finds it worthwhile to become a stationary bandit by protecting the idea (the entrepreneurial discovery and the economic asset) from organized theft by other governments and settling down as a monopoly exploiter.

This sort of public action on behalf of private citizens should not be confused with rent seeking or cronyism, where agents seek private benefit from insider markets to public power. This is important, because the standard model of intellectual property is in effect a benevolent form of rent creation—in which the private citizen is incentivized to produce novelty and make it public, in return for a government-granted monopoly—thus creating a mechanism to transfer resources from other agents in the economy to the target. But in our model of intellectual property, when private agents seek protection, they are not seeking to exploit insider political markets or to transfer resources from other groups of citizens (Olson 1965), but rather to have their economic property rights represented and protected as they venture through the world seeking to trade.[3] In return, they grant that protector government an exclusive right to exploit them through perpetual taxation of the property. The other side of the intellectual property is not market failure and the creation of a monopoly form of intellectual property right, as Arrow (1962) claimed, but rather market making in return for a monopoly on taxation.

A better model of this exchange is gunboat diplomacy. Governments have long used force to open markets when acting as agents on behalf of their merchant citizens seeking to extend their

[3]Because of fixed costs associated in setting up intellectual property protection, larger jurisdictions are predicted to offer greater protection than smaller jurisdictions (Mulligan and Shleifer 2005).

property rights and opportunities for exchange and commerce into new territories.[4] Widely misrepresented as imperialist venturing, implying a political empire-building motive, these actions are better understood as venturesome tax policy, because the public displays of force—through the posturing of naval assets, say—are on behalf of those over whom the government has monopoly tax rights and therefore a vested interest in their market success. The origin and nature of intellectual property rights can be similarly construed.

The Market Failure Model of Intellectual Property

In the standard account, the economics of intellectual property is a two-sided ledger, balancing the benefit of an investment incentive against the cost of a monopoly. When economists argue for or against, or for stronger or weaker intellectual property, they are arguing about the relative size of the entries in this ledger. The prime argument that the benefit exceeds the cost is that of market failure in the production of information owing to a fixed cost (e.g., R&D) that is unrecoverable under perfect competition (Arrow 1962). The opposite argument is that there is no such market failure and that innovation is possible under perfect competition, and so the benefit can be obtained without paying the monopoly cost (Boldrin and Levine 2008, 2013). Dourado and Tabarrok (2015) show how the rent-seeking costs of the monopoly at some margins will be greater than the public benefit. The existence of such a tradeoff implies an optimal size and duration of that government-created rent (Nordhaus 1969, Gilbert and Shapiro 1990, Romer 2002, Landes and Posner 2003), including the prospect that private or informal institutions may be more efficient (Hall et al. 2014), or that more efficient mechanisms might be designed (Wright 1983, Kremer 1998) or could evolve (Ostrom and Hess 2006, Benkler 2006).

In the market failure model of intellectual property, economic efficiency requires government support for private creative and

[4]Examples include the arrival of the U.S. Commodore Perry fleet in Tokyo Bay in 1853 to open Japan to trade with the West; or the British Naval fleet ending the Opium Wars with China in 1839–42, resulting in the Treaty of Nanking that opened Hong Kong to foreign trade; or the U.S. defeat of the Barbary States, ending tribute to the pirates under protection of the Ottoman empire, enabling U.S. and European merchants to trade without harassment along the North African coast.

innovative activity (Arrow 1962, Scotchmer 2004). Two premises are baked into this argument. The first is that intellectual property—patents, copyrights, trademarks, as well as industrial design rights, trade dress, plant varieties, and trade secrets—is a government-granted monopoly. This can certainly be read in the history of intellectual property, which grew out of royal grants and privileges, with patents from the *Statute of Monopolies* (1624) and copyright from the *Statute of Anne* (1710). The property right exists because the government grants a temporary monopoly to create an incentive for the private production of new ideas, information, and technology. The second premise is that the natural domain of the artificially created property right, as a monopoly on trade, is therefore identical to the domain of the government that granted it. The market failure model of intellectual property is told from the perspective of a large nation state. While rarely stated in this way, this is a *social contract* view of intellectual property in which governments act on behalf of the collective will of the people, who benefit overall from having such an incentive within their civil jurisdiction and are willing to grant special privileges to those among them who furnish what will eventually become public goods (Nordhaus 1969).[5] This is a model of intellectual property in which government is a benevolent agent acting to maximize social value for its citizens by dispensing temporary rents. Overall, most thinking about intellectual property is grounded in the idea that it is a legal property right granted by a nation-state rather than an economic right with global expression.

The Stationary Bandit Model of Intellectual Property

The stationary bandit model of government (Olson 1993, 1995; McGuire and Olson 1996) is an anti-social-contract[6] theory approach. A stationary bandit is a roving bandit who settles down to monopolize theft over his subjects at a given rate of tax, which is preferable to roving bandits who steal everything. The model arises from the combination of a domain of "encompassing interest"

[5]This model then extends to international agreements (e.g. TRIPS) to replicate this logic through harmonized intellectual property provisions in trade treaties.

[6]"Government for groups larger than tribes normally arises, not because of social contracts or voluntary transactions of any kind, but rather because of rational self-interest among those who can organize the greatest capacity for violence" (Olson 1993: 568).

(Olson 1982) interacting with the private value of peace, order and public goods, including protection (Olson 1993).

In our alternative model, it is not the government that grants the monopoly right to the inventor, now recognized as a public benefactor, from the many grateful citizens, who consent to that transfer. Rather, in our Olson-inspired model,[7] intellectual property arises when a citizen *who already has property* in an anarchic environment that is vulnerable to theft seeks protection from a strong government in exchange for an exclusive *monopoly right to tax*. This is not a one-sided gift from government. Rather, there is a mutual exchange of monopoly rights, in which the inventive and venturesome citizen creates an encompassing interest—as a monopoly right to tax—in return for enforcing the economic property rights in the idea as they seek to create and grow market opportunities throughout the world. A strong government does this by opening an area for trade and protecting the property right over a potentially global context from the depredations of roving bandits. This is not market failure and rent creation, but market making through negotiation backed with threats of force (a.k.a., "gunboat diplomacy"). This is not an internal transfer from citizen to citizen within a nation-state, but the government protection of one citizen to freely go about their peaceable business with their new idea in a disordered and potentially hostile world, in return for exclusive license to tax that flow of new trade.

In the old view of the economics of intellectual property, the role of government is that of a benevolent agent enforcing a social contract in which a collective agrees to grant special privilege (intellectual property, as a monopoly right) in order to incentivize private investment in new knowledge creation and to mandate free revealing. The government is simply an agent to facilitate an internal transfer between the citizens of a nation-state, who have collectively consented to an artificial monopoly, in order to create the monopoly rent incentive to spur its inventive citizens to action. But in our new view of the economics of intellectual property, the role

[7]This is closely related to Barzel's (2000, 2002) theory of the state as a violent monopoly enforcement mechanism of economic rights (as distinct from legal rights) to facilitate third-party agreements about economic rights that are otherwise difficult to enforce. Barzel argues that citizens will only consent when further agreements are in place to prevent abuse of the citizen by the state (what Barzel calls dictatorship), which in our model is equivalent to self-predation.

of government is one side of a mutually rewarding strategic alliance, in which inventive and venturesome citizens (including corporate citizens) seek government protection as they engage in trade and seek to expand their markets, in exchange for a share of surplus created, taken as tax. And what they need protection from is other powerful bandits, who for the most part are other governments or the cronies they enable (Olson 1993).

It is misleading to think of the origin of new ideas and new technologies as a consequence of "investment" that is incentivized within a political jurisdiction. New ideas with economic value are born into markets, not into nation-states. New ideas that make valuable information property arise in the process of economic activity, or in the expectation of economic activity, as a by-product of entrepreneurial discovery and venturing. This is why the information property that is of value—the intellectual property—is often much more than the technical invention, extending through the ways in which value is created and extracted. In this sense, intellectual property is, in practice, a much larger part of commercial and market activity than the accounting value of the parts that can be registered and protected. The role of government from this perspective is not a benevolent grant of an artificial form of property within the borders of the nation-state; rather, it is to protect property that already exists in consequence of economic activity and to extend that protection as far as possible. Moreover, a strong government will do so from its own self-interest in order to maximize its share of the gains from trade.

The stationary bandit model of intellectual property expresses the evolutionary conception of market capitalism advanced by Schumpeter (1912, 1942), Baumol (2002), and Phelps (2013). In their view, firms compete through innovation, and the role of government is to supply the institutional conditions for such competition to flourish. However, most innovation policy, including much in the Schumpeterian tradition (Nelson 1959, 1993; Mowery and Rosenberg 1989; Mowery 1995), sees a more active role for government premised on the market failure model of science and technology (Arrow 1962, Romer 2002, Steinmueller 2010). The very notion of a market failure diagnosis in the aggregate social welfare model sets up a government solution. Those arguing against this "solution" have variously argued that there is no market failure under perfectly competitive innovation (Boldrin and Levine 2008), or that government failures loom as large as market failures (Davidson and Potts

2016a, 2016b). But the argument here is neither a market failure (neoclassical microeconomics) nor a government failure (public choice economics) approach, but rather a Schumpeterian political economy model set in a global entrepreneurial and institutional context. The key insight is that firms compete on innovation, not on price, which implies there are private rents to be captured, and that the value of the property right is limited by the extent of the market. Governments do not create rents in this model; rather, they expropriate them.[8] How do they do that?

Governments as Bandits: Public Theft Is Worse Than Private Theft

In the standard model of intellectual property, government is benevolent, making gifts and protecting the good (inventors) from the bad (pirates and thieves). The malevolent forces are the corporations that can only compete through conspiracy and theft (Lessig 2004) and the consumers who seek products for free (Goodenough and Decker 2009). In this view, governments are very much on the side of intellectual property (Posner 2005). Yet this view is fundamentally misleading because it represents government as largely being on the supply side—that is, as part of the infrastructure that creates intellectual property, which is, of course, technically correct, but ignores government's larger role on the demand side of intellectual property—for example, in public health services—where its incentives are those of a purchaser, albeit one with military and naval assets at its disposal.

Government, in this view, is far from a friend of intellectual property, but its biggest and most devastating predator. Public theft by *governments*, as a simple matter of scale, is a far worse problem than private theft by corporations and citizens. Yet this tends to be opaque, because governments rarely steal outright, but systematically use a complement of tactics to devalue and extract rents from

[8]A corollary of our theory is that governments should be systematically hostile to alternative incentive systems for innovation—such as rewards, bounties, and contracts for research—that economists have long favored (e.g., Wright 1983, Shavell and Ypersele 2001). From a social contract view of innovation, incentives, prizes, and output rewards should be preferred. But from a stationary bandit perspective, these are not viable because there are no monopoly rents for government to capture.

intellectual property in areas where they are on the demand side. So they can appear to be protectors in the courts while a different branch of government (the health ministry, for instance, or department of agriculture) engages in activities that devalue intellectual property through the rest of the world (Wu and Ezell 2016).

Our argument is both that public theft is a bigger problem than conventionally realized and that private theft is a smaller problem. First, private theft concerns are often about the economics of enforcement costs of existing property rights and the rents accruing to particular business models. Music and movie piracy, for instance, has seen an evolutionary response through innovation in business models (e.g., streaming, or product placement), while companies seeking government protection are often effectively seeking protection from competition to their extant business models. Second, corporate theft of ideas is often constrained by the tacitness of technical knowledge, particularly where knowledge is carried in groups, as is typical in manufacturing (Kogut and Zander 1992). Technical piracy as a commercial strategy is often simply not that effective. However, governments can also predate on intellectual property and, often, very effectively. They can seek to devalue intellectual property in order to reduce the prices they pay, or the scope over which they pay, or to increase their range of control over the commercial product, and therefore to improve their bargaining position (Davidson 2015). In biopharma, for instance, governments can act as a single desk-purchaser of a drug, particularly through a schedule of subsidized drugs that allows a government to blockade entry or force sales below cost as a condition of entry for other products. Governments can also demand offsets, such as domestic production, or location of research facilities, regardless of the commercial merit of such decisions. They can also enforce "social license" conditions, demanding free supply of particular products to politically favored or particular needy groups, thus off-loading government responsibility to provide such health care services (Stiglitz 2008: 1694). Governments can also work to maximize the taxation of intellectual property within a jurisdiction by making market entry conditional upon profits being also domiciled, as well as extracting rents from intellectual property through bargaining over regulations (Holcombe and Boudreaux 2015). And of course governments can just outright steal through protection of cronies (who presumably reward governments through side payments) who directly engage in intellectual property theft.

From a perspective of new ideas born global, in which governments are mostly on the demand side of intellectual property, and private appropriation is rarely an effective market strategy, the main predator on intellectual property is public theft by governments. Following Olson (1993), we call these roving bandits. To protect yourself from that magnitude of threat, you grant monopoly rights to taxation in perpetuity to a now stationary bandit. You both benefit from this exchange. We now consider why.

The Stationary Bandit's Offer

The stationary bandit model predicts maximum predation, as a rational monopolist. As Olson (1993: 569) argues, "Exactly the same rational self-interest that makes a roving bandit settle down and provide government for his subjects also makes him extract the maximum possible amount from the society for himself." In this case, the taxes and tributes the government will extract at each point in time from that agent seeking intellectual property protection will be consistent with maximizing the flow of future tax revenues. The holders of intellectual property are allowed to make enough profit that they survive to be taxed again (McGuire and Olson 1996), but are still safe from appropriation by another bandit through tax competition.

Yet from the perspective of the agent seeking protection "the rational monopolization of theft also leaves the bandits' subjects in better shape" (McGuire and Olson 1996: 38). Even though the bandits' subjects are maximally exploited, a situation of uncoordinated competitive theft from all other governments is even worse because it exposes them to total theft, whereas under a monopoly model of theft, in which the stationary bandit preys on the intellectual property producer but as a *monopoly predator*, the subject experiences only partial theft. Of course, "there is little or no production in the absence of a peaceful order" (Olson 1993: 567), and with the protection of a strong bandit, the subject also receives security to invest and to engage in long-term contracting.

The measure and value of this security and service offered to the subject is proportional to the credible threat the stationary bandit can make against the roving bandits. Which is to say that some stationary bandits are better than others. This appears on both sides of the exchange. A strong bandit that can make a credible threat against other bandits (perhaps it has the most powerful naval assets) and

offer maximum protection can also extract a larger rent from the subject, because if they choose some lesser bandit for protection, then the stronger bandit is now a potential predator. So we would expect to see intellectual property gravitate to the protection of the strongest bandit (e.g., the United Kingdom in the 18th and 19th centuries, and the United States in the 20th century to the present). Yet a sufficiently strong bandit will be powerful enough to steal from its own subjects, creating a dynamic inconsistency problem. A strong bandit will therefore also need *a good reputation for self-control* (Barzel 2000: 34), which is why intellectual property industries tend to concentrate in nations such as the United States, Switzerland, Singapore, and other constitutionally secure and well-governed states (sometimes known as tax havens). Furthermore, patents are more highly valued in institutionally stronger jurisdictions where the local government is not a predator (Hall, Thoma, and Torrisi 2007).

Our theory connects the global distribution of intellectual property with the global distribution of tax havens. It is well known that intellectual property is mobile and can be relocated to minimize tax liability. Yet it is also a robust finding that intellectual property tends to locate in relatively high tax jurisdictions (Griffith, Miller, and O'Connell 2011, 2014). This seeming paradox is resolved in our model: it predicts both maximum predation by the host government (i.e., high taxes) in return for extremely strong institutions, particularly self-control. Tax havens, then, are jurisdictions with extremely strong institutions that can effectively protect against the predation of other governments, as well as to credibly promise minimal risk of self-predation (Desai, Foley, and Hines 2006; Dharmapala and Hines 2009). Thus tax havens will also tend to be intellectual property hubs.

Not all countries are bandits all of the time, but most are some of the time. Wu and Ezell (2016) examined price controls on biopharmaceutical products, ranking 56 countries low, moderate, or high in terms of forced price reductions. Only nine countries scored a low rating—Switzerland, Israel, and the United States among them. A broader study by Ezell, Nager, and Atkinson (2016) looked at the extent of free riding on overall innovation policy finding substantial evidence of what they call "innovation mercantilism" through failing to protect intellectual property through weak enforcement, compulsory licensing or outright theft, or through balkanized producer and consumer markets. Indeed it is widely observed that governments

can insist on compulsory licensing to force intellectual property rights holders to reveal information to the government's cronies or through the lobbying efforts of NGOs as a condition of entering a market (Sell and Prakash 2004). Multinational firms can often find themselves targeted for theft by firms in other countries, with their own governments unwilling or unable to help, and so needing to seek private enforcement that may pit them against the foreign government (Yang, Sonmez, and Bosworth 2004).

Intellectual Property as Policy

In the standard economic model, the institution of intellectual property rights solves a market failure problem in the private production of new knowledge, connecting innovation policy to economic growth policy to incentivize socially efficient levels of private domestic investment in new idea creation (Nelson 1959, 1993; Arrow 1962; Romer 1990; Scotchmer 2004). The core economic argument is that the government creation of these monopoly rents is an effective and efficient institutional mechanism and, evaluated from a social welfare perspective, is a good public investment (Gallini and Scotchmer 2002, Romer 2002, Aghion et al. 2005). This argument that an effective intellectual property rights regime is part of a nation's innovation system is a self-contained logic at the level of the nation-state. Because new knowledge is treated as a positive externality, this logic then generalizes to all nations, and indeed offers a model of a global innovation policy based around harmonized and strong intellectual property in order to drive global economic growth and development (Deardorff 1992, Gould and Gruben 1996, Maskus 2000, McCalman 2001). The global context, therefore, merely complicates a nation-state model in which intellectual property rights are economically justified to resolve a market failure problem, into a system of interconnected intellectual property rights systems. This logic then underscores the importance of global governance organizations, such as the World Intellectual Property Organization, to coordinate this complexity.

Yet according to the stationary bandit model, this notion that the global context of intellectual property is an additional complication to what is essentially a national story is a post hoc rationalization invented in the mid-20th century to justify government actions that had already been occurring for many centuries. Rather, the inherent

context of intellectual property rights and protection is, and has always been, global vulnerability to expropriation. From this perspective, the subject's stationary bandit is pitted in conflict in an arena of roving bandits. What is being fought for is the spoils of plunder of the roving bandits and monopoly tax rights for the stationary bandit. Intellectual property is not a form of innovation policy in this view, but rather a species of foreign policy in the proximate instance but ultimately as tax policy for the stationary bandit and retail politics for the roving bandit.

There are several institutional pathways through this potential global Hobbesean jungle. One way is to minimize the importance of intellectual property by developing and promoting commons-based and open-access institutions that broadly seek to weaken intellectual property and push back against closed-access regimes. This can be observed in the mounting intellectual critique of intellectual property. Some economists argue against the costs of monopoly and point to the weakness of the incentives subsequently created (e.g., Boldrin and Levine 2008), emphasizing the rent-seeking aspects, particularly around extensions to existing intellectual property regimes. Legal scholars have sought to create alternative institutions based around common property, led by the work of Richard Stallman (2010), Larry Lessig (2004), and Yochai Benkler (2006). Other law and economic scholars have simply argued the ineffectiveness of the current regimes from the comparative perspective of parts of economic production flourishing under weak or nonexistent intellectual property (Kinsella 2001, Raustilia and Sprigman 2012). This institutional model of open access intellectual property seeks to disarm the roving bandits by, in effect, making everyone a bandit, eliminating the ability of stationary bandits to seek monopoly exploitation.

An alternative approach harnesses the costs and benefits of global reputation. Nations are expected to self-moderate because of the broader costs of a poor reputation, or to cooperate in order to benefit from a reputation as a good global citizen (for instance, via preferential access to trade or security treaties). Ezell, Nager, and Atkinson (2016) have developed a measure to identify countries as contributors or detractors. There are strong moral or natural rights arguments in favor of intellectual property that are particularly effective on the supply or incentive side. However, these tend to substantially weaken in the case when monopoly gets in the way of public goods provision, such as health care (De George 2005), revealing intellectual property

rights to be a political institution that can be changed by political expediency (Sened 1997).[9]

Conclusion

The standard *economic model* of intellectual property is market failure, corrected with a monopoly rent; the standard *legal model* of intellectual property is a property right bound up with a moral right, issued and enforced by government, and in which most theft is private; and the standard *political economy model* of intellectual property is the expression of a social contract within a nation-state. We have proposed an alternative view of intellectual property that makes none of these claims. Instead, we have an *anti-social-contract model* (Olson 1993) in which governments are bandits; a legal model based around taxation, and in which most theft is public; and an economic model in which intellectual property is a product of entrepreneurial discovery in a global context, and is thus a story of trade.

Intellectual property, in our new approach, is not some quirky institution only optimized in the late 20th century, once governments got serious about growth policy, and moreover that might be replaced by more efficient mechanisms of discovery, such as public science or prizes. Rather, intellectual property, we argue, is an inevitable evolutionary product of a global trading economy in a world of powerful predators. It is the result of a grand bargain struck between a citizen, seeking to maximize the value of the ideas associated with their property, and a powerful state, maximizing the value of their share—formalized as a monopoly tax right—by extending protection as they trade near and far. Intellectual property thus arises in the normal course of entrepreneurial activity in pursuit of the creation of ideas with value. These are vulnerable on a global scale, and the main predators are governments or their cronies, which we call roving bandits following Olson (1993). Intellectual property emerges when a roving bandit finds it worthwhile to become a stationary bandit by protecting the idea from competitive theft from other governments and, in the process, becoming a monopoly exploiter. This monopoly exploitation

[9]A related instance occurred in Australia, when the antitobacco lobby successfully blocked the use of company trademarks, forcing "plain packaging" and devaluing the tobacco company's intellectual property by justifying the legislative change as a public health issue (Davidson 2015).

aspect is widely overlooked in the standard model of a benevolent government, but the implication, also following from the theory (McGuire and Olson 1996), is that the stationary bandit will tax their subjects to extract the maximum rent they will bear.

We have only sketched the underlying idea and are currently developing this into a testable formal (Hotelling type) model. We can nevertheless immediately see several basic predictions that we hope to test using historical data. Specifically, our model predicts a relationship between ability of a stationary bandit to project power (and willingness to use this to open markets) and the growth of the intellectual property sector, which in turn implies a non-Ricardian theory of the global pattern of trade; that a nation's innovation system is more significantly shaped by tax policy than science and technology policy; and that the stationary bandit will have a strong incentive to have a noncorrupt and globally competitive tax system. Our future research program aims to test our theory (and its particular predictions about institutional quality, trade patterns, and relative government priorities) against the alternative hypothesis that intellectual property rights emerged in order to solve the market failure problem in private investment in innovation.

Rather, the key starting insight here is that governments predate on intellectual property. Where they protect it, and seek to do so globally—as the United States does when tying intellectual property agreements into trade treaties or other foreign policy sanctions—they do so not out of moral respect for the creative rights of its citizens, but because the offer of protection maximizes the government's future tax stream (the corollary is that U.S. citizens are taxed by the U.S. government wherever they are domiciled). But where they predate directly, through theft or various mechanisms to minimize its value, they do so because governments are most everywhere the predominant consumers of intellectual property—examples are in socialized health care, infrastructure, media and communications, and defense, all of which are technology intensive. The global distribution of intellectual property, and of firms and industries that are heavily reliant on it, seems to conform to our model's predictions. Intellectual property intensive industries predominantly locate in institutionally robust tax havens, such as the United States, Switzerland, and more recently Singapore and Ireland.

Finally, our approach argues that the societal benefits of intellectual property are not the product of some far-thinking design by a

benevolent government, which is the entire focus of the standard model; rather, it is a corollary of the "hidden hand theorem," in which the stationary bandit, who is "a rational self-interested actor with unquestioned coercive power and has an encompassing and stable interest in the domain over which the power is exercised, . . . is led to act in ways that are, to a surprising degree, consistent with the interests of society and of those subject to power" (McGuire and Olson 1996: 73). Innovation policy in the form of intellectual property institutions is a self-organizing product of the invisible hand and does not actually require any more than government self-interest in maximizing its tax revenue through the creation of high quality property rights protection institutions (Cooter 2005). What does need protection, in our new view, are intellectual property producing firms and individuals predated on by governments. These actions constrain the profitability of these firms, and therefore their ability to reinvest and grow, which of course includes the discovery of further intellectual property.

References

Aghion, P.; Bloom, N.; Blundell, R.; Griffith, R.; and Howitt, P. (2005) "Competition and Innovation: An Inverted U-Relationship." *Quarterly Journal of Economics* 120 (2): 701–28.

Alchian, A., and Allen, W. R. (1977) *Exchange and Production: Competition, Coordination, and Control.* 2nd ed. Belmont, Calif.: Wadsworth.

Arrow, K. (1962) "Economic Welfare and the Allocation of Resources for Invention." In R. Nelson (ed.), *The Rate and Direction of Inventive Activity,* 609–26. Princeton: Princeton University Press.

Barzel, Y. (2000) "Property Rights and the Evolution of the State." *Economics of Governance* 1 (1): 25–51.

_____ (2002) *A Theory of the State*. New York: Cambridge University Press.

Baumol, W. (2002) *The Free-Market Innovation Machine*. Princeton: Princeton University Press.

Benkler, Y. (2006) *The Wealth of Networks*. New Haven: Yale University Press.

Boldrin, M., and Levine, D. (2008) "Perfectly Competitive Innovation." *Journal of Monetary Economics* 55 (3): 435–53.

_____ (2013) "The Case against Patents." *Journal of Economic Perspectives* 27 (1): 3–22.

Cooter, R. (2005) "Innovation, Information, and the Poverty of Nations." *Florida State University Law Review* 33: 373–93.

Davidson, S. (2015) "Antimarket Bias and Its Threat to Property." *SEANET Perspectives* No. 4.

Davidson, S., and Potts, J. (2016a) "A New Institutional Approach to Innovation Policy." *Australian Economic Review* 49 (2): 200–7.

_____ (2016b) "Social Costs and the Institutions of Innovation Policy." *Economic Affairs* 36 (3): 282–93.

Deardorff, A. (1992) "Welfare Effects of Global Patent Protection." *Economica* 59: 35–51.

De George, R. (2005) "Intellectual Property and Pharmaceutical Drugs: An Ethical Analysis." *Business Ethics Quarterly* 15 (4): 549–75.

Desai, M.; Foley, C.; and Hines, J. (2006) "The Demand for Tax Haven Operations." *Journal of Public Economics* 90 (3): 513–31.

Dharmapala, D., and Hines, J. (2009) "Which Countries Become Tax Havens?" *Journal of Public Economics* 93 (9): 1058–68.

Dourado, E., and Tabarrok, A. (2015) "Public Choice Perspectives on Intellectual Property." *Public Choice* 163 (1–2): 129–51.

Ezell, S.; Nager, A.; and Atkinson, R. (2016) "Contributors and Detractors: Ranking Countries Impact on Global Innovation." Available at https://itif.org/publications/2016/01/20/contributors-and-detractors-ranking-countries%E2%80%99-impact-global-innovation.

Gallini, N., and Scotchmer, S. (2002) "Intellectual Property: When is it the Best Incentive System?" *Innovation Policy and the Economy* 2: 51–78.

Gilbert, R., and Shapiro, C. (1990) "Optimal Patent Length and Breadth." *RAND Journal of Economics* 21: 106–12.

Goodenough, O., and Decker, G. J. (2009) "Why Do Good People Steal Intellectual Property?" In M. D. A. Freeman and O. R. Goodenough (eds.), *Law, Mind and Brain*, chap. 15. Burlington, Vt.: Ashgate.

Gould, D., and Gruben, W. (1996) "The Role of Intellectual Property Rights in Economic Growth." *Journal of Development Economics* 48 (2): 323–50.

Griffith, R.; Miller, H.; and O'Connell, M. (2011) "Corporate Taxes and the Location of Intellectual Property." Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1865574.

_____ (2014) "Ownership of Intellectual Property and Corporate Taxation." *Journal of Public Economics* 112: 12–23.

Haber S. (2016) "Patents and the Wealth of Nations." *George Mason Law Review* 23 (4): 811–35.

Hall, B.; Helmers, C.; Rogers, M.; and Sena, V. (2014) "The Choice between Formal and Informal Intellectual Property: A Review." *Journal of Economic Literature* 52 (2): 375–423.

Hall, B.; Thoma, G.; and Torrisi, S. (2007) "The Market Value of Patents and R&D: Evidence from European Firms." *Academy of Management Proceedings* 2007 (1): 1–6.

Holcombe, R., and Boudreaux, C. (2015) "Regulation and Corruption." *Public Choice* 164 (1–2): 75–85.

Kinsella, N. S. (2001) "Against Intellectual Property." *Journal of Libertarian Studies* 15 (2): 1–53.

Kogut, B., and Zander, U. (1992) "Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology." *Organization Science* 3 (3): 383–97.

Kremer, M. (1998) "Patent Buyouts: A Mechanism for Encouraging Innovation." *Quarterly Journal of Economics* 113 (4): 1137–67.

Landes, W., and Posner, R. (2003) *The Economic Structure of Intellectual Property Law*. Cambridge: Harvard University Press.

Lessig, L. (2004) *Free Culture*. New York: Penguin Press.

Maskus, K. (2000) *Intellectual Property Rights in the Global Economy*. Washington: Peterson Institute.

McCalman, P. (2001) "Reaping What You Sow: An Empirical Analysis of International Patent Harmonization." *Journal of International Economics* 55 (1): 161–85.

McGuire, M., and Olson, M. (1996) "The Economics of Autocracy and Majority Rule: The Invisible Hand and the Use of Force." *Journal of Economic Literature* 34 (1): 72–96.

Mowery, D. (1995) "The Practice of Technology Policy." In Stoneman, P. (ed.), *Handbook of the Economics of Innovation and Technological Change*, 513–57. Oxford: Blackwell.

Mowery, D., and Rosenberg, N. (1989) *Technology and the Pursuit of Economic Growth*. New York: Cambridge University Press.

Mulligan, C., and Shleifer, A. (2005) "The Extent of the Market and the Supply of Regulation." *Quarterly Journal of Economics* 120 (4): 1445–73.

Nelson, R. (1959) "The Simple Economics of Basic Scientific Research." *Journal of Political Economy* 67 (3): 297–306.

_____, ed. (1993) *National Innovation Systems*. Oxford: Oxford University Press.

Nordhaus, W. (1969) *Invention, Growth and Welfare*. Cambridge: MIT Press.

Olson, M. (1965) *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge, Mass.: Harvard University Press.

_____ (1982) *The Rise and Decline of Nations*. New Haven: Yale University Press.

_____ (1993) "Dictatorship, Democracy, and Development." *American Political Science Review* 87 (3): 567–76.

_____ (1995) "Why the Transition from Communism Is So Difficult." *Eastern Economic Journal* 21 (4): 437–61.

Ostrom, E., and Hess, C. (2006) *Understanding Knowledge as a Commons*. Cambridge: MIT Press.

Phelps, E. (2013) *Mass Flourishing*. Princeton: Princeton University Press.

Posner, R. (2005) "Intellectual Property: The Law and Economics Approach." *Journal of Economic Perspectives* 19 (2): 57–73.

Raustiala, K., and Sprigman, C. (2012) *The Knockoff Economy*. Oxford: Oxford University Press.

Romer, P. M. (1990) "Endogenous Technological Change." *Journal of Political Economy* 98 (5, Part 2): S71–S102.

_____ (2002) "When Should We Use Intellectual Property Rights?" *American Economic Review* 92 (2): 213–16.

Schumpeter, J. ([1912] 1934) *Theory of Economic Development*. Cambridge: Belknap Press.

_____ (1942) *Capitalism, Socialism and Democracy*. New York: Harper & Bros.

Scotchmer, S. (2004) *Innovation and Incentives*. Cambridge: MIT Press.

Sell, S., and Prakash, A. (2004) "Using Ideas Strategically: The Contest between Business and NGO Networks in Intellectual Property Rights." *International Studies Quarterly* 48 (1): 143–75.

Sened, I. (1997) *The Political Institution of Private Property*. New York: Cambridge University Press.

Shavell, S., and Ypersele, T. (2001) "Rewards versus Intellectual Property Rights." *Journal of Law and Economics* 44 (2): 525–47.

Stallman, R. M. (2010) *Free Software, Free Society: Selected Essays of Richard M. Stallman*. 2nd ed. Boston: GNU Press.

Steinmueller, W. E. (2010) "Economics of Technology Policy." In B. Hall and N. Rosenberg (eds.), *Handbook of Economics of Innovation*, vol. 2, 1181–1218 Amsterdam: Elsevier.

Stiglitz, J. (2008) "Economic Foundations of Intellectual Property Rights." *Duke Law Journal* 57 (6): 1693–724.

Wright, B. (1983) "The Economics of Invention Incentives: Patents, Prizes and Research Contracts." *American Economic Review* 73 (4): 691–707.

Wu, J., and Ezell, S. (2016) "How National Innovation Policies Impact on Global Biopharma Innovation: A Worldwide Ranking." Washington: ITIF. Available at www2.itif.org/2016-national-policies-global-biopharma.pdf?_ga=1.105395782.499446021.1460352948.

Yang, D.; Sonmez, M.; and Bosworth, D. (2004) "Intellectual Property Abuses: How Should Multinationals Respond?" *Long Range Planning* 37 (5): 459–75.

Margaret Thatcher's Privatization Legacy

Chris Edwards

Economic policy has taken an antimarket turn in recent years, with many nations increasing regulations, running large deficits, and embracing repeated stimulus actions by central banks. There is, however, one good-news story in economic policy that is often overlooked: the ongoing privatization revolution that has swept the world since the 1980s. Governments in more than 100 countries have moved thousands of state-owned businesses to the private sector. Airlines, railroads, postal services, electric utilities, and many other types of businesses valued at more than $3.3 trillion have been privatized over the last three decades (Megginson 2015).

The revolution was launched by Margaret Thatcher, British prime minister from 1979 to 1990. She came to power determined to revive the stagnant British economy with market-based reforms. Her government deregulated, cut marginal tax rates, repealed exchange controls, and tamed militant labor unions. But it was privatization that became her most important and enduring economic legacy. Thatcher popularized the word *privatization,* and she oversaw the sale of many major businesses, including British Airways, British Telecom, British Steel, and British Gas.

Spurred by the success of Thatcher's reforms, privatization swept through developed and developing nations in Europe, Latin America, and elsewhere. Other nations followed Britain's lead

---

Cato Journal, Vol. 37, No. 1 (Winter 2017).

Chris Edwards is Editor of www.DownsizingGovernment.org at the Cato Institute.

because of "disillusionment with the generally poor performance of state-owned enterprises and the desire to improve efficiency of bloated and often failing companies," noted a report on privatization by the Organisation for Economic Cooperation and Development (OECD 2003: 21).

Privatization has had a huge effect on the global economy. It has spurred economic growth and improved living standards as privatized businesses cut costs, increased service quality, and innovated. The reforms also "massively increased the size and efficiency of the world's capital markets," argues William Megginson in his book, *The Financial Economics of Privatization* (2005: 4). Many of the largest share offerings in world history have been privatizations, and a large share of global stock market capitalization is from privatized companies.

It is inspiring to look back at Margaret Thatcher's privatization triumph. But for U.S. policymakers, there are practical lessons as well. Many types of businesses that Britain privatized are still partly or wholly in government hands in the United States, including airports, seaports, postal services, air traffic control, electric utilities, and passenger rail. To tackle lackluster U.S. growth, policymakers should pursue privatization in order to increase productivity and inject more dynamism into the economy.

Britain Blazes the Trail

In a 1969 essay, management expert Peter Drucker said that politicians in the 20th century had been "hypnotized by government . . . in love with it and saw no limits to its abilities" (1969: 4). But he said that the love affair was coming to an end as the mismanagement of state-owned businesses was becoming more apparent everywhere. Drucker called for a "reprivatization" of government activities. But he was ahead of his time, as many developed economies struggled through years of stagflation before new leaders emerged to begin making pro-market reforms.

Margaret Thatcher was elected Conservative Party leader in 1975, and her party gained a parliamentary majority in 1979. Prime Minister Thatcher came into office promising to "denationalize" the government-dominated economy. However, she faced numerous crises her first few years in office that limited her privatization efforts, including a deep recession, high inflation, labor union strife, and the Falklands War.

At first, Thatcher and the Conservatives were politically cautious about privatization, and they did not have a detailed agenda to pursue it. But they learned as they went, and some early successes generated momentum for further reforms. One early reform was the popular "Right to Buy" law, which allowed people to buy the government-owned "council" houses that they lived in. With that successful reform, the share of British households in government council housing plunged from 31 percent in 1981 to just 7 percent today (Department for Communities and Local Government 2016: Annex Table 1.1).

With the economy recovering in the early 1980s, and with Thatcher reelected with a large majority in 1983, the British privatization program kicked into high gear. Campaigning in 1983, the Conservatives promised widespread privatizations, and that created a strong mandate for them to move boldly after their landslide election victory.

Thatcher had a strong personal belief in privatization. Privatization was crucial for "reversing the corrosive and corrupting effects of socialism," she said, and central to "reclaiming territory for freedom" (Thatcher 1993: 676). The purpose of privatization was to ensure "the state's power is reduced and the power of the people enhanced" (Thatcher 1993: 676). Thatcher was heavily influenced by economist F. A. Hayek, as well as by her key adviser Keith Joseph.

Thatcher blazed the trail, but there were some precedents for her reforms. In the 1950s, the British Conservatives privatized some industries—including the steel industry—that had been nationalized by the previous Labour government. And in the 1950s and 1960s, West German political leaders Konrad Adenauer and Ludwig Erhard began "denationalizing" industries to improve efficiency and broaden public share ownership. The German government, for example, sold a majority stake in Volkswagen in a public share offering in 1961.

Another influence on Thatcher's government was a Canadian privatization effort. Some of Thatcher's key advisers, including Alan Walters, were familiar with the privatization of the British Columbia Resources Investment Corporation in 1979 (Milke 2012). That process included a distribution of free shares to all citizens in the largest share offering in Canadian history to that date. A 1980 book describing that reform was the first with the word *privatization* in its title (Ohashi and Roth 1980).

Numerous privatization methods have been used in Britain and subsequently in reforms elsewhere. The dominant method has been share issue privatization. The government proceeds with an initial public offering (IPO) of all or a portion of company shares, which is usually followed by a later sale of the remaining shares. British Aerospace was privatized in 1981 with an IPO of 52 percent of its shares, with remaining shares unloaded in later years.

The British Telecom (BT) IPO in 1984 was a mass share offering, which "did more than anything else to lay the basis for a share-owning popular capitalism in Britain," said Thatcher (1993: 680). The government ran high-profile television ads to encourage the purchase of BT shares, and more than two million citizens participated in the largest share offering in world history to that date.

Selling the 250,000-worker BT was a bold decision, and its success generated momentum for further reforms. The OECD (2003: 24) called the BT privatization "the harbinger of the launch of large-scale privatisations" internationally. In subsequent years, the British government proceeded with large public share offerings in British Gas, British Steel, electric utilities, and other companies. In the gas privatization, two million individuals who bought shares had never held corporate equities before (Moore 2016: 211).

A second privatization method is a direct sale or trade sale, which involves the sale of a company to an existing private company through negotiations or competitive bidding. For example, the British government sold Rover automobiles and Royal Ordnance to British Aerospace. Other privatizations through direct sale included British Shipbuilders, Sealink Ferries, and The Tote.

A third privatization method is an employee or management buyout. Britain's National Freight Corporation was sold to company employees in 1982, and London's bus services were sold to company managers and employees in 1994. Management and employee buyouts were also popular in Eastern Europe after the fall of communism. The mass issuance to citizens of free or low-cost share vouchers was also a popular privatization method in Eastern Europe.

In most cases, British privatizations went hand-in-hand with reforms of regulatory structures. The government understood that privatization should be combined with open competition when possible. British Telecom, for example, was split from the post office and set up as an arms-length government corporation before shares were

sold to the public. Then, over time, the government opened BT up to competition.

The British government opened up intercity bus services to competition beginning in 1980. That move was followed by the privatization of state-owned bus lines, such as National Express. Numerous British seaports were privatized during the 1980s, and the government also reformed labor union laws that had stifled performance in the industry.

Studies in Britain and elsewhere have found that opening industries to competition is important to maximizing the productivity gains from privatization (Parker 2004: 12). When possible, privatization should be paired with the removal of entry barriers because open competition is preferable to either government or private monopoly. However, British experience also shows that even when industries have natural monopoly elements, privatization combined with improved regulatory oversight spurs gains to efficiency and transparency.

After a leadership challenge from within her party, Margaret Thatcher resigned as prime minister in 1990. Privatization, however, lived on. John Major's Conservative government, for example, privatized British Rail. Tony Blair's Labour government privatized air traffic control. And David Cameron's Conservative government privatized the Royal Mail.

Table 1 lists the major British privatizations since 1979, which was compiled from various sources including Rhodes, Hough, and Butcher (2014). It shows company names at the time of privatization and the year that the first portion of each business was privatized. For less-familiar companies, the industry is noted in parentheses.

Effects of Privatization

Privatization transformed the British economy. Bloated workforces at many formerly state-owned firms were slashed. Employment in the electricity and gas industries was cut in half between the mid-1980s before privatization and mid-1990s after privatization (Lloyd and Nevala 2007: 14). As workforces declined, labor productivity increased. Labor productivity roughly doubled in the electricity and gas industries in the decade after privatization (Department of Energy and Climate Change 2015: 10). Productivity

TABLE 1
Major British Privatizations

Year	Company or Asset
1979	British Petroleum, government council housing
1981	British Aerospace, Cable & Wireless, British Sugar Corporation
1982	Britoil, National Freight Corporation, Amersham International (radioactive materials)
1983	Associated British Ports (seaports), British Shipbuilders, British Transport Hotels
1984	British Telecom, Jaguar, Enterprise Oil, Sealink Ferries
1986	British Gas, National Bus Company
1987	British Airways, British Airports Authority, Rolls Royce, Rover (trucks), Royal Ordnance (military products), Royal Dockyards
1988	British Steel, Rover (automobiles), National Express (intercity busing)
1989	The 10 regional water agencies, Short Brothers (aircraft)
1990	National grid and the 12 regional electricity distribution firms, Girobank
1991	National Power, PowerGen, Scottish Power, Scottish Hydro, Forth Ports (seaports)
1992	Trust seaports, motorway service stops, British Technology Group
1993	Northern Ireland electricity
1994	British Rail, British Coal, London bus services
1996	British Energy (nuclear generation), AEA Technology (nuclear research)
2001	National Air Traffic Services (NATS)
2003	Qinetiq (defense technology)
2006	British Nuclear Fuel
2009	UKAEA Limited (environmental management)
2011	The Tote (retail betting shops)
2013	Royal Mail
2015	Eurostar rail service

Notes: A portion of British Petroleum had been sold in 1977 as part of a deal with the International Monetary Fund. The Thatcher government sold the rest of the shares beginning in 1979. Also, Britain has sold its 40 percent stake in Eurostar, but the rest of the firm is held by the French state-owned rail firm.

increases were particularly pronounced for firms in competitive industries such as British Steel, British Coal, British Telecom, British Airways, and Associated British Ports.

Just knowing that privatization was coming spurred reforms in many companies. British Steel chopped its workforce and improved its productivity leading up to its 1988 privatization, as did British Airways before its 1987 privatization. After privatization, with revenues and profitability rising, British Airways increased its employment to serve expanding markets. That pattern of cost cutting, increased efficiency, and then growth is common among privatized firms.

British consumers benefited as privatization and competition reduced prices and improved service quality. A Treasury study found that real prices after a decade of privatization had fallen 50 percent for telecommunications, 50 percent for industrial gas, and 25 percent for residential gas (HM Treasury 1997: 13). And a decade after electricity privatization, real prices were down more than 25 percent (Parker 2004: 14). The environment gained from the electricity reform as well because the privatized industry moved rapidly to replace coal as a fuel source with natural gas.

The Treasury study (1997: 14) found that "most indicators of service quality have improved" in privatized businesses. Economist David Parker (2004: 16) found, "There is no substantial evidence that lower manning and price reductions in the public utilities have been at the expense of service quality." The share of British Telecom service calls completed within eight days soared from 59 percent to 97 percent in the decade after privatization (Holder 1998: 20). Before privatization, it had taken months and sometimes a bribe to get a new telephone line. By various measures, safety also improved in the privatized industries, including gas, electricity, and water (Bourne and Knox 2013).

Millions of individuals gained from investing in the privatized companies. The government made share offerings appealing to small investors, which fit with Thatcher's belief in "popular capitalism." She wanted to create a "capital-owning democracy . . . a state in which people own houses, shares, and have a stake in society, and in which they have wealth to pass on to future generations" (Yergin and Stanislaw 1998: 97). Under Thatcher, the share of British citizens owning equities soared from 7 percent to 25 percent (Simon 2013).

The British experience in improving industry performance from privatization has been repeated in many other countries. An OECD report reviewed the research and found "overwhelming support for the notion that privatization brings about a significant increase in the profitability, real output and efficiency of privatised companies" (OECD 2003: 9). And a review of dozens of academic studies in the *Journal of Economic Literature* concluded that privatization "appears to improve performance measured in many different ways, in many different countries" (Megginson and Netter 2001: 25). Further academic assessments of privatization are summarized in Edwards (2016).

Rail and Water Controversies

Despite the general success of British privatization, some of the reforms were quite controversial at the time, such as the rail and water privatizations of the 1990s.

State-owned British Rail had been experiencing a long-term decline in its transportation market share, and it was consuming large taxpayer subsidies. In 1994, the government split up the company and privatized separate pieces: Railtrack took control of tracks and stations, 3 separate firms took control of rail freight, and 25 firms received franchises to operate passenger services. The British rail industry went from being vertically integrated to being split into pieces.

In the late 1990s, a few high-profile rail accidents raised concerns about the industry's new structure. Some of the accidents may have been due to insufficient track maintenance in the years before and the years after privatization. Those problems prompted the renationalization of Railtrack in 2002 as Network Rail.

Some experts believe that undoing the industry's vertical integration was a mistake. Before nationalization in the 1940s, British passenger rail was vertically integrated as four regional private rail firms owning both track and rolling stock. So there continues to be uncertainty about the industry's optimal structure.

Nonetheless, passenger rail has flourished since privatization. Productivity has substantially improved, with passenger journeys per employee increasing 37 percent (Association of Train Operating Companies 2013) since the reforms. And, unlike elsewhere in Europe, total rail ridership in Britain has soared. By 2014, total

passenger trips had more than doubled since privatization, from 740 million to 1.5 billion (Department of Trade and Investment 2014). Rail ridership is now hitting levels not seen since the early 1920s (Vranich 2004: 144).

Despite the growth in passengers, the on-time performance of British passenger rail improved in the years immediately following privatization (Merkert and Nash 2006: 82). And despite a few high-profile accidents in the 1990s, the overall safety record of British rail has steadily improved since privatization (Merkert and Nash 2006: 83). Surveys find fairly high levels of customer satisfaction with British rail travel today (Association of Train Operating Companies 2013).

In 2013, the European Commission found that Britain's railways were the "most improved" in all of Europe since the 1990s and were second only to Finland's in customer satisfaction (Department of Trade and Investment 2014). U.S. passenger rail expert Joseph Vranich (2004: 147) noted that "private operators [in Britain] have demonstrated more initiative, imagination, and visionary planning than state-run British Rail did in its prime or Amtrak does today." In sum, British rail reform has been a success, not the failure that some critics have claimed.

Turning to water industry reforms, the government privatized 10 regional water and sewer agencies in 1989 and created a new regulatory authority to oversee them. After the reforms, people complained that water prices rose. And, indeed, water prices did initially rise. But those increases stemmed from new private owners increasing capital investment to modernize very old government infrastructure. Privatization gave the companies access to the capital they needed to upgrade.

Put another way, water prices had been kept artificially low under government ownership, which led to underinvestment and inefficient overconsumption. After increases in the first six years following privatization, British water prices have risen just 9 percent in real terms over the past two decades (National Audit Office 2015: 7).

Efficiency and service quality have increased in the British water industry since privatization. Wasteful leaks have fallen by one-third since privatization, supply interruptions are down, and the number of customers with low water pressure has plummeted (Ofwat 2006, Day 2012, and National Audit Office 2015). Drinking water quality has improved, and pollution has fallen. In sum, water service privatization has increased both efficiency and environmental stewardship.

Global Influence

Since Margaret Thatcher got the ball rolling in 1979, more than 100 countries have privatized many thousands of state-owned businesses. In France, the Jacques Chirac government sold 22 major companies in 1986 and 1987 (Megginson 2005: 17). Then in the 1990s and 2000s, both conservative and socialist governments in France continued to privatize. The number of companies in which the French government holds a majority stake has plunged from 3,000 in the early 1990s to about 1,500 mainly smaller companies today (Brauninger 2013, 2015).

In New Zealand, a Labour government elected in 1984 privatized dozens of state-owned companies including airports, banks, energy companies, forests, and the national airline and telecommunications companies. In Australia, a series of governments privatized dozens of companies in the 1990s and 2000s, generating proceeds of more than $100 billion.

During the 1980s and 1990s, Canada privatized more than 50 major businesses, including electric utilities, energy companies, the national railway, and the national airline (Boardman and Vining 2012). Perhaps Canada's most innovative privatization was the 1996 transfer of its air traffic control (ATC) system to a nonprofit corporation, Nav Canada. In recent years, the company has become a global leader in ATC innovation and technologies. The system is self-financing, raising revenues from charges on aviation users. Nav Canada has cut its workforce 30 percent since privatization, even though it is handling 50 percent more traffic (Robyn 2015).

Privatization swept through many developing nations. In Latin America, Chile, Mexico, and Panama had particularly large and successful privatization programs. Mexico, for example, slashed the number of state-owned firms from 1,155 in the early 1980s to just 210 by the early 2000s (Chong and López-de-Silanes 2014: 8). In Eastern Europe, huge privatizations were pursued after the fall of communism, and the government share of total economic output in that region fell from about three-quarters in 1990 to about one-quarter today (Borrmann et al. 2013: 18).

Privatization has attracted opposition from the public in many countries, but very rarely have reforms been undone once they have been put in place, at least in the developed nations. In Canada, for example, none of the more than 50 major privatizations have been reversed. The

reason is that privatization simply works, and so reforms have generally lasted through both liberal and conservative governments.

Today, many countries have privatized the "lowest hanging fruit." But there is much left to sell, and global privatization is continuing at a robust pace. Over the past four years, governments worldwide have sold an average $203 billion of state-owned businesses annually (Megginson 2015). China is now the largest privatizer, but some developed nations continue to pursue reforms as well.

What about the United States? Despite the global success of privatization, reforms have largely bypassed our federal government. President Ronald Reagan's administration explored privatizing the U.S. Postal Service, Amtrak, the Tennessee Valley Authority, the air traffic control system, and federal land, but those efforts stalled. President Bill Clinton's administration was more successful: it oversaw the sale of the Alaska Power Administration, the Elk Hills Naval Petroleum Reserve, the U.S. Enrichment Corporation, and Intelsat.

But little action on privatization has been pursued since then, even though Britain and other countries have shown that postal systems, passenger rail, electric utilities, air traffic control, and other "public" services could be run better privately. The same is true for numerous business activities run by U.S. state and local governments, such as seaports and airports.

The United States has always been a land teeming with bold entrepreneurs. Privatization would allow those innovators to inject fresh capital, new ideas, and dynamism into a range of industries currently stifled by political control and bureaucracy.

References

Association of Train Operating Companies (2013) "Growth and Prosperity: How Franchising Helped Transform the Railway into a British Success Story" (July). Available at ATOC.org.

Boardman, A., and Vining, A. (2012) "A Review and Assessment of Privatization in Canada." School of Public Policy, University of Calgary (January).

Borrmann, J.; Alt, R.; Helmenstein, C.; and Berrer, H. (2013) "The Privatisation Goldmine." New Direction Foundation, Brussels (June).

Bourne, R., and Knox, T. (2013) "What Did Privatisation Do for Us?" Centre for Policy Studies, *Growth Bulletin* No. 28 (April 11).

Brauninger, D. (2013) "Privatisation in the Euro Area: Differing Attitudes Towards Public Assets." Research Briefing, Deutsche Bank (August 20).

_____ (2015) "Privatisation in the Euro Area: Governments Should Grasp Opportunities." Research Briefing, Deutsche Bank (July 31).

Chong, A., and López-de-Silanes, F. (2014) "Privatization in Mexico." Inter-American Development Bank (August).

Day, G. (2012) "An Insider's View of U.K. Privatization." *Global Water Intelligence* 3 (1) (January).

Department for Communities and Local Government (2016) "English Housing Survey: Headline Report 2014–2015" (February).

Department of Energy and Climate Change (2015) "U.K. Energy Sector Indicators 2015" (October 29).

Department of Trade and Investment (2014) "The U.K. Rail Industry: A Showcase of Excellence."

Drucker, P. F. (1969) "The Sickness of Government." *The Public Interest* 14 (Winter): 3–23.

Edwards, C. (2016) "Options for Federal Privatization and Reform Lessons from Abroad." Cato Institute Policy Analysis No. 794 (June 28).

Her Majesty's Treasury (1997) "Implementing Privatisation: The UK Experience."

Holder, S. (1998) "Privatisation and Competition: The Evidence from Utility and Infrastructure Privatisation in the UK." Presented at Twelfth Plenary Session of the OECD Advisory Group on Privatisation, Helsinki (September 17–18).

Lloyd, P., and Nevala, A. (2007) "The Employment Impact of the Opening of Electricity and Gas Markets." European Commission (March).

Megginson, W. L. (2005) *The Financial Economics of Privatization*. Oxford: Oxford University Press.

_____ (2015) "Privatization Trends and Major Deals in 2014 and Two-Thirds 2015." *The PB Report 2014/2015. Privatization Barometer*. Available at www.privatizationbarometer.net.

Megginson, W. L., and Netter, J. M. (2001) "From State to Market: A Survey of Empirical Studies on Privatization." *Journal of Economic Literature* 39 (2): 321–89.

Merkert, R., and Nash, C. (2006) "The Restructuring of the Rail System in Britain: An Assessment of Recent Developments." In *The Railways, the Market and the Government,* 71–104. London: Institute of Economic Affairs.

Milke, M. (2012) "The Iron Lady for a New Generation: Why Margaret Thatcher Mattered." *Fraser Forum*, Fraser Institute (March/April).

Moore, C. (2016) *Margaret Thatcher: The Authorized Biography, At Her Zenith in London, Washington and Moscow*. New York: Alfred A. Knopf.

National Audit Office. (2015) "The Economic Regulation of the Water Sector" (October 14).

OECD. (2003) *Privatising State-Owned Enterprises: An Overview of Policies and Practices in OECD Countries.* Paris: OECD Publishing.

Ofwat. (2006) "The Development of the Water Industry in England and Wales."

Ohashi, T. M., and Roth, T. P., eds. (1980) *Privatization: Theory and Practice*. Vancouver, British Columbia: Fraser Institute.

Parker, D. (2004) "The UK's Privatisation Experiment." CESifo (Munich) Working Paper No. 1126 (February).

Rhodes, C.; Hough, D.; and Butcher, L. (2014) "Privatisation." Research Paper 14/61, British House of Commons Library (November 20).

Robyn, D. (2015) "Alternative Governance Models for the Air Traffic Control System." Brookings Institution (April 6).

Simon, E. (2013) "Thatcher's Legacy: How Has Privatisation Fared?" *Telegraph* (April 12).

Thatcher, M. (1993) *The Downing Street Years*. New York: HarperCollins.

Vranich, J. (2004) *End of the Line: The Failure of Amtrak Reform and the Future of America's Passenger Trains*. Washington: American Enterprise Institute.

Yergin, D., and Stanislaw, J. (1998) *The Commanding Heights*. New York: Free Press.

Have Multilateral Conventions Lowered Bribery Around the World?

Mushfiq Swaleheen and Marcus T. Allen

Corruption is generally defined as the use of public office for private gains (Bardhan 1997). Jain (2001) provides an overview of the agency model of the origin and spread of corruption in an economy. The economy consists of three groups of actors: firms and households, government leaders, and appointed public officials. The firms and households are the principals who are utility/wealth maximizers. They employ two groups of agents: government leaders and public officials. The government leaders formulate the regulatory laws and processes in the country. The appointed public officials interpret, implement, and uphold the regulatory laws and processes. Acceptance of the position of a government leader or appointed public official indicates that the incumbent has agreed that the pay is sufficient reward for his/her effort. The principals' well-being is impacted by actions (or inactions) of their agents.

Firms and households competitively lobby government leaders to align regulations in their favor. They also seek to influence the interpretation and implementation of these regulations by the bureaucracy, judiciary, and law enforcement to get a competitive edge over rivals. Such rent seeking has an efficiency cost for the economy (Tullock 1993, Krueger 1974, Bhagwati 1982). However, as long as it

---

Cato Journal, Vol. 37, No. 1 (Winter 2017).

Mushfiq Swaleheen is Associate Professor of Economics, and Marcus T. Allen is Eminent Scholar and Alico Chair in Finance, at the Lutgert College of Business, Florida Gulf Coast University.

is transparent and open (and thereby precludes secret payments to public officials), there is no corruption involved. Corruption occurs when the agents' response to lobbying depends on secret private payments received. The literature labels it *grand or political corruption* when secret payments involve high-level government officials; *legislative corruption* when it involves elected representatives; and *petty, administrative, bureaucratic, or judicial corruption* when administrative or judicial officials are involved. Agency problems can only be resolved by monitoring, which is costly. Complete elimination of corruption, therefore, can be very costly. Indeed, there is scant historical or contemporary evidence of a country where the economy is completely free of corruption.

There is a view that although morally repugnant, corruption may play a useful role. For example, Leff (1964), Huntington (1968), and Leys (1965) argue that in countries burdened with cumbersome economic controls, selective tolerance of corruption is beneficial.[1] In such countries, corruption may "grease the wheels" of the economy so that it is "market expanding," according to Osterfeld (1992: 209). Rose-Ackerman (1997) questions that view and doubts that it is possible to contain corruption at a level where it is "efficient." Meanwhile, Kurer (1993) holds that corrupt officials have an incentive to create more opportunities for bribes. Finally, Murphy, Shleifer, and Vishny (1991) and Pecorino (1992) present the theoretical argument that corruption restricts the growth of an economy. Their view is supported by the empirical evidence in Mauro (1995), Mo (2001), Pelligrini and Gerlagh (2004), Méon and Sekkat (2005), and Swaleheen (2011).

The degree to which public officials in a country have the opportunity to resort to corruption depends on historical and homegrown factors—for example, the colonial past, level of economic development, electoral system, breadth and depth of regulations, decentralization, female participation, trade openness, inflation, and freedom of the press (Treisman 2000, Goel and Nelson 1998, Brunetti and Weder 2003, Swamy et al. 2001), as well as factors that originate in

[1]Aidt (2003) cites the example provided by Leff (1964) of corruption actually benefiting an economy: both Chile and Brazil introduced price controls for food products during a period of high inflation in the early 1960s. Chilean public officials enforced the price controls, which reduced food production. In Brazil, corruption weakened the price controls and food production increased.

other countries (Lambsdorff 1998, Getz 2006, Cuervo-Cazurra 2008). National anticorruption policies typically aim at regulatory and administrative reforms with varying degrees of success.

The focus of this article is on factors that originate in other countries and, therefore, are beyond the scope of national policies. Until 1977, the encouragements to public officials to act corruptly that came from foreign sources were largely ignored, and the bribing of foreign officials did not attract any sanctions. The Foreign Corrupt Practices Act of 1977 (FCPA) was the first instance of a home country attempting to put legal limits on business practices employed by its residents abroad. This unilateral step expanded into a multilateral effort criminalizing the bribing of foreign officials by multinational enterprises (MNEs) from countries in the Americas in 1997 and Europe in 1999. However, perceived corruption and the use of bribes around the world was not significantly reduced (Getz 2006). This situation led to suggestions that a uniform sanction against the use of bribes for all MNEs, from all countries doing business anywhere in the world, is needed to prevent some MNEs from gaining by using bribes at the expense of others who do not (Cuervo-Cazurra 2008). In 2006, a full global sanction against the bribing of foreign officials by MNEs was agreed upon. Table 1 presents a chronology of the expansion of the anti-bribery regime in the watershed years: 1997, 1999, 2003, 2005, and 2006. This article examines whether the global anti-bribery regime has led MNEs to cut back on the use of bribes. In the following sections, we first present the sources of data and the measures of bribes and corruption used in the study and then examine the available data for evidence of changes in the trajectory of bribing by MNEs and corruption in the MNE host countries corresponding to the aforementioned watershed years.

Measures of Bribes and Corruption

The MNEs that engage in international business are numerous and their business practices in foreign countries are not publicly known. A practical way of organizing information about global supply of bribes is, therefore, by country of origin of firms active in paying bribes. Transparency International's Bribe Payers Index (BPI) ranks countries in terms of how engaged MNEs from these countries are in offering bribes to foreign public officials. The

TABLE 1
Sequence of Multilateral Conventions against Corruption

Name of Convention	Signatories	Year Effective
The Inter-American Convention against Corruption (OAS Convention)	Organization of American States (OAS)	1997
Organisation for Economic Cooperation & Development Convention (OECD Convention)	OECD countries and Argentina, Brazil, Chile, Bulgaria, Estonia, Slovenia, and South Africa	1999
United Nations Convention against Transnational Organized Crime	All UN members	2003
UN Convention against Corruption	All UN members	2005
African Union (AU) Convention on Preventing and Combating Corruption	All AU members	2006

Source: Transparency International (2011).

index covers between 14 and 28 countries in a given year (Table 2). These countries account for more than half of global imports and exports or foreign direct investment flows for the corresponding year.

It is safe to assume that the countries that are not represented in Table 2 are each insignificant sources of supply of bribes to foreign officials. Thus, we treat them as the host countries to MNEs and destinations for bribes from sources listed in Table 2 (MNE home countries). Strictly speaking, BPIs from different years are not comparable, except in terms of the relative ranking of the MNE home countries. Transparency International (2011) does indicate, however, that some comparison over time is permissible—for example, in the case of the indices for 2008 and 2011, as they contain a largely overlapping set of countries.

TABLE 2
Bribe Payers Index (BPI) 1999–2011

Country	1999	2002	2006	2008	2011
Argentina					2.7
Australia	1.9	1.5	2.41	1.5	1.5
Austria	2.2	1.8	2.5		
Belgium	3.2	2.2	2.78	1.2	1.3
Brazil			4.35	2.6	2.3
Canada	1.9	1.9	2.54	1.2	1.5
China	6.9	6.5	5.06	3.5	3.5
France	4.8	4.2	3.5	1.9	2
Germany	3.8	3.7	2.66	1.4	1.4
Hong Kong		5.7	3.99	2.4	2.4
India			5.38	3.2	2.5
Indonesia					2.9
Israel			3.99		
Italy	6.3	5.9	4.06	2.6	2.4
Japan	4.9	4.7	2.9	1.4	1.4
Malaysia	6.1	5.7	4.41		2.4
Mexico			3.55	3.4	3
Netherlands	2.6	2.2	2.72	1.3	1.2
Portugal			3.53		
Russia		6.8	4.84	4.1	3.9
Saudi Arabia			4.25		2.6
Singapore	4.3	3.7	3.22	1.9	1.7
South Africa			4.39	2.5	2.4
South Korea	6.6	6.1	4.17	2.5	2.1
Spain	4.7	4.2	3.67	2.1	2
Sweden	1.7	1.6	2.38		
Switzerland	2.3	1.6	2.19	1.3	1.2
Taiwan	6.5	6.2	4.59	2.5	2.5
Turkey			4.77		2.5
U.A.E.			3.38		2.7
U.K.	2.8	3.1	2.61	1.4	1.7
U.S.	3.8	4.7	2.78	1.9	1.9
Share in global trade/ FDI (all countries)	60%	60%	82%	54%	78%

Notes: The BPI was rescaled so that a higher value indicates a higher propensity to bribe. A blank indicates data is not available.
Source: Transparency International (2011).

Bribes from MNEs to foreign officials show up as corruption in the MNE host countries. When it comes to measuring the level of corruption at the host country level, various problems emerge. In particular, corruption occurs in secret and is not directly observable. Reliability of the measurement of corruption is, therefore, a key issue in any empirical study. A good measure must be able to convey the frequency and the depth of corruption and be comparable over time.

Possible objective measures of corruption—like the number of convictions or the number of corruption cases reported in the press—do not accurately reflect the incidence of corruption alone; they also measure the performance of the judiciary and freedom of the press. Both vary from country to country. Also, measures based on such objective definitions are difficult to compare because legal definitions of bribery and corruption in different countries are yet to converge to the language of the multilateral statutes. Separating fraud and embezzlement from corruption is also difficult.

In the absence of a suitable objective measure of corruption that can be used for a cross section of countries, we use the Corruption in Government index from the International Country Risk Guide prepared by Political Risk Services (referred to as the ICRG index hereafter). It is a subjective measure of corruption based on surveys that has wide usage in the literature. The ICRG index ranges from 0 to 6 with a higher number signifying lower corruption and is created from a proprietary model that is based on survey results from respondents drawn from the same population year after year. The ICRG index is a good proxy for corruption under the reasonable assumptions that (1) perceived threat of political instability to foreign investors owing to corruption increases linearly with the incidence of corruption in the country, (2) corrupt public officials make no distinction at the margin between foreign and domestic firms when it comes to extracting bribes, and (3) the same range of information is used for all countries when assessing country risks (Swaleheen 2011). The ICRG index allows for cross-country comparison and is the longest running series of data on corruption that is available.

The Trend in Bribing of Foreign Officials

We first consider some readily available descriptive data that confirm previous findings (D'Souza 2012)—notably, that gradual globalization of the criminalization of bribery has changed the behavior of MNEs from OECD countries, and that MNEs from non-OECD

convention signatories may have picked up the business and the bribing left behind by their departed colleagues. The OAS and the OECD convention signatories have a longer record of criminalizing the bribing of foreign officials compared to other countries.[2] A comparison of the average of BPI for 1999 (BPI = 3.28) and the average for the post-1999 years (BPI = 2.41) for this set of countries indicates the use of bribes was lower.[3] Yet, in the post-2006 world where all countries have adopted a common anti-bribery regime, a comparison of the means of the BPI for the same set of countries in 2008 and 2011 does not show a significant reduction in the use of bribes.[4] Thus, even after full globalization, the efficacy of anti-bribery regime is not fully confirmed.

The data on bribe payments by MNEs that are presented in Table 2 are far from adequate for a rigorous and robust test of the hypothesis that a progressive global expansion of the anti-bribery regime will lower the supply of bribe payments in the countries that host MNEs. Faced with this problem, we examine the data that are available to test whether a direct corollary of the aforementioned hypothesis is supported—namely, whether the progressive globalization of the anti-bribery regime led to concurrent changes in the incidence of corruption in the host countries where the MNEs operate. In other words, as the global anti-bribery regime tightened over time for MNEs from countries listed in Table 2, was there a concurrent decline in corruption in the rest of the world that hosted the MNEs?

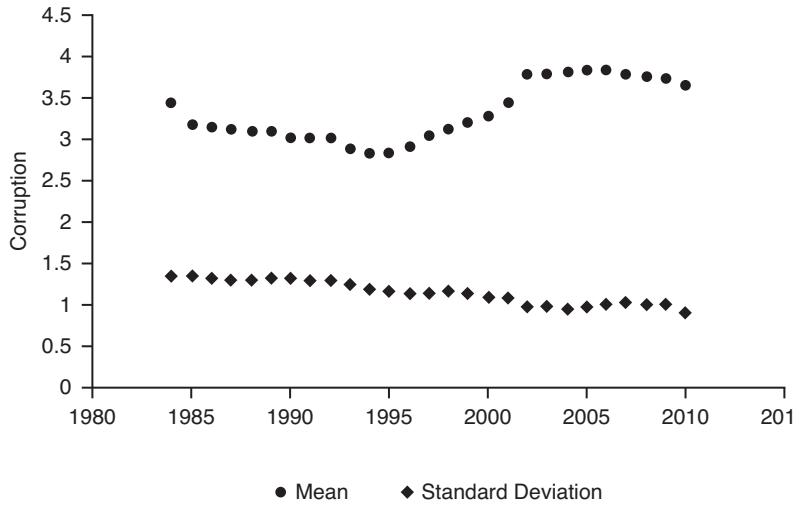
Data from the ICRG index are used to test this corollary. Given that the ICRG survey respondents are drawn from the same population every year, we hypothesize that unless there is a structural break, the distribution of the ICRG index in terms of the first and the second moments is repeated every year around an unknown trend that is driven by the process that determines corruption in each host country relative to every other host country. Note that the focus is not as much on the trend in corruption in all host countries as it is on

[2]These are Argentina, Australia, Austria, Belgium, Brazil, Canada, Germany, Japan, Mexico, Sweden, and the United States .

[3]The *t*-statistics for the null hypothesis that the two averages are equal is 1.71 (1.64 if unequal variances and independent distributions are assumed). The hypothesis that the average of BPI is lower in post-1999 years is accepted at a 6 percent level of significance.

[4]The *t*-statistic for the null hypothesis that the averages of the BPI for 2008 and 2011 are equal is −0.13.

FIGURE 1
Yearly Mean and Standard Deviation of Corruption in Host Countries



whether there is a break in this trend corresponding to the years when the anti-bribery regime went through an expansion as more countries signed on to multilateral conventions that criminalize bribing foreign officials.

Figure 1 plots the sequence of the mean and the standard deviation of the distribution of the ICRG index of corruption (rescaled so that a higher value indicates more corruption) for each year during 1984–2010 for the 146 countries in the ICRG panel that are not home to the MNEs that the data in Table 2 refer to. The graph shows that outside of the 32 MNE home countries listed in Table 2, the long-term trend in corruption in the rest of the world is rising. Corruption declined between 1984 and 1995, increased at a brisk pace between 1995 and 2002, and has remained somewhat stable thereafter. It is clear that the trend in average level of corruption in the host countries where MNEs do business has shifted over time, but, in a direction and to a level that is opposite to what one would expect from less bribing by MNEs owing to an expansion of the anti-corruption regime across the world.

Visual inspection of the trend presented in Figure 1 indicates the presence of an autoregressive process. Our examinations of the autocorrelation and the partial autocorrelation functions indicate that corruption is almost perfectly correlated with its own one-period lagged value.[5]

An augmented Dickey-Fuller test for the presence of unit root in the sequence of the mean level of corruption reveals that through the course of the sample period (1984–2010), the null hypothesis of a presence of breaks in the sequence is accepted. It is known that with structural breaks, the Dickey-Fuller test is biased toward non-rejection of the null hypothesis of a break (Enders 2010: 227). As we anticipate structural breaks in the sequence, we performed Perron's test, and it led to the same conclusion: there are structural breaks in the trend of the average level of corruption presented in Figure 1.

The question now is: Did the breaks occur in any of the years (1997, 1999, 2003, 2005, and 2006) when the coverage of global anti-bribery expanded? If the multilateral conventions against bribery worked, there should be evidence of structural breaks in one or more of these years of expansion. We perform Chow tests for structural breaks in each of these years by using the first-order autoregressive model in Equation 1.

$$(1)\quad M_t = \beta_0 + \beta_1 M_{t-1} + \gamma_0 D_t(\tau) + \gamma_1[D_t(\tau) \times M_{t-1}] + u_t,$$

where M is the mean of ICRG index in countries other than MNE home countries, τ is the hypothesized break year, and $D_t(\tau)$ is a binary variable that equals zero before the break date and 1 after. Thus, under the null hypothesis of no break at year τ, $\gamma_0 = \gamma_1 = 0$ against the hypothesis that at least one of the γ's is nonzero. The results are summarized in Table 3. The test results do not reveal any evidence of structural breaks corresponding to the years that marked the high-water levels in the expansion of the prohibition of illegal payments to foreign officials.

Conclusion

The United States was the first country to criminalize the bribing of foreign officials by its MNEs in 1977. The rest of the world joined

[5]Results are available from the authors.

TABLE 3
Chow Tests for Structural Break

Null Hypothesis	Break Year				
	1997	1999	2003	2005	2006
No break ($\gamma_0 = 0$)	Accept	Accept	Accept	Accept	Accept
(p value)	(0.5)	(0.8)	(0.9)	(0.9)	(0.9)
No break ($\gamma_1 = 0$)	Accept	Accept	Accept	Accept	Accept
(p value)	(0.5)	(0.8)	(0.9)	(0.9)	(0.9)
$\gamma_0 = \gamma_1 = 0$	Accept	Accept	Accept	Accept	Accept
(p value)	(0.83)	(0.95)	(0.99)	(0.99)	(0.99)

in in phases, and in 2006, a global anti-bribery sanction was agreed upon. This global prohibition was expected to lower the transnational supply of bribes and other illicit inducements to foreign officials and lead to a significant decline in the incidence of corruption around the globe. There is little evidence that these expectations have been realized. Data from Transparency International on bribe payments by MNEs before and after a fully globalized anti-bribery regime do not indicate a decline in the use of bribes to influence foreign officials. There is not one instance of a dip in the trajectory of public corruption in the countries where the MNEs do business that corresponds to any one of the five multilateral conventions that expanded and tightened the global anti-bribery regime. Indeed, the problem of corruption in the countries where MNEs operate has continued to grow.

One possible, albeit partial, explanation is that the U.S. Foreign Corruption Practices Act of 1977 and the OAS (effective 1997) and OECD (1999) conventions compelled MNEs from respective member countries to shift business to host countries where corruption is less acute (Cuervo-Cazurra 2008, D'Souza 2012), while MNEs from other countries (e.g., China, India, Turkey) filled the resulting vacuum. Further research may reveal that the quality of institutions in the countries that are home to the latter group of MNEs is not sufficiently strong to ensure the same degree of enforcement of anti-bribery laws as in OAS and OECD member countries. If that is the case, the United States and fellow OAS and

OECD member countries may need to consider limiting the geographical coverage of multilateral anti-bribery law to countries with strong institutions and level the playing field for all MNEs.

References

Aidt, T. S. (2003) "Economic Analysis of Corruption: A Survey." *The Economic Journal* 113: F632–52.

Bardhan, P. (1997) "Corruption and Development: A Review of Issues." *Journal of Economic Literature* 35: 1320–46.

Bhagwati, J. (1982) "Directly Unproductive Profit-Seeking (DUP)Activities." *Journal of Political Economy* 90: 988–1002.

Brunetti A., and Weder, B. (2003) "A Free Press Is Bad News for Corruption." *Journal of Public Economics* 87 (7): 1801–24.

Cuervo-Cazurra, A. (2008) "The Effectiveness of Laws against Bribery Abroad." *Journal of International Business Studies* 39 (4): 634–51.

D'Souza, A. (2012) "The OECD Anti-Bribery Convention: Changing the Currents of Trade." *Journal of Development Economics* 97: 73–87.

Enders, W. (2010) *Applied Time Series Econometrics*. 3rd ed. New York: John Wiley & Sons.

Getz, K. A. (2006) "The Effectiveness of Global Prohibition Regimes: Corruption and the Anti-Bribery Convention." *Business and Society* 45 (3): 254–81.

Goel, R. K., and Nelson, M. A. (1998) "Corruption and Government Size: A Disaggregated Analysis." *Public Choice*, 97: 107–20.

Huntington, S. P. (1968) *Political Order in Changing Societies*. New Haven: Yale University Press.

Jain, A. K. (2001) "Corruption: A Review." *Journal of Economic Surveys* 15: 71–121.

Krueger, A. O. (1974) "The Political Economy of the Rent-Seeking Society." *American Economic Review* 64: 291–303.

Kurer, O. (1993) "Clientelism, Corruption and the Allocation of Resources." *Public Choice* 77: 259–73.

Lambsdorff, J. G. (1998) "An Empirical Investigation of Bribery in International Trade." *European Journal of Development Research* 10 (1): 40–59.

Leff, N. H. (1964) "Economic Development through Bureaucratic Corruption." *American Behavioral Scientist* 8: 8–14.

Leys, C. (1965) "What Is the Problem about Corruption? *Journal of Modern African Studies* 3: 215–30.

Mauro, P. (1995) "Corruption and Growth." *Quarterly Journal of Economics* 110: 681–712.

Méon, P., and Sekkat, K. (2005) "Does Corruption Grease or Sand the Wheels of Growth?" *Public Choice* 122: 69–97.

Mo, P. H. (2001) "Corruption and Economic Growth." *Journal of Comparative Economics* 29: 66–79.

Murphy, K. M.; Shleifer, A.; and Vishny, R.W. (1991) "Allocation of Talent: Implications for Growth." *Quarterly Journal of Economics* 106 (2): 503–30.

Osterfeld, D. (1992) *Prosperity versus Planning: How Government Stifles Economic Growth*. New York: Oxford University Press.

Pecorino, P. (1992) "Rent Seeking and Growth: The Case of Growth Through Human Capital Accumulation." *Canadian Journal of Economics* 25 (4): 944–56.

Pelligrini, L., and Gerlagh, R. (2004) "Corruption's Effect on Growth and Its Transmission Channels." *Kyklos* 57 (3): 429–56.

Rose-Ackerman, S. (1997) "The Political Economy of Corruption." In K. A. Elliot (ed.), *Corruption and the Global Economy*, 31–66. Washington: Institute for International Economics.

Swaleheen, M. (2011) "Economic Growth with Endogenous Corruption: An Empirical Study." *Public Choice* 146: 23–41.

Swamy, A.; Knack, S.; Lee, Y.; and Azfar, O. (2001) "Gender and Corruption." *Journal of Development Economics* 64 (1): 25–55.

Transparency International (2011) "Bribe Payers Index, 2011." Available at www.transparency.org/bpi2011.

Treisman, D. (2000) "The Causes of Corruption: A Cross-National Study." *Journal of Public Economics* 76 (3): 399–457.

Tullock, G. (1993) "Rent Seeking." *The Shaftesbury Papers*, Vol. 2. Northhampton, Mass.: Edward Elgar.

Religious Liberty and Economic Prosperity: Four Lessons from the Past

Anthony Gill and John M. Owen IV

Many political scientists and economists have argued that expanded civil liberties and general political freedom are conducive to economic growth (e.g., Smith [1776] 1976; North, Wallis, and Weingast 2009). One subset of these civil liberties—religious freedom—is frequently neglected in such discussions, with scholars tending to focus on the rights of private property ownership, contracting, assembly, and access to political decisionmaking. Religious liberty is often seen as an isolated freedom that directly impacts parishioners and clerics but has little spillover effects on the general economy or society. But can the right to worship freely also have positive consequences for economic growth? A casual glance suggests nations that have developed strong legal guarantees of religious freedom (and a concomitant culture of religious toleration) are also ones that have had long-term sustained economic growth (Grim and Finke 2011).

Within the modern world, the Netherlands, Britain, and the United States were the first adopters of toleration and liberties for religious minorities, and these countries also became the loci of rapid

---

Cato Journal, Vol. 37, No. 1 (Winter 2017).

Anthony Gill is Professor of Political Science at the University of Washington and Distinguished Senior Fellow at Baylor University's Institute for Studies of Religion. John M. Owen IV is Ambassador Henry J. and Mrs. Marion R. Taylor Professor of Politics at the University of Virginia and Faculty Fellow at the Institute for Advanced Studies in Culture.

growth in entrepreneurial activity in the 17th century and beyond, not to mention some of the first societies to set the modern standards for constitutional democracies.

Might there be a causal connection between religious freedom, on the one hand, and societal prosperity, on the other? If so, what are the precise mechanisms linking the two? We take a historical approach to this puzzle and a set of related questions. Namely, if religious freedom has remained historically elusive, how did it ever originate? Can the experience of the past inform us about the process to achieve expanded civil liberties in the area of religious belief and practice today? Answers to these questions will enable us to fill out our understanding of civil and economic liberties and the complex relationship between general freedom and economic prosperity.

Our examination of these issues takes us back to the 17th and 18th century, with the first emergence of religious freedom in the modern era in the Netherlands, Britain, and British American colonies, eventually culminating in the First Amendment of the United States Constitution—not a be-all-and-end-all of religious freedom but certainly an important historical marker. We assert that history is a salutary teacher and that current struggles for enhanced democratic freedoms can learn from the processes of the past. Furthermore, we argue that the emergence of religious toleration coincided with, and is causally related to, the growth in prosperity that was witnessed in these nations during the same period. While not asserting that a respect for religious freedom is the only reason for increased economic prosperity, we do consider it an important catalyst that should be taken seriously in contemporary policy debates. Understanding that the promotion of a core set of civil liberties—namely, the rights of religious conscience and practice—are linked to material prosperity helps us understand the more holistic nature of social flourishing, both economic and political. We make our case by laying out the generally understood causes for economic growth and illustrate it with four historical lessons.

Conditions for Economic Growth: Institutional and Ideational

Economic historians have long noted that the conditions giving rise to the Industrial Revolution of the 1800s began in Northwest Europe several centuries earlier (North and Thomas 1976, Weber

[1930] 2001). This region began witnessing a steady increase in economic productivity and living standards in the late medieval period that set it apart from other civilizations around the globe (Kuran 2011). Given the seemingly endogenous relationship between economic growth and political liberalization (North, Wallis, and Weingast 2009), it behooves us to examine some of the basic factors promoting the relationship between prosperity and democratization. Writing on the cusp of the Industrial Revolution, Adam Smith was well-placed to observe what was giving rise to an increasingly obvious change in European economy and society. Though his *Inquiry into the Nature and Causes of the Wealth of Nations* is long, his basic idea is very simple: wealth arises from the reciprocal relationship between the division of labor (i.e., specialization) and expanded trade (Smith [1776] 1976: 25–27). Specialization allows one to become more efficient at producing goods and services, while trade allows people to specialize in one thing while still having their needs and desires met by the production of others. Anything that inhibits the ability to trade decreases the benefits of specialization and, thus, retards long-term economic growth. Trade being critical here, the next logical question is: What lowers barriers to trade, both domestic and international?[1]

The factors contributing to increased trade and concomitant economic production have been numerous, with most explanations centering on secular concerns. Popular among a new breed of political economists have been political institutions that establish a predictable set of property rights that allow producers to capture the fruits of their production (North 1990, Olson 1993, Weingast 1995).[2] To the extent that the rapacious hands of political rulers are limited by checks upon concentrated power and constitutional guarantees of civil liberties, individual entrepreneurs will have an incentive to invest in long-term projects that increase efficiency and promote

[1]It is common to assume that we are talking about cross-national trade when mentioning "barriers to trade." However, trade occurs among at least two individuals irrespective of nationality and things such as local zoning laws can play an important role in limiting or encouraging mutual exchange.

[2]The institutional literature on economic growth is too expansive to do justice to here. Our main point is that economists have mostly focused on secular rules of governance to explain economic growth, relegating concerns over religious issues to the sidelines.

expanded specialization and trade. The secular ruler benefits from such an arrangement by being able to tax an expanding economic pie. Finding a balance that allows rulers to maximize their tax revenue without damaging the incentive of private entrepreneurs to expand production is at the crux of economic growth for these theories. To the extent that government leaders understand that any barriers to trade negatively affect their ability to increase tax revenue over the long term, private actors will be able to leverage greater freedom of operation for themselves. Such freedoms may be specifically related to physical manufacturing and commerce but may also extend to other important matters such as the ability to worship freely, as we will argue below.

More recently, economists have revived the notion that culture and ideas serve to reduce impediments to free exchange and efficient production (McCloskey 2006, Mokyr 2009). The insights here are that the intellectual conditions of the Enlightenment provided for a greater free flow of ideas, which in turn prompted innovation and reduced the stigma attached to various entrepreneurial behaviors. It was no coincidence that rapid industrial growth followed on the heels of philosophical ideals of greater toleration, limited government, and the virtues of thrift. Stark (2005) and Aquilina and Papandrea (2015) take this argument deeper historically by noting that it was Christianity, and not just a Calvinistic ethic per se (à la Weber), that promoted egalitarian ideals and civil rights that eventually blossomed into historically unparalleled prosperity.

It is into this debate that we insert our contention that the desire of individuals to practice their faith freely, and the willingness of governments to allow such liberty, enhanced long-term economic development. Religious liberty is not the only factor responsible for economic growth, but it is an important ingredient that makes the economic pie larger—and all the more sweet. The idea that individuals holding different beliefs should be tolerated in society, and the incentive of these individuals to promote institutions that allow them to organize and worship freely, contribute greatly to an environment that promotes a wide variety of civil liberties that concurrently facilitate a number of secular relationships, which in turn promote greater interaction (trade) among people. In short, religious liberty is a catalyst for the freedoms that constitute democratic civil society and promote prosperity over the long haul.

Four Lessons from an Enlightened Time and Place

Our empirical journey to look for places where religious liberty may have had an impact on economic development and democratization takes us back to Northwest Europe in the 17th and 18th centuries, specifically to the Netherlands and England (and by extension the British American colonies). It was here that we see the beginnings of modern democratic governance, the rise of a liberal commercial culture, and the first steps toward industrialization. It was also a period of increasing religious pluralism and scholarly innovation, with the likes of Locke, Hobbes, and Smith laying the intellectual bedrock for liberal societies. While it may be that such cultural changes were coincidental to economic growth, exploring the potential causal linkages is worthwhile. And if this relationship held true four centuries ago, there is reason to think it holds true today. To that end, we offer four lessons from the historical experiences of the Netherlands, Britain, and the British American colonies.

Lesson 1: Religious Freedom Promotes Diversity, Security, and Prosperity

Freedom of religion was pioneered slowly by religious people who accepted the inevitable permanence of religious diversity and strongly desired a cohesive, secure, and prosperous society.

Religious pluralism is an essential ingredient to religious liberty. If all individuals within a society shared the same beliefs and set of worship practices, there would be little, if any, need to ensure the rights of religious dissenters; they simply would not exist! Looking at cultures writ large, we may be tempted to think that religious homogeneity is a defining feature of a culture, but upon closer inspection, there is a great deal of diversity that can be discerned. Stark (2003) argues that monotheism is inherently schismatic as there will invariably be differing interpretations of the "one true God" and various forms of worship styles that appeal to different groups in society (Owen 2015: 46–66). Even prior to the Protestant Reformation, Christianity was bursting at the seams with various theological flavors ranging from Franciscan monasteries to wandering Waldensians to the pre-Reformation reformers John Wycliffe and Jan Hus. While some reformist monastic orders were given freedom in certain areas, there were simultaneous attempts to suppress other movements—such as those instigated by Jan Hus, who paid for his

dissent with his life. This came to a head in 1517 with Martin Luther leading a definitive break from the dominant church of the day, followed shortly thereafter by John Calvin and others. Protestantism, as the alternative to Catholicism became known, created a fissiparous system of ongoing schisms and religious minorities. Not surprisingly, religious dissent brought about conflict, as witnessed by the Thirty Years War (1618–48) and the English Civil War (1642–51). But too much war prompted many intellectuals and diplomats to understand the futility in such conflict and promote a "balance of power" tolerance that would avoid the economic devastation of perpetual war (Philpott 2001).

Consider the Dutch Republic. While not inherently blessed with natural mineral wealth, the residents of the Low Countries became known for their enterprise throughout the late medieval period by exploiting the Rhine and Ems rivers to begin developing a comparative advantage in commerce. But they were hampered at sea by their swampy coastline until Dutch farmers began to move into the marshy areas and develop dikes to keep the water out. This technological innovation yielded two important results. First, the large areas of new land lacked feudal titles, and so Dutch peasants became the freest in Europe, able to capture a greater portion of the profits from their own labor. Second, the new land enabled the Dutch to expand their comparative advantage in commerce by dominating maritime trade in the North and Baltic seas (Maddison 2001: 77–80). The Dutch, from the lowliest peasant to the highest noble, learned that their prosperity hinged upon keeping the barriers to trade low.

Although technology allowed them to conquer geographic barriers to trade, another obstacle remained—religious differences. When the Protestant Reformation swept Europe in the 1510s, the Netherlands were ruled by the Habsburgs, who also controlled Spain and most of Central Europe, regarding themselves as guardians of Catholic fidelity. Calvinism spread into the Netherlands rapidly in the 1550s, creating a clear religious split that threatened to divide the nation beyond its already variegated ethnicities. Philip II of Spain not-so-kindly responded to Protestant advances with an Inquisition. In 1566, the Dutch incited an 80-year revolt against Habsburg rule with both Catholics and Protestants finding cause in independence. However, in an effort to keep Catholics from defecting to the Habsburgs, Calvinists had to send a credible signal that their right to worship would be protected. Likewise, Protestants wanted assurances that

Catholic hegemony would not be reasserted after victory. The result was the Union of Utrecht (1579), which functioned as an informal constitution and included a provision stating that each Dutch province could regulate religious matters as it saw fit, that there would be no official establishment of religion, and that persecution based upon religious belief would be prohibited (Haefeli 2012: 20–21). Two features of this pact stand out: first, a presumption of provincial rather than national control regarding religion—analogous to federalism, and essential in a land with seven distinct provinces; and second, a proviso prohibiting religious persecution. These provisions, which help to keep a diverse nation unified against a common enemy, would be echoed two centuries later in the U.S. Constitution.

The emergence of religious toleration took many decades, however. A number of Dutch thinkers followed the great humanist Erasmus of Rotterdam (1466–1536) in arguing for liberty concerning nonessential doctrines, and noted that Habsburg religious repression was one reason why the Dutch fought for independence (Witte 2007: 171–73). Others disagreed strongly (Mout 1997: 39). In most provinces, Calvinists made up at most one-quarter of the population, but they were the best organized and most disciplined group and spent the next four decades gaining political predominance, with the approval of state authorities (Israel 1997: 3–4). In the early decades, where Calvinists were in power, they prohibited Catholic worship. At the same time, in some cities in the state of Holland (the largest Dutch state), civil authorities began to defy the Reformed clergy by allowing Lutherans to have small churches. In 1612, the Jews of Amsterdam began building a large synagogue, but that crossed a line: pressured by the clergy, the city halted construction. In the meantime, Catholics conducted worship in large houses but could not construct church buildings (Israel 1997: 8–9). The Union of Utrecht may have codified religious toleration, but the virtue of real toleration still needed to be intellectually absorbed in the culture.

A traumatic split within the Reformed church itself helped to convince more and more Dutch citizens that toleration was key to peace and prosperity. The Arminians or Remonstrants rejected certain Calvinist doctrines and attracted many followers. The Synod of Dordrecht (1618–19) defeated the Remonstrants, entrenching the orthodox Reformed as the established, state-subsidized church. But in contrast to the Anglican settlement in England, the Dutch state did not control the church and Dutch citizens were not required to

join it (Mout 1997: 41–43). An entire category of believer emerged: the *liefhebber* (sympathizer or, literally, "lover"), who attended Reformed services but eschewed membership (Kooi 2002: 32–33). And in the cities of Holland, magistrates allowed more and more Lutherans, Arminians, and even Jews and Catholics, to hold services and construct houses of worship. Amsterdam, Rotterdam, and other major cities were developing a strong burgher class of shopkeepers, professionals, teachers, artisans, merchants, and bankers. Wealthy burgher families took to paying city officials to leave them alone to practice their religion (Haefeli 2012: 55). Included in the bargain was that minority communities would look after their own poor and sick, and not draw on the public coffers (Kooi 2002: 40–43). A symbiotic relationship took hold between city officials and religious minorities, with the benefits being realized in bustling trade. However, there was variation in the tolerance observed, and one particular case proves this lesson in the obverse. "Lacking Amsterdam's dependence on international trade, Leiden was much less hospitable to both Jews and Lutherans and generally a strong supporter of the power of the Dutch Reformed Church and Calvinism" (Haefeli 2012: 77). Eventually, though, the entire political economy of Holland's cities came to reap the benefits of toleration. As Joke Spaans writes, "Tolerance was after all conducive to the trade interests of the merchant elite that ruled the cities and eventually the state itself" (Spaans 2002: 78).

As many scholars have argued, the Netherlands served as the birthplace for modern notions of religious liberty. The necessity for several regional ethnic groups to band together for commerce initially set the table. The burgeoning of religious pluralism fostered by the Protestant Reformation threatened to tear the fragile unity of the Low Countries apart, which would have likely led to defeat of their rebellion against the Habsburgs. Moreover, domestic conflict amongst different confessions threatened a civil war that would destroy the advantage in peaceful trade the Dutch had built over time. Slowly, a social consensus evolved around a "live and let live" attitude that allowed individuals of different faiths to interact with one another and capture the gains from trade, making the Netherlands one of the great commercial empires of the time. This lesson was not lost upon other observers who realized that the Netherlands was a nice place to set down roots and build a business (especially if one was fleeing from religious persecution elsewhere),

and even if one did not plan to reside amongst the Dutch, it was nonetheless a safe haven to trade with others without fear of being persecuted for dissenting spiritual beliefs. England, having experienced its own religious turmoil during a bloody civil war, picked up the ideas of the Dutch (not to mention a king) during their Glorious Revolution, a matter we lay out below. This leads us neatly to our next two historical lessons: religious liberty attracts entrepreneurial individuals and fosters peaceful trade amongst strangers who may not necessarily think alike.

Lesson 2: Religious Freedom Attracts Entrepreneurs Who Foster Economic Growth

Freedom of religion attracts creative, risk-taking, pioneering, entrepreneurial individuals and groups who increase the rate of economic growth.

Economies need people: people to produce, people to consume, and people to innovate (a key engine of economic growth). As Adam Smith noted,[3] and the experience with import-substitution policies of the mid-20th century proved, autarkic societies do not grow over the long term. Countries and territories that encourage entrepreneurial individuals to settle, manufacture, truck, and barter will thrive over time.[4] This lesson also runs in reverse: societies that enact policies odious to the most entrepreneurial will see these individuals flee and witness a corresponding lag in productivity.

Nowhere was this lesson clearer than in France's religious history. Beset with bloody religious conflict following the Protestant Reformation, France's King Henry IV decreed the Edict of Nantes in 1598, giving French Protestants (Huguenots) a tolerated freedom to practice their "dissenting" faith. As Armstrong (2004: 11–12) notes, this decree recognized that "outright violence would only devastate the king's forces and *the French economy*" and that "Catholics were willing to endure the presence of Protestants for the sake of peace" (emphasis added). An uneasy religious truce endured and

[3]Smith's argument to this effect is threaded throughout *The Wealth of Nations*, including in an often-overlooked history of the European economy in Book III.

[4]The reference to "truck and barter" is a hat tip to Adam Smith, who noted the innate tendency of all individuals to "wheel and deal" (Smith [1776] 1976: 25).

France grew until roughly a century later when Catholics began reasserting their cultural authority. King Louis XIV revoked the decree in 1685, resulting in an increased wave of persecution against the Huguenots. The French Calvinists fled in the hundreds of thousands to neighboring countries including England and the Netherlands, where they helped boost the economy (Scoville 1952). Not surprisingly, it was the most industrious and wealthy who were able to exit France and put down roots elsewhere, the type of people that governments love to attract. The Netherlands also offered refuge to Jews who were fleeing persecution in Portugal at the time and "brought valuable trade connections and knowledge to Amsterdam just as the Dutch began expanding their trade networks overseas" (Haefeli 2012: 111).

One area of Dutch expansion was North America, where they developed an outpost on a small island they would call New Amsterdam. With the home country bustling, it was difficult to attract settlers to migrate to the rugged Americas, a situation that eventually led the colony to be taken over by the British, who named it New York. However, before this happened, the directors of the Dutch West India Company issued a direct order to respect different religions so as not to chase people away and destroy any hopes of economic fortune. In response to a reluctance to trade with a Quaker merchant, they wrote: "We heartily desire that these and other sectarians remained away from there, yet as they do not, we doubt very much, whether we can proceed against them rigorously without diminishing the population and stopping immigration, which must be favored at so tender stage of the country's existence" (cited in Smith 1973: 230).

The English were quick to understand the importance of religious freedom for the construction of prosperous colonies. None other than the great champion of early American religious freedom, William Penn, advertised for colonists to come to Pennsylvania based upon an explicit appeal to religious freedom (Sweet 1935: 50). And in the Carolinas, the famed American religious historian Perry Miller argued that "the [religious] uniformity for which the noble proprietors hoped was impossible, unless they were prepared to expel nine-tenths of their settlers. So religious principle gave way to economic interest; practical toleration became the rule" (Miller 1935: 60). Ironically, the lesson that would be learned by the French after revoking the Edict of Nantes and chasing away valuable Protestant

merchants would prove to be helpful in building the Virginia colony. While persecution of religious dissenters in the early 17th century pushed Puritans to flee and settle New England, the English Civil War (1642–51) turned the tables and Puritan persecution of High Church Anglicans pushed between 40,000 and 50,000 English to Virginia between 1645 and 1670 (Fischer 1989: 207–29). These settlers built the colony into North America's most populous and prosperous by the time of American Independence. As all the British American colonies grew, it was difficult to suppress the denominational pluralism resulting from de facto policies of toleration. Migrants, particularly the risk-taking entrepreneurial types, are a necessary component of economic growth. As diverse people learned to live in harmony with one another, toleration eventually became constitutionally enshrined religious liberty (Gill 2008).

Lesson 3: Religious Freedom Fosters a Commercial Society

Religious toleration and liberty encourage trade and commerce by decreasing a cultural barrier to trade, thereby enriching the coffers of the king.

Gaining entrepreneurial immigrants is an important aspect of building a prosperous economy, but fostering trade with those who wish to remain in foreign lands and occasionally visit yours is equally important, as Adam Smith reminded us. The efficiency of specialization and comparative advantage hinges on the ability to develop a wide trade network. Foreclosing commercial interactions with individuals with whom you have spiritual disagreements simply shutters a country off from the wealth of others—people who would buy your nation's products and could offer desirable things in return. If foreign traders are concerned with being imprisoned or disproportionately taxed for adhering to a different faith, they are much less likely to come to your shores. As noted above, the Netherlands thrived because of its people's ability to trade with people of other nations, irrespective of their religious beliefs. Indeed, free trade was the "product" in which the Dutch realized their greatest comparative advantage.

The lesson that religious toleration and liberty fostered trade was also clearly apparent in the British American colonies. Historians of colonial America have realized this: "Trade tended to distract colonies from their absorbing preoccupation with an exclusiveness in the matter of religion and encouraged their thinking relatively less of

the Church and more of the State and of commerce. The colonists began, in turn, to see the enormous advantage commerce would derive from liberty" (Stokes and Pfeffer 1964: 29). A case in point was colonial Boston under the control of a fairly strict Congregationalist-based government. Although the Massachusetts Bay Colony Puritans were rather restrictive of dissenting sects (e.g., Quakers) and, early on, forced them to pay mandatory tithes and even incarcerated some of them, the policy toward merchants in Boston was much more liberal. As early as the 1650s, visiting merchants and their settled representatives in that port city were exempted from mandatory tithing to the Congregational church (McLoughlin 1971: 118). Such a mandatory tax would have represented a "spiritual tariff," and the governors of that colonial outpost understood that such taxes would only divert trade to other more favorable locations. These exemptions were extended within the colonies as well, with a potential trade war between Rhode Island and the other New England colonies being averted when it was agreed that Quakers should not be detained in their travels between Pennsylvania and Providence solely for religious reasons (Curry 1986: 22–23).

Pennsylvania once again played a major role in fostering American religious liberty. Its founder, William Penn, was an intellectual who understood both the inherent benefit of religious tolerance and its instrumental value in achieving wealth for all. While successfully advocating for greater freedoms for all confessions in the colony, he explicitly made appeals to trade and economic growth:

> But as [religious persecution] has many Arguments for it, that are drawn from the Advantages that have and would come to the Publick by it, so there are divers Mischiefs that must unavoidably follow the Persecution of Dissenters, that may reasonably disswade from such Severity. For they must either be ruined, fly, or conform; and perhaps the last is not the Safest. If they are Ruin'd in their Estates and their Persons Imprisoned, modestly compute, a Fourth of the Trade and Manufactury of the Kingdom sinks; and those that have helped to maintain the Poor, must come upon the Poor's Book for Maintenance [Penn 2002: 317].

It was quite obvious to Penn that killing and incarcerating people who want to trade goods with you is a bad way to foster commerce. Yet Penn's genius rested not only upon this argument, but also in

connecting it with the wealth of the English crown, which he had hoped would enforce laws of toleration:

> Consider Peace, Plenty, and Safety, the three great Inducements to any Country to Honour the Prince, and Love the Government, as well as the best Allurements to Foreigners to trade with it and transport themselves to it, are utterly lost by such [persecution]. . . . Men of Virtue, good Contrivance, Great Industry; whose Labours, not only keep the Parishes from the Trouble and Charge of maintaining them and theirs, but help to maintain the Poor, are great contributors to the King's Revenue by their Traffick [Penn 2002: 58].

If you want the king to enforce freedoms you desire, convince him it is good for his treasury. Samuel Davies, an itinerant Presbyterian preacher who would one day be the president of Princeton University, would make a similar appeal for religious freedom by petitioning none other than the Lords of Trade in New Jersey arguing "a free Exercise of Religion is so valuable a branch of true liberty, and so essential to the enriching and improving of a Trading Nation, it should ever be held sacred in His Majesty's Colonies" (quoted in Isaac 1973: 27). While religious toleration does foster economic growth by lowering the barriers to trade among people of different denominations and providing a safe haven for entrepreneurs of minority faiths, we do recognize the endogeneity in the relationship; success with trading amongst individuals of different religions will further encourage greater toleration as it becomes apparent to others that incivility toward minorities is not worth the loss of wealth due to sectarian conflict.

The insight that intolerance toward people of other faiths would dampen their desire to interact with you seems rather obvious. Nonetheless, theological preferences run deep and are held innately valuable by many persons, a fact that social scientists should not underestimate. Restrictions on religious belief and practice have not been uncommon throughout history and have served as barriers to immigration and commerce. The need of William Penn, Samuel Davies, the burghers of the Netherlands, and others[5] to argue that

[5]Gill (2008) provides a number of other instances where religious liberty was linked positively to immigration and trade in Europe and the colonial Americas, as well as Latin America and the Baltics.

religious freedom can be conducive to social peace and economic prosperity demonstrates that the connection is not obvious, and needs to be argued throughout time. Fortunately, manifest success can breed more success, which brings us to our fourth and final lesson.

Lesson 4: Religious Freedom Tends to Spread

Once entrenched in one country, religious freedom tends to spread to neighboring countries.

Social scientists have noted that ideas and policies often diffuse across countries (Simmons and Elkins 2005). These scholars have come up with various mechanisms by which institutions and practices spread (Cederman and Gleditsch 2004, Gleditsch and Ward 2006). Diffusion can be propelled by the states with the new ideas—as when a state promotes its institutions in other countries (Owen 2010, Boix 2011). Recipient states can be attracted to the ideas and policies of their neighbors, particularly if those neighbors are successful (secure, powerful, or wealthy); scholars term this a "demonstration effect." Recipients also can feel pressure ("externalities") to imitate the policies of successful neighboring states; for example, a large economy that practices free trade gives incentives to its neighbors to emphasize exports over imports.

Religious freedom spread from the early modern Dutch Republic to other polities in Europe and across the Atlantic by all three mechanisms: economic pressure, demonstration effects, and even imposition. The 17th century was known as the Netherland's Golden Century, in which the Dutch Republic defied widely held prophecies of its doom and not only remained independent from Spain but also became Europe's largest economy and greatest trader, with the biggest merchant fleet and imperial holdings in the Americas, Africa, and Asia. The Netherlands' neighbors were threatened and perplexed by its successes. Some responded with war: England and the Dutch Republic fought three wars over trade routes and colonies (1652–54, 1665–67, and 1672–74), and France tried out-and-out conquest (1672–78). Notwithstanding their geographical exposure and relatively small population, the Dutch survived and indeed continued to prosper, and foreigners inquired into why they enjoyed such national success.

Economic pressure and demonstration effects from the spectacular Dutch economy worked together to lead England to imitate the Netherlands' religious freedom. Religious dissenters (non-Anglicans)

were especially prone to attribute Dutch success to its toleration of religious diversity. Thomas Helwys (1575–1616), founder of the Baptist denomination, relocated to the Netherlands in 1608. Persuaded by what he saw that state persecution of religion was never justified, Helwys wrote: "Behold the Nations where freedome of Religion is permitte, and you may see there are not more florishinge and prosperous Nations under the heavens than they are." At roughly the same time, English Separatists, known more colloquially as the Pilgrims, settled in the Netherlands, eventually making their way to Plymouth, Massachusetts. In Plymouth, the Separatists were models of toleration (in contrast to their Puritan brethren, who exiled and occasionally executed religious dissenters), perhaps because they had seen toleration modeled in Holland (Cobb 1970: 133–48). Two generations later Sir William Temple (1628–99) visited the Netherlands and published his *Observations upon the United Provinces of the Netherlands* (1673). Openly admiring of Dutch toleration, Temple wrote that in the Netherlands even Catholics could practice openly, and "it is hardly to be imagined how all the violence and sharpness which accompanies the differences in religion in other countries seems to be appeased or softened here [in the Netherlands], by the general freedom which all men enjoy, either by allowance or connivance; nor how faction and ambition are thereby disabled to color their interested and seditious designs with the pretenses of religion, which has cost the Christian world so much blood for these last hundred and fifty years" (Owen 2015: 136).

John Locke, the English philosopher, was another admirer of the Dutch. In 1669, he coauthored the Fundamental Constitutions of Carolina, the basic law for a new colony (now North and South Carolina). Articles 96 and 97 are especially interesting. The former, not written by Locke, mandated that "the only true and orthodox religion," the Church of England, would be the sole religious institution publicly supported. But Locke's Article 97 followed up by stating that, since immigrants "will unavoidably be of different opinions concerning matters of religion, the liberty whereof they will expect to have allowed them . . . it will not be reasonable for us, on this account, to keep them out, that civil peace may be maintained amidst diversity of opinions." Locke went on to assert that "Jews, heathens, and other dissenters" could be won to pure Christianity by seeing the meekness of the faithful. The upshot was that "any seven or more persons agreeing in any religion, shall constitute a church or

profession, to which they shall give some name, to distinguish it from others" (*Fundamental Constitutions of Carolina* 1669). Locke later spent five years in the Netherlands (1683–88), a country he admired greatly. The year after his return to England he published his famous *Letter Concerning Toleration* (Locke [1689] 2010).

Locke came back to England with William of Orange, who was invited by Parliament to become King William III of England. The Glorious Revolution in England was, among other things, a foreign (Dutch) imposition of religious toleration upon England, done for geopolitical reasons. When the Catholic James II was ousted from the English throne, Louis XIV of France, the superpower of the time, lost a close ally. He declared war on the Netherlands and accused William of planning to eradicate Catholicism from England. To placate Louis, William struck a deal with a reluctant English Parliament: the Toleration Act of 1689 allowed Presbyterians, Quakers, Baptists, and Independents to worship in public, build chapels, and contradict Anglican preachers. As king, William went further, using his royal prerogatives to "direct judges and curb popular and ecclesiastical interference and opposition" to Catholics and Jews (Israel 1991: 140–54). In sum, England became more tolerant of minority religions because of the example, pressure, and direct intervention of the already tolerant Netherlands.

The connection between the Netherlands, John Locke, the English Toleration Act, and the subsequent religious freedoms that flourished in the British American colonies cannot be underestimated. Good ideas do not manifest themselves inherently, and they require proponents to carry them forward throughout time and space. Adam Smith, whom we earlier credited with laying the foundational path toward economic prosperity—specialization and trade—knew full well the importance of religious toleration, devoting a whole section of Book V of *The Wealth of Nations* to the free exercise of religion. Smith ([1776] 1976: 793) even noted the socioeconomic importance of religious liberty in Pennsylvania. Without the demonstration effects and free flow of ideas, many of the social benefits provided by religious toleration would remain provincial.

Conclusion

Religious freedom has once again taken center stage in policy debates around the world. *Dignitatis Humanae*, celebrating its 50th anniversary, argued for this essential civil liberty based upon the

inherent dignity it brings to the human person. This appeal alone may be cause enough to justify its implementation and protection in democratic constitutions and specific policy measures. However, history informs us that such appeals often go unheeded, as the political landscape is often strewn with competing social and economic interests. Religious intolerance still exists and threatens to divide communities that would otherwise be strong partners in producing wealth and engaging in peaceful trade. To that end, the case for religious toleration and institutionalized liberties may lie in noting the connection they have for social flourishing as well. Nearly all individuals seek a safe and prosperous life, free from conflict. Fortunately, history also provides evidence that religious tolerance and freedom enhances social well-being in many other realms, including long-term economic growth and democratic governance. Just as tolerance promoted security, immigration, and trade in the past, it can continue to do so today. Religious minorities continue to flock to nations that provide them a safe haven, and in doing so, they bring their particular talents that enhance their new societies. Such tolerance also fosters increased trade across cultural boundaries, reminding all of us of our common humanity. As noted in our fourth lesson above, the diffusion of good ideas is critical for the enactment of good policies. History is a teacher and we would do well to heed its lessons.

References

Aquilina, M., and Papandrea, J. (2015) *Seven Revolutions: How Christianity Changed the World and Can Change It Again*. New York: Image.

Armstrong, M. C. (2004) *The Politics of Piety: Franciscan Preachers during the Wars of Religion, 1560–1600*. Rochester, N.Y.: University of Rochester Press.

Boix, C. (2011) "Democracy, Development, and the International System." *American Political Science Review* 105 (4): 809–28.

Cederman, L., and Gleditsch, K. S. (2004) "Conquest and Regime Change: An Evolutionary Model of the Spread of War and Peace." *International Studies Quarterly* 48 (3): 603–29.

Cobb, S. H. ([1902] 1970) *The Rise of Religious Liberty in America: A History*. New York: Johnson Reprint.

Curry, T. J. (1986) *The First Freedoms: Church and State in America to the Passage of the First Amendment*. Oxford: Oxford University Press.

Fischer, D. H. (1989) *Albion's Seed: Four British Folkways in America*. New York: Oxford University Press.

Fundamental Constitutions of Carolina (1669) Available at http://avalon.law.yale.edu/17th_century/nc05.asp.

Gill, A. (2008) *The Political Origins of Religious Liberty*. New York: Cambridge University Press.

Gleditsch, K. S., and Ward, M. D. (2006) "Diffusion and the International Context of Democratization." *International Organization* 60 (4): 911–33.

Grim, B., and Finke, R. (2011) *The Price of Freedom Denied: Religious Persecution and Conflict in the Twenty-First Century*. New York: Cambridge University Press.

Haefeli, E. (2012) *New Netherland and the Dutch Origins of American Religious Liberty*. Philadelphia: University of Pennsylvania Press.

Isaac, R. (1973) "Religion and Authority: Problems of the Anglican Establishment in Virginia in the Era of the Great Awakening and the Parsons' Cause." *The William and Mary Quarterly* 30 (1): 3–36.

Israel, J. I. (1991) "William III and Toleration." In O. P. Grell, J. I. Israel, and N. Tyacke (eds.), *From Persecution to Toleration: The Glorious Revolution and Religion in England*, 129–70. Oxford: Clarendon Press.

_____ (1997) "The Intellectual Debate about Toleration in the Dutch Republic." In C. Berkvens-Stevelinck, J. Israel, and G.H.M. Posthumus Meyjes (eds.), *The Emergence of Tolerance in the Dutch Republic*, 3–36. New York: Brill.

Kooi, C. (2002) "Paying off the Sheriff: Strategies of Catholic Toleration in Golden Age Holland." In R. P. Hsia and H. F. K. van Nierop (eds.), *Calvinism and Religious Toleration in the Dutch Golden Age*, 87–101. New York: Cambridge University Press.

_____ (2012) *Calvinists and Catholics during Holland's Golden Age*. New York: Cambridge University Press.

Kuran, T. (2011) *The Long Divergence: How Islamic Law Held Back the Middle East*. Princeton: Princeton University Press.

Locke, J. ([1689] 2010) *A Letter Concerning Toleration and Other Writings*. Edited by M. Goldie. Indianapolis: Liberty Fund.

Maddison, A. (2001) *The World Economy: A Millennial Perspective.* Paris: OECD Development Centre Studies.

McCloskey, D. (2006) *The Bourgeois Virtues: Ethics for an Age of Commerce*. Chicago: University of Chicago Press.

McLoughlin, W. G. (1971) *New England Dissent 1630–1833: The Baptists and the Separation of Church and State*. 2 vols. Cambridge: Harvard University Press.

Miller, P. (1935) "The Contribution of the Protestant Churches to Religious Liberty in Colonial America." *Church History* 4 (1): 57–66.

Mokyr, J. (2009) *The Enlightened Economy: An Economic History of Britain 1700–1850*. New Haven: Yale University Press.

Mout, M. E. H. N. (1997) "Limits and Debates: A Comparative View of Dutch Toleration in the Sixteenth and Early Seventeenth Centuries." In C. Berkvens-Stevelinck, J. Israel, and G. H. M. Posthumus Meyjes (eds.), *The Emergence of Tolerance in the Dutch Republic*, 37–48. New York: Brill.

North, D. C. (1990) *Institutions, Institutional Change, and Economic Performance*. New York: Cambridge University Press.

North, D. C., and Thomas, R. P. (1976) *The Rise of the Western World: A New Economic History*. New York: Cambridge University Press.

North, D. C.; Wallis, J. J.; and Weingast, B. R. (2009) *Violence and Social Orders: A Conceptual Framework for Interpreting Recorded Human History*. New York: Cambridge University Press.

Olson, M. (1993) "Democracy, Dictatorship, and Development." *American Political Science Review* 87 (3): 567–75.

Owen, J. M., IV. (2010) *The Clash of Ideas in World Politics: Transnational Networks, States, and Regime Change 1510–2010*. Princeton: Princeton University Press.

_____ (2015) *Confronting Political Islam: Six Lessons from the West's Past*. Princeton: Princeton University Press.

Penn, W. (2002) *The Political Writings of William Penn*. Introduction and annotations by A. R. Murphy. Indianapolis: Liberty Fund.

Philpott, D. (2001) *Revolutions in Sovereignty: How Ideas Shaped Modern International Relations*. Princeton: Princeton University Press.

Scoville, W. C. (1952) "The Huguenots and the Diffusion of Technology." *Journal of Political Economy* 60 (4): 294–311.

Simmons, B., and Elkins, Z. (2005) "On Waves, Clusters, and Diffusion: A Conceptual Framework." *The Annals of the American Academy of Political and Social Science* 598 (1): 33–51.

Smith, A. ([1776] 1976) *An Inquiry into the Nature and Causes of the Wealth of Nations*. 2 vols. Indianapolis: Liberty Fund.

Smith, G. L. (1973) *Religion and Trade in New Netherland: Dutch Origins and American Development*. Ithaca: Cornell University Press.

Spaans, J. (2002) "Religious Policies in the Seventeenth-Century Dutch Republic." In R. P. Hsia and H. F. K. van Nierop (eds.), *Calvinism and Religious Toleration in the Dutch Golden Age*, 72–86. New York: Cambridge University Press.

Stark, R. (2003) *For the Glory of God: How Monotheism Led to Reformations, Science, Witch-Hunts, and the End of Slavery*. Princeton: Princeton University Press.

_____ (2005) *The Victory of Reason: How Christianity Led to Freedom, Capitalism, and Western Success*. New York: Random House.

Stokes, A. P., and Pfeffer, L. (1964) *Church and State in the United States*. New York: Harper and Row.

Sweet, W. W. (1935) "The American Colonial Environment and Religious Liberty" *Church History* 4 (1): 43–56.

Weber, M. ([1930] 2001) *The Protestant Ethic and the Spirit of Capitalism*. New York: Routledge Classics.

Weingast, B. R. (1995) "The Economic Role of Political Institutions: Market-Preserving Federalism and Economic Development." *Journal of Law, Economics, and Organization* 11: 1–31.

Witte, J., Jr. (2007) *The Reformation of Rights: Law, Religion, and Human Rights in Early Modern Calvinism*. New York: Cambridge University Press.

Making Inroads: China's New Silk Road Initiative

Christine R. Guluzian

China's New Silk Road initiative is a multistate commercial project as grandiose as it is ambitious. Comprised of an overland economic "belt" and a maritime transit component, it envisages the development of a trade network traversing numerous countries and continents. Major investments in infrastructure are to establish new commercial hubs along the route, linking regions together via railroads, ports, energy transit systems, and technology. A relatively novel concept introduced by China's President Xi Jinping in 2013, several projects related to the New Silk Road initiative—also called "One Belt, One Road" (OBOR, or B&R)—are being planned, are under construction, or have been recently completed. The New Silk Road is a fluid concept in its formative stages: it encompasses a variety of projects and is all-inclusive in terms of countries welcomed to participate. For these reasons, it has been labeled an abstract or visionary project. However, those in the region can attest that the New Silk Road is a reality, backed by Chinese hard currency. Thus, while Washington continues to deliberate on an overarching policy toward Asia, Beijing is making inroads—literally and figuratively—across the region and beyond.

Cato Journal, Vol. 37, No. 1 (Winter 2017).

Christine Guluzian is a post-doctoral Visiting Research Fellow in Defense and Foreign Policy Studies at the Cato Institute. The author thanks Christopher Preble, James Dorn, and Tom Clougherty for their helpful comments.

An Ancient Trade Network Modernized

The New Silk Road is a modern-day revival of the ancient Silk Road initiated under the Western Han dynasty. The original network opened up various regions to trade by land and by sea, until advancements in maritime transport ultimately rendered land routes economically uncompetitive.

The Silk Road route traversed lands called *xi yu*, or "western regions." These lay to the west of the Yumen Pass and include today's Xinjiang region as well as Central Asia, both considered strategically significant from at least the 3rd century BC. In its broadest sense, *xi yu* encompassed areas further to the west and as far reaching as the Indian subcontinent, Europe, the Middle East, and Africa.

As is the case with most revivals, China's New Silk Road takes certain notions from the original while adapting to the present-day circumstances. Most notably, the New Silk Road initiative places emphasis on "hard" infrastructure projects (i.e., construction of roads, railways, and energy pipelines) as well as "soft" (or "smart-iron") projects such as e-commerce platforms. Trucks, trains, and pipelines carrying cargo and crude are thus to replace caravans of camels laden with silk and amber.

Another modern aspect of the New Silk Road initiative is the institutionalization of its funding mechanism. Namely, Beijing established a purpose-built $40 billion Silk Road Fund in December 2014 to support investments as part of the New Silk Road, pooling together resources from the State Administration of Foreign Exchange, the China Investment Corporation, the Export-Import Bank of China, and the China Development Bank. Beijing has also initiated a multinational funding body: the $100 billion China-initiated Asian Infrastructure Investment Bank (AIIB)—boasting 57 members including Germany, the United Kingdom, France, and Russia, but not the United States—is a development bank allocating funds for infrastructure construction projects as part of the New Silk Road. However, even China's state coffers are insufficient to meet the over $1 trillion cost associated with the New Silk Road. To bridge the gap, international and regional development banks (such as the European Bank for Reconstruction and Development, the Asian Development Bank, and the World Bank), host governments, and private-sector actors are also providing financing.

Assessing "Win-Win"

China is marketing its New Silk Road initiative as a "win-win" (*shuang ying*) scenario for all partners involved. For Beijing, benefits come in the form of a boost to the domestic economy by offsetting industrial overcapacity and opening up new markets abroad for China's new consumer-driven growth model. This is of particular importance as China adjusts to a "new normal" of single-digit growth following decades of double-digit growth. The project will also deliver greater energy security for China by providing alternate hydrocarbon transport routes aside from those it traditionally depends upon, such as the Strait of Malacca and the South China Sea. Additionally, a core aim of the New Silk Road initiative is to close the investment gap between the east and west of China in an attempt to equalize regional disparities in economic development. Economic growth is also seen as a remedy to counteract ethnic separatism in certain regions, namely Xinjiang province. All of these factors allude to a strong domestic imperative behind the New Silk Road initiative. At a time when China is undergoing a massive reform program, the New Silk Road initiative is therefore heralded as a vital component of China's evolving economic and foreign policy. In fact, it is a project mainly under the purview of the National Development and Reform Commission (NDRC).

For host governments, the New Silk Road initiative is projected to foster development in otherwise economically marginalized regions or countries. For instance, Pakistan's president has lauded the China–Pakistan Economic Corridor component of the New Silk Road as a "monument of the century" (Iftikhar 2016) benefiting billions of people in the region through the construction of new roads, an 1,800-kilometer railway line, oil pipelines, and a multi-billion dollar port at Gwadar. The New Silk Road also brands itself as an all-inclusive project, thereby benefiting emerging economies as well as established ones: a remote village on the outskirts of Kyrgyzstan could stand to gain economically as much as a European Union member-state. Germany and Poland are building key New Silk Road-related dry ports and corresponding industrial zones, for instance (Shepard 2016). Thus, on the one hand, for host governments, the New Silk Road represents an opportunity for job creation, infrastructural and economic development, and becoming a part of the global supply chain. On the other hand, the local content aspect

of New Silk Road projects are wanting: Chinese companies generally prefer using a Chinese workforce as opposed to local laborers, which may offset local job growth creation in host countries. Environmental and safety standards are also of sizable concern for host governments, as is the protection of intellectual property. Limited reciprocal market access for foreign business into China is another area of concern.

For the private sector, the New Silk Road offers new business opportunities. According to PricewaterhouseCoopers (2016), the private entities set to benefit most include businesses focused on building infrastructure, such as suppliers of technology, raw materials, equipment, and components—as well as foreign engineering, procurement, and construction companies that could partner with Chinese players in overseas markets. The preferred method for private-sector participation looks to be in the form of public-private partnerships (PPPs).[1] Indeed, Chinese companies are increasingly experimenting with public-private partnerships in infrastructure projects and bringing substantial Chinese public-sector financing to the table under the New Silk Road banner (PricewaterhouseCoopers 2016).

While China's New Silk Road project seems to promote a "win-win" scenario in terms of economic growth, infrastructure development, job creation, and interregional and international connectivity, participation in New Silk Road projects ultimately requires partnering with China's state-sponsored banks, companies, and other state actors. The potential ramifications of this will be covered in the next section.

Public- and Private-Sector Participation

As of late, private businesses in China—particularly small and medium-sized enterprises (SMEs)—have been making slow and

[1]In fact, on September 23, 2014, China's Ministry of Finance issued a notice to promote public-private partnerships in order to encourage capital investment in infrastructure.

According to the World Bank, there is no one widely accepted definition of public-private partnerships. An increasing number of countries are enshrining a definition in their laws, each tailoring the definition to their institutional and legal particularities. The PPP Knowledge Lab defines a public-private partnership as "a long-term contract between a private party and a government entity, for providing a public asset or service, in which the private party bears significant risk and management responsibility, and remuneration is linked to performance." Available at ppp.worldbank.org/public-private-partnership/overview/what-are-public-private-partnerships.

steady headway into traditionally state-dominated sectors. However, state-owned enterprises (SOEs) dominate Chinese outward foreign direct investment (FDI) (Backaler 2014). This trend will continue as the central government pushes for SOEs to play a prominent role within New Silk Road projects. In fact, the Chinese government's role in FDI has been gradually changing from that of regulator to facilitator (KPMG 2015), and SOEs are the state's commercial conduit of choice. As China's FDI increasingly channels into New Silk Road projects,[2] participation by SOEs will therefore be an inherent component of the initiative.

Within corporatist states,[3] SOEs may offer the only pathway for conducting business—at least until private companies and/or free market principles gain further ground.[4] Until then, government backing via SOEs could prove critical for investing within particular sectors, such as energy, telecommunications, or defense.[5]

[2]According to *The Economist* (2016), official estimates suggest that in 2015, China's FDI in New Silk Road countries rose twice as fast as the increase in total FDI. Forty-four percent of China's new engineering projects were signed with New Silk Road countries, and 52 percent in the first five months of 2016.

[3]The significant role played by SOEs within domestic and international commercial activities in China is comparable to the "state-corporatist" paradigm in Russia, where heavily state-backed national champions typically wield a privileged position within outward and inward FDI projects. Schmitter (1974) expounds on "state corporatism" as "a defining element of, if not structural necessity for, the antiliberal, delayed capitalist, authoritarian, neomercantilist state." Another theorist of corporatism, Mihail Manoilesco (Schmitter 1974), distinguishes between two subtypes of corporatism: *corporatism pur*, in which "the legitimacy and functioning of the state were primarily or exclusively dependent on the activity of singular, noncompetitive, hierarchically ordered representative corporations" and *corporatisme subordonne*, whereby corporations were "created by and kept as auxiliary and dependent organs of the state which founded its legitimacy and effective functioning."

[4]On the possibility of small and medium-sized enterprises gaining further ground, according to Backaler (2014), the Third Plenum of the 18th Central Committee in November 2013 expressed the new administration's intention to build a more even playing field between state-owned and private enterprises.

[5]This is especially the case when it comes to conducting commercial activity in "transition" countries, where informal socioeconomic structures often vie with formal state institutions. In China, for instance, it is a long-standing presumption that social connections (*guanxi*), and particularly personal connections to political authorities, remain indispensable for a wide range of activities, including conducting business, using courts, obtaining bureaucratic protection, boosting a company's reputation, and securing bank loans (Wang and Zhang 2014).

A similar concept, *blat,* also exists in Russia. For a good comparative study of informal practices in China and in Russia, see Ledeneva (2008).

However, SOEs present their own challenges and often operate at a loss. In China, in particular, SOEs have "more debt and worse repayment ability than private-owned firms, although they can borrow longer term because of government support" (Bloomberg 2016).[6] Nontransparency within SOEs is an additional obstacle, allowing for pervasive rent-seeking behavior, increased opportunities for corruption and bribery, and unreliable figures pertaining to costs, revenues, and operational performance. All of these factors significantly increase the risk profile of conducting business with SOEs as opposed to private-sector entities. The quality and cost of infrastructure projects could also be impacted. According to a recent study by Oxford University's Saïd Business School (Ansar et al. 2016) sampling 95 large Chinese road and rail transport projects over the last three decades, project completion times were by-and-large on time or ahead of schedule; however, actual construction costs averaged 30.6 percent higher than estimated costs, with three-quarters of transport projects in China coming in over budget. The study surmises that incentives for Chinese cadres and contractors are such that building as quickly as possible is rewarded even if performance in other areas such as cost, quality, safety, environmental impact, or public consultation processes suffer.

Partnering with SOEs undoubtedly presents a unique set of challenges for host governments and private-sector actors. While overarching reforms to state corporatist structures in China would provide a more effective long-term solution, the reality is that China's SOEs are considered of strategic importance to the state and are thus likely to retain their monopolistic positions within industry, both domestically and in FDI projects. Nevertheless, the New Silk Road presents opportunities along with challenges.

For instance, the ambitious New Silk Road initiative might incentivize SOEs to implement international best-practice standards within their commercial activities. For China, which is still acclimating to conducting business abroad, implementing best-practice standards may even become a necessary step when faced

[6]According to *The Economist* (2016), many SOEs even have a department established specifically for the New Silk Road initiative "if only in the hope of getting money for their projects."

with new business environments in foreign countries. For instance, the OECD's "Guidelines on Corporate Governance of State-Owned Enterprises" introduced in 2005 addresses issues particular to SOEs and offers advice on governance aimed at increasing transparency and efficiency.[7] Likewise, the UN's recently introduced Global Partnership for Sustainable Transport aims to assist in setting best-practice standards for environmental and safety issues particular to the transport sector—which will feature prominently in New Silk Road projects. Multilateral and bilateral treaties, joint venture contracts, and/or public-private partnership agreements also provide plausible methods for stipulating best-practice standards with regard to SOE involvement in commercial projects.[8]

For host countries, the same could hold true. The New Silk Road may motivate host governments not only to adopt international best-practice standards, but also to improve their overall investment climates in order to attract New Silk Road–oriented FDI. A good case study can be found in the Almaty Ring Road Concession Project (also known by its Russian acronym, BAKAD)—part of the "Western China–Western Europe" transnational highway and the first public-private partnership project of its type not only in Kazakhstan but in the whole of Central Asia. BAKAD was implemented as a pilot project for the New Silk Road and served as the first public-private partnership structured under a new regulatory framework, with substantial amendments to the existing legislation made in 2014 as part of a new reform program in Kazakhstan called "*Nurly Zhol*."[9] Making the effort to enact legislation in favor of public-private

[7]Available at www.oecd.org/daf/ca/OECD-Guidelines-Corporate-Governance-SOEs-2015.pdf. Also see the 2015 "G20/OECD Principles of Corporate Governance" at www.oecd-ilibrary.org/docserver/download/2615021epdf?expires=1478208852&id=id&accname=guest&checksum=58BA07C3D057A719F3828ED51DEA76DE.

[8]For example, several countries participating in B&R are also signatories to the Energy Charter Treaty, which sets provisions on dispute resolution, nondiscriminatory conditions for trade in energy materials, the protection of foreign investments, and the promotion of energy efficiency.

[9]Kazakhstan's new economic policy, "*Nurly Zhol,*" or "Bright Path," was announced by President Nursultan Nazarbayev in November 2014 to correspond with the development plans of the New Silk Road initiative.

partnerships created the conditions that enabled international investors—such as the European Bank for Reconstruction and Development—to invest in BAKAD. For instance, in May 2015,[10] Kazakhstan's "Law on Natural Monopolies" was amended to increase the transparency of tariff calculations and the activities of monopolies, streamline tariff procedures, ensure consumer protection, and facilitate the processing of bids and the elimination of unnecessary costs. In fact, BAKAD could serve as a blueprint for future public-private partnerships in sectors that need to attract funding as part of the New Silk Road initiative.

Thus, while the substantial role of public-sector actors in the New Silk Road brings challenges, the New Silk Road initiative could help pave the way for the adoption of more market-oriented corporate governance measures, which may be crucial for private-sector investors considering placing funds into the costly and long-term projects espoused by the New Silk Road.[11]

International and U.S. Implications

"Economic diplomacy" is Beijing's novel approach for engaging with the international community (Ewert, Poeter, and Fermont 2016), and the New Silk Road initiative is a key mechanism for doing so. Accordingly, Beijing and host governments consider the New Silk Road initiative and the AIIB as high-priority projects.

From Beijing's perspective, there is more to gain by pushing for soft "economic diplomacy" than by exercising "hard power" abroad. Beijing views the costliness and unpredictability of military conflict as contrary to its policy preference for cultivating sustained domestic

[10]In May 2015, Kazakhstan's President Nazarbayev also set out the "100 Concrete Steps" plan intended to implement five institutional reforms in accordance with "*Nurly Zhol,*" including: (1) the creation of a modern and professional civil service, (2) ensuring the rule of law, (3) industrialization, (4) economic growth, and (5) transparency and accountability of the state.

[11]Additional steps that private-sector actors could take to ensure best-practice standards and/or to help mitigate risks include: conducting thorough due-diligence on all potential project partners, considering political risk insurance, establishing arbitration and/or dispute resolution mechanisms from the outset, and ensuring contracts and/or treaties clearly delineate between states' rights and investors' rights.

economic growth,[12] which, in turn, it sees as crucial for ensuring domestic stability and regime continuity.[13]

Hence, Beijing is keen to stress that the New Silk Road is not a foreign policy project, but an economically oriented one. This is especially important for Beijing to emphasize because of the strategically significant geographic areas through which the New Silk Road passes.

Just as in ancient Silk Road times, Xinjiang (particularly Ürümqi and Kashgar) and Central Asia are areas of central importance for the New Silk Road initiative. Indeed, a main impetus behind the New Silk Road is to develop and engage with the Xinjiang Autonomous Region and to quell separatist tendencies and religious radicalization. Much to Russia's chagrin, Central Asia is also of core importance for the New Silk Road project: this region, traditionally considered part of "Russia's backyard," may be increasingly vied over not only for its massive energy reserves but also for its strategic location as the "buckle" within the New Silk Road's "economic belt" connecting trade links between the east and west. Kazakhstan and Kyrgyzstan are also members of the Eurasian Economic Union (EEU) along with Russia.[14] Therefore, unless mutual areas for cooperation are fostered, a geopolitical tug-of-war between Russia and China could

[12]Whether China's adherence to a state-corporatist model is an effective method for achieving sustainable economic growth is a separate topic for debate.

[13]For Beijing, economic growth trumps military confrontation, unless strategic economic interests or territorial issues are concerned. Shaojun Li, a professor at the Institute of World Economics and Politics at the Chinese Academy of Social Sciences (an institution directly under the State Council and noted as the highest academic research organization in China in the field of social sciences) states:

> For China, if there is no order in its external relations, its economic construction will suffer. In order to ensure an orderly and peaceful environment, therefore, China needs to embark on peaceful means as the preferred route for it to resolve its territorial disputes. . . . Focusing on developmental goals rather than hegemony determines China's defensive and inward military policy. The purpose of China's military construction is to serve its development. Once the military buildup exceeds the limits of self-defense, however, it tends to have a negative impact on national development because it not only utilizes domestic resources needed for development, but also creates unnecessary tensions in foreign interactions. [Accordingly,] as China changes its mode of development and expands domestic demand, it will not transform its economic development achievements into military-based power [Li 2014: 66].

[14]Other member states of the EEU include Belarus and Armenia.

ensue.[15] For its part, Central Asia's interest in trade and FDI is particularly high at present due to the global economic slowdown, lower energy exports, and economic sanctions placed upon Russia, negatively impacting Central Asian workers' remittances back to their home countries.[16] Accordingly, Central Asian countries are looking at diversifying their economies, engaging in and implementing WTO guidelines in a more expedient manner, and improving their overall investment and trade environments[17] to attract FDI, as the BAKAD project and recent legislative reform program in Kazakhstan demonstrate.

However, the participation of the Chinese state via SOEs and other state-backed entities could challenge the notion of the New Silk Road as a purely commercial, and not political, project. Should there be excessive state intervention by Beijing, or should military activity be exercised in relation to the New Silk Road, international perceptions of the New Silk Road could quickly shift from it being viewed as a trade and investment initiative to an alliance-building and/or geopolitical expansion project. Such a perception would contribute to, rather than assuage, the foreign policy debate regarding a "rising China" and its perceived threat to the international order.[18] The international community will therefore look beyond the "win-win" rhetoric driving the New Silk Road initiative and focus on the facts.

As for the United States, a substantive foreign policy toward Beijing is yet to crystallize. While Washington is still to form a collective approach toward China's New Silk Road initiative, the current outlook leans toward wholly ignoring, "countering," or "containing" China's New Silk Road initiative.[19] Donald Trump's position toward

[15]Relations between the New Silk Road and the Eurasian Economic Union will be expounded upon by the author in a forthcoming Cato Institute Policy Analysis.

[16]According to the World Bank's 2016 Migration and Development Brief, within Europe and Central Asia, estimated declines in remittance receipts in 2015 were greatest among the Central Asian countries: 47 percent in both Turkmenistan and Uzbekistan, 25 percent in the Kyrgyz Republic, 24 percent in Tajikistan, and 23 percent in Kazakhstan. Available at pubdocs.worldbank.org/en/661301460400427908/MigrationandDevelopmentBrief26.pdf.

[17]Though these measures are not evenly applicable across all Central Asian states.

[18]The chief literature presenting this perspective include Mearsheimer (2006, 2010) and Allison (2015).

[19]The author will further analyze these approaches and offer policy recommendations in a forthcoming Cato Institute Policy Analysis.

China during his presidential campaign exhibited a strongly protectionist and economically isolationist rhetoric, and it is yet to be seen whether candidate Trump's views will be in line with President Trump's actual policies. A better approach would be to look for areas of cooperation where U.S. companies could potentially benefit from New Silk Road projects.

Conclusion

In conjunction with the New Silk Road initiative, Beijing is espousing a new "economic diplomacy" model to assist in spurring long-term, sustainable domestic economic growth. In addition, China's New Silk Road initiative is offering its partners job creation opportunities, FDI, infrastructure building, and bolstered commercial exchange. The end goal, according to China, is to establish a "win-win" scenario for all partners involved.

Yet, this ambitious multinational project comes with serious obstacles: unstable political regimes within host countries; subpar international business practice standards, including nontransparency and corruption; and the potential for Chinese state involvement to politicize commercial relations via SOEs. These could all thwart positive trade relations and investment environments. Such challenges could be mitigated if private- and public-sector participants take precautionary steps, such as exercising due diligence on projects and partners, establishing clear contractual or treaty terms on dispute and arbitration mechanisms, and insisting on the application of international best-practice standards. As for host countries, the New Silk Road initiative could incentivize governments to implement free market principles within their own economies in order to better attract FDI. This must include removing or reducing tariffs, simplifying tax codes, limiting bureaucracy, providing for the protection of private property, and strengthening the rule of law.

The New Silk Road is an imperfect project in its formative stages. It is a large-scale initiative projected to span several countries and continents and is backed by the world's second largest economy: if proven successful, it would be too large a project to ignore or to "contain." The United States should approach the New Silk Road initiative cautiously yet constructively and as a potentially positive opportunity for cultivating mutually beneficial trade and relationship-building ties with China and New Silk Road participant states.

References

Allison, G. T. (2015) "The Thucydides Trap: Are the U.S. and China Headed for War?" *The Atlantic* (September 24). Available at www.theatlantic.com/international/archive/2015/09/united-states-china-war-thucydides-trap/406756.

Ansar, A.; Flyvbjerg, B.; Budzier, A.; and Lunn, D. (2016) "Does Infrastructure Investment Lead to Economic Growth or Economic Fragility? Evidence from China." *Oxford Review of Economic Policy* 32 (3): 360–90.

Backaler, J. (2014) "Hearing on US-China Economic Challenges: Panel III: Chinese State-Owned Enterprises and Non-Market Economics—Competitive Challenges." Testimony before the U.S.–China Economic and Security Review Commission (February 21).

Bloomberg. (2016) "China Provinces Meddle in Bank Loan Choices to Keep Firms Alive." Bloomberg.com (July 7). Available at www.bloomberg.com/news/articles/2016-07-07/china-provinces-meddle-in-bank-loan-choices-to-keep-firms-alive.

Ewert, I.; Poeter, J. P.; and Fermont, M. (2016) "Economics, Law and Politics: The Divisive Question of China's Market Economy Status." *EU–China Observer* 1 (16): 4–8.

Iftikhar, M. (2016) "CPEC: A monument of century." *The Nation* (August 21). Available at nation.com.pk/business/21-Aug-2016/cpec-a-monument-of-century.

KPMG. (2015) "China Outlook 2015." Available at www.kpmg.com/ES/es/Internacionalizacion-KPMG/Documents/China-Outlook-2015.pdf.

Ledeneva, A. V. (2008) "'Blat' and 'Guanxi': Informal Practices in Russia and China." *Comparative Studies in Society and History* 50 (1): 118–44.

Li, S. (2014) "China's Dual-Identity Dilemma and its Countermeasures." In S. Binhong (ed.), *The World in 2020 According to China: Chinese Foreign Policy Elites Discuss Emerging Trends in International Politics*, 51–68. Leiden: Brill.

Mearsheimer, J. J. (2006) "China's Unpeaceful Rise." *Current History* 105 (690): 160–62.

_____ (2010) "The Gathering Storm: China's Challenge to U.S. Power in Asia." *Chinese Journal of International Politics* 3 (4): 381–96.

PricewaterhouseCoopers. (2016) "China's New Silk Route: The Long and Winding Road." PwC's Growth Markets Centre (Febuary). Available at www.pwc.com/gx/en/growth-markets-center/assets/pdf/china-new-silk-route.pdf.

Schmitter, P. C. (1974) "Still the Century of Corporatism?" *The Review of Politics* 36 (1): 85–131.

Shepard, W. (2016) "The New Silk Road Is Not Chinese, It's International." *Forbes* (October 14). Available at www.forbes.com/sites/wadeshepard/2016/10/14/regardless-of-what-beijing-says-the-new-silk-road-is-not-chinese/#76405ef67289.

The Economist. (2016) "Our Bulldozers, Our Rules." *The Economist* (July 2). Available at www.economist.com/news/china/21701505-chinas-foreign-policy-could-reshape-good-part-world-economy-our-bulldozers-our-rules.

Wang, Y., and Zhang, D. (2014) "Can Clientelism Subvert Authoritarianism? Evidence from China's Housing Market." APSA 2014 Annual Meeting Paper. Available at ssrn.com/abstract=2453289.

China's Future and the Determining Role of the Market for Ideas

Ning Wang

The ultimate success of China's search for economic prosperity, cultural renaissance, and a "peaceful rise" depends, in large part, on whether a free market for ideas can reemerge and flourish in China. The concept of the "market for ideas" (*sixian shichang*) was first introduced to a Chinese audience by Ronald Coase and myself in *How China Became Capitalist* (Coase and Wang 2012, see also Coase 1974). It quickly won acceptance among academics and the media. China is the only leading economy where the production and communication of ideas remains under strict state control. Universities, the primary venue where new ideas are produced, are run by the state. Newspapers, radio and TV stations, and publishers are all controlled by the state; ideas unwelcome by the state have a hard time to see the light of day. Because the freedom to supply ideas, choose ideas, and criticize ideas is severely limited, the creativity of the Chinese people is underutilized and their innovative potential undertapped.

In the past several years, our argument has been picked up and further developed in China—most consistently and prominently by

---

Cato Journal, Vol. 37, No. 1 (Winter 2017).

Ning Wang is a Senior Fellow at the Ronald Coase Institute and is the International Director of the Ronald Coase Center for the Study of the Economy at Zhejiang University. This article is a revised and expanded version of a speech first delivered at the Cato Institute on October 6, 2016. The author thanks Jim Dorn and Tom Clougherty for comments and suggestions; the usual caveat applies.

Weiying Zhang (2015).[1] In public speeches and writings, Zhang (e.g., 2014) highlights the leading role played by ideas in energizing and transforming the Chinese economy and emphasizes the importance of a free market for ideas in facilitating political reform and sustaining economic development. Wu Jinglian (2016) is another prominent Chinese economist who has come to appreciate the importance of the market for ideas in determining China's future. Outside economics, Chinese legal scholars (e.g., Guo Daohui 2015) have also recognized the market for ideas as a critical check on state power and as a prelude to the rule of law.

The Chinese translation of our book was published in January 2013, with a different title, 变革中国：市场经济的中国之路. If translated back into English, it reads: "China under Transformation: China's Road to the Market Economy." That the Chinese publisher had to erase "capitalist" from the title and substitute "market economy" or "market system" for "capitalism" in the book is an act of self-censorship. This is a delicate art of compromise between reality and integrity, between the pressure of power and the pursuit of truth—a critical skill of survival in a society where a free market for ideas is lacking. Nonetheless, the contents and arguments of the book are kept intact. A major argument that Coase and I put forward in the book is that China's future crucially hinges upon whether it can embrace a free market for ideas.

This is what we wrote:

> As remarkable as the Chinese market transformation is, capitalism with Chinese characteristics is impoverished by the lack of a free market for ideas; this deficiency has become the most restrictive bottleneck in China's economic and social development. Ever since the start of economic reform, the Chinese government has been persistently calling for the "emancipation of the mind," but nothing is more effective than an active market for ideas in freeing people's minds.

[1]Our argument for the market for ideas, including the significance we place on the importance of ideas in the working of the modern economy, is greatly elaborated upon by Deirdre McCloskey (2016), who singles out ideas, or rather, changes in ideas, and not capital or institutions, as being mainly responsible for what she calls the "Great Enrichment" and the beginning of the modern world (see also McCloskey 2015).

> Indeed, without this, any "emancipation of the mind" is doomed. The creative minds of the Chinese people and their inventive power have been underexploited. This is unfortunate since capitalism with Chinese characteristics could definitely be more innovative and more driven by quality rather than quantity. As the largest producer of PhDs in the world, China could have contributed much more to the growth of human knowledge. In today's world, new products and industries, novel ideas and practices, flexible and innovative organizations and institutions urgently need to tackle global challenges, from poverty and disease to war, from energy conservation and water shortage to environmental protection. We simply cannot afford to set aside the human potential of one-fifth of humanity [Coase and Wang 2012: 199].

The rest of the article is organized as follows. First, I distinguish the market for ideas from democracy; this separation obtains special significance in the Chinese context. Next, I defend the freedom to partake in an open market for ideas as a basic natural right; the prevailing practice to associate the market for ideas as a bundle of political rights mischaracterizes the relationship between the state and the market for ideas. Following the defense and elaboration of the market for ideas, I then sketch a new, rather Hayekian, vision of the economy, which portrays the modern economy as an enterprise of knowledge. I conclude with a battle cry for the market for ideas.

Priority of the Market for Ideas over Democracy

Among many factors, Coase and I singled out the lack of a market for ideas as China's most vital defect; this was a novel and quite unconventional position. When our book manuscript was under review, we were censured by several reviewers for our overarching stress on this market for ideas. At the time, many China experts instead placed their emphasis on democratization—that is, on opening up the political system and introducing multiparty political competition to replace the existing party-state—as the most critical challenge facing China. When and how China will embrace democracy, and whether the Chinese Communist Party can survive democratization, were the main questions asked about China's political future. That we did not engage with these questions certainly left

some readers disappointed. Yet with no pretense of pleasing everyone, and absent any pressure to maximize readership, we stuck to our position and offered in the book a different diagnosis of the main flaw of the Chinese market economy: China has developed a robust market for goods, but it still lacks a free market for ideas. For China to become a normal country, we argued, it has to embrace—or, rather, reembrace—this market for ideas.

A market for ideas flourished much earlier in China, at the time of Confucius. During the so-called Axial Age, Confucius, Lao-zi, and Mo-zi and their followers, each established competing schools of Chinese thought. These thinkers lived in an era when the Zhou dynasty was disintegrating and China was divided into many small states, each competing with the others for wealth, power, and human talent. As learning was no longer a privilege confined to the royal house of Zhou, knowledge began to spread out in society, transmitted by private tutor houses, somewhat similar to the academies in ancient Greece. New ideas about the nature of man and society, and competing views and strategies about the pursuit of wealth and power, all burst into life. From this emerging marketplace for ideas, all the Chinese schools of philosophy were born; together these gave life and character to Chinese civilization.

Since Chinese civilization first emerged out of "competition among one hundred schools of thought" (*baijia zhengming*), the idea of a free market for ideas has acquired sacred status in Chinese history. It has remained an inspiration for the Chinese literati and is accepted as a golden benchmark by which to judge the merits and legitimacy of political regimes. The first emperor of Qin, despite his historical role in unifying China, is forever condemned as a "tyrant" (*baozheng*) for "burning books and burying Confucian scholars alive" (*fenshu kengru*). A thousand years later, the first emperor of the Song dynasty set a rule that no scholars or critics of government should be killed, laying a crucial platform for the glory of Song China.[2] Modern China

[2]While it was weak militarily, and was eventually conquered by the powerful army of Genghis Khan, Song China achieved significant advances in art, literature, science, and technology; the Chinese economy was more commercialized and life more urbanized in Song than in any previous or following dynasty. Technological advances in printing enabled mass production of books. As the cost of book printing fell dramatically, books were no longer a luxury good for the wealthy and became, for the first time in Chinese history, readily affordable. Private universities also flourished. All this contributed to a lively market for ideas in Song China.

witnessed a brief resurgence of this free market for ideas during the early decades of the 20th century, after the collapse of the Qing dynasty and before the socialist revolution. During this period, while the nation was ravaged by foreign invasion and civil wars, a modern free press and private universities boomed. It may seem ironic, but even Mao himself, probably the most damaging enemy of the free market for ideas in modern China, who executed many critics of his policies and imprisoned hundreds of thousands of intellectuals, admitted the need to "let one hundred flowers bloom and one hundred schools of thought contend." If the long history of Chinese political thought can be distilled into a single piece of wisdom, it may well be that attributed to Wei Zheng, the chief minister of the first emperor of the Tang dynasty: "listening to all sides makes you enlightened, heeding only one side leaves you in the dark" (*jianting zeming, piianting ze'an*). This dictum puts the market for ideas at the foundation of good governance and social harmony.

In contrast to the market for ideas, multiparty competition has virtually no precedent in Chinese history. Political parties emerged for the first time in China only at the beginning of the 20th century when the Qing dynasty was falling apart. Sandwiched by the obstinacy of residual imperial power and the violence of modern revolutionary ideology, political parties had little breathing room and quickly collapsed, leaving few footprints in Chinese political history. What is worse, the Chinese word for "party" has a strong negative connotation in traditional Chinese political thinking.[3]

In the West, any political party assumes as its mission securing and defending the interests of its members. In China, however, "forming a party to pursue its own self-interest" (*jiedang yingsi*) has been consistently condemned as violating justice and undermining social harmony. The Book of History (*shangshu*), one of the five Confucian classics, says, "In the absence of partialities and parties, the kingly way is broad and vast. In the absence of parties and partialities, the kingly way is level and smooth" (*wupian wudang, wangdao dang-*

[3]Since the political party did not exist in Chinese society, the Chinese lexicon had no term for it. "Dang" ( ) was first appropriated by Japanese scholars to translate the English term "party" after the Meiji Restoration. It was later accepted by the Chinese before they knew anything about the nature and role of political parties in the modern world. This poor choice of word has cast a long shadow on the fate of political parties in modern China.

dang; wudang wupian, wangdao pingping). Ouyang Xiu, a famous scholar-official of the Song dynasty, reinforced the long-standing critique of political parties. In his celebrated article, "On Parties" (*pengdang lun*), he attributed the sudden collapse of the almighty Tang dynasty to its failure to check the rise of parties.

In addition to this uniquely Chinese historical factor, there is another general reason for us to prioritize the market for ideas over democracy. Namely, that the market for ideas is, in one sense, more fundamental than democracy.

First and foremost, functioning democracy requires the presence of an effective market for ideas. Without such a market, democracy inevitably degenerates into straightforward majority rule. Indeed, without freedom of speech and the press, genuine democracy would be hardly possible. Democracy works only when genuine political debate is honored by most citizens as part of their civic duty and valued by them as the only legitimate way to resolve conflict in political life. Democracy survives only when the losing side concedes graciously, trusting their voice can still be heard in an open market for ideas and their life, property, and pursuit of happiness will be equally protected. By contrast, once people get used to accepting the rule of the majority as the triumph of democracy, democracy itself is at serious risk: the tyranny of majority is on the horizon.

Second, the market for ideas is a precondition for order and prosperity in any political regime, whether it is a democracy or not. After waves of democratization, democracy has become a dominant form of organizing political life all over the world (Dahl 1998, Tilly 2007). Even nondemocratic countries, such as China, are compelled to recognize democracy as a "good thing" (*hao dongxi*) (Yu 2009). This, however, does not change the fact that democracy is a newcomer to political life. In contrast, the market for ideas has a much longer and richer history. Baghdad stood out as the global center of learning and knowledge in the 8th century, with a booming market for ideas reaching far beyond the Arabic world, yet democracy was alien to Islam. Britain, too, was not in any meaningful way a democracy during Adam Smith's time; yet Smith had no difficulty enjoying a lively social life, engaging in debates with other scholars, and publishing books, including *The Wealth of Nations*, which in due course would change British economic policy forever. China offers another example of a place where the market for ideas once flourished with no hint of democracy.

Natural Right versus Political Right

In a society with a free market for ideas, every individual enjoys the right to create new ideas and share them with others; to criticize and debate ideas in coffeehouses, tearooms, classrooms, and newspapers, as well as at seminars, conferences, and forums, on TV programs, and on the internet; to promote and defend ideas one endorses; to ignore or repudiate those one finds misleading or wrong; to be inspired and informed by ideas; and to come up with still better ones. Such a right exists prior to the birth of the state or any government. Before the rise of politics, before the birth of the state, people must have learned to think, to record their thinking and put it down as ideas. If natural rights exist, the right to think independently and communicate freely must surely count as one of them. It follows, then, that a free market for ideas must be a basic human right—a natural right for any person as a human being.

Here we diverge from the existing economics literature, in which the market for ideas is often treated as a bundle of political rights, such as freedom of expression and the press.[4] Aaron Director (1964) was probably the first modern economist to explicitly use the term "the market for ideas," in contrast to the "the market for economic goods and services." A decade later, Coase (1974) used the term in a similar fashion in his influential paper "The Market for Goods and the Market for Ideas." As Coase (1974: 384) explained, in the context of the United States, the market for ideas is concerned with the expression of opinion in speech, writing, and other channels—that is, "activities protected by the First Amendment." For both Director and Coase, it is clear that the market for ideas refers to a set of political rights enjoyed by people in a free society.

While it is true that wherever it exists, the market for ideas is protected by the state via laws and regulations, and wherever it does not exist, the market for ideas is banned by government, we must not make the mistake of treating the market for ideas as a creation of the state or as a gift handed down from government. The main reason for emphasizing that the market for ideas is a natural right rather than a political right is to steer us away from this error.

[4]Another common usage of "the market for ideas" in economics is rather narrow, referring to the market for patentable or salable innovation and invention (e.g., Gans and Sterns 2010, Chatterjee and Rossi-Hansberg 2012).

The market for ideas belongs in a special category of human inventions that flourished before the rise of any government, because their presence and operation, notwithstanding crude forms and primitive manners, must have played a critical role in man's rise from the jungles to build human civilization. While the details of human evolution remain sketchy, and there are still many gaps in the story of human beginnings, it is clear that modern Homo sapiens faced several competing species of hominid, including Neanderthals—who branched out from the Homo sapiens lineage 500,000 years ago and became extinct only about 30,000 years ago—and Denisovans, who separated from the Homo sapiens lineage 700,000 years ago. Over a long stretch of time—much longer, anyway, than recorded human history—Homo sapiens coexisted with Neanderthals and Denisovans in the same environment; occasionally they interbred. We know little about the genetic and epigenetic advantages Homo sapiens enjoyed over their competitors.[5] But what we do know suggests that it was the social use of brainpower, rather than the brain itself, that enabled Homo sapiens to outcompete other hominids. The use of the brain in various social settings, and for various emotional and cognitive tasks, further stimulated the growth of brainpower, which in turn enabled our ancestors to conduct a lively communal life, to master the cumulative growth of knowledge, and thus to increase adaptability in the face of rapidly changing environments. In this long process, the market for ideas, in its most crude manner, allowed our ancestors to exchange information, to winnow truth from rumor, and to pass knowledge on to the next generation—which could in turn constantly revise, improve, and expand in a cumulative fashion the existing stock

[5]With the recent publication of the genomes of the two extinct Homo species, scientists can now compare the genomes of Homo sapiens, Neanderthals, and Denisovans. Apparently, the genetic differences between Homo sapiens and other hominids are much smaller than previously thought. For example, the two amino acid changes in gene FOXP2, widely associated with the emergence of modern language and other unique human cognitive capabilities, are found in Homo sapiens and Neanderthals. The KLK8 protein, which is preferentially expressed in the central nervous system and is involved in learning and memory, and which was until recently thought to be unique to Homo sapiens, has now been found in both the Neanderthals and Denisovans. According to a recent study, "*H.* sapiens, Neanderthals, and Denisova are virtually equal when the molecular aspects involved in the cognitive processes considered here are compared" (Paixao-Cortes et al. 2013: 2).

of knowledge. Long before the rise of any government, the market for ideas gave Homo sapiens an evolutionary advantage through social learning and the accumulation of knowledge.

Unlike stone tools and other artifacts that our ancestors used in their daily lives, many of which have survived and are available for us to inspect, the market for ideas that was in operation before human civilization took off remains intangible. Its nature is often abstract and its working opaque and complex. It left few footprints with which we can trace its development over time. Nonetheless, its existence, and the central role it played in empowering Homo sapiens to prosper and conquer the Earth, cannot be denied.

Besides the market for ideas, other members of this category of human inventions include the market itself, law, money, language, and knowledge.[6] With the rise of centralized state bureaucracy, these easily fell prey to naked political power. Almost without exception, they were enslaved, maimed, and deformed in the hands of political power—if they were lucky enough to survive at all. It was only with the advent of the Enlightenment—and particularly following the rise of what Adam Smith called "commercial society"—that the West entered what Deirdre McCloskey (2016) calls "the Bourgeois Era." Only then did these vital institutions gradually liberate themselves from overbearing political power and regain their independence.[7] This process is still ongoing and subject to fluctuation and reversal.

Political power can, and does, suppress the market for ideas, but it is far from its only enemy. Bad ideas often cast a long and dark shadow. The market is often put at risk by our ignorance or worse, knowledge we hold dear but is no better than half-truth. For example, the concept of perfect competition helps to breed a pervasive misunderstanding of the nature of knowledge in the working of the

[6]The social sciences rarely treat it as an institution, but knowledge is itself a human invention that came into existence long before the rise of any government. And just like the market, law, money, and language, knowledge has been subject to state intervention. The expanding role of the state in public education, as well as in research and development, is today often taken for granted, even in most developed market economies. For an early critique, see Friedman (1962: Ch. 6) and Coase (1974).

[7]In most cases, the autonomy of these institutions is now protected by laws. This has misled many to credit the state as their ultimate creator. In reality, however, these institutions emerged and developed spontaneously, without the direction of a central authority.

market system (Hayek 1937, 1945). Perfect competition presumes perfect information. With perfect information, there is little need for the market for ideas. No wonder Hayek's messages have fallen to deaf ears and mainstream economics has little room for knowledge, let alone the market for ideas. When economics leaves knowledge out of the equation, it is bound to degenerate into what Coase (1992) called "blackboard economics," an intellectual exercise detached from the real world economy.

The Economy as an Enterprise of Knowledge

Our belief that the market for ideas plays a determining role in China's future rests on our view of the economy as an enterprise of knowledge, and our conviction that the market for ideas determines the growth of that knowledge. To view the economy as an enterprise of knowledge is to come to terms with the simple fact that the growth of knowledge—broadly understood to include scientific, technical, and institutional knowledge—has always been, and will always be, the most critical force driving economic development.

In hunter-gatherer society, economic output is mainly determined by biological processes, which determine what fruits and plants are available to pick and what animals to hunt. In such a society, one's livelihood depends on local or folk knowledge. Where to fetch water and find food, how to build shelters, avoid predators and recover from illnesses, how to raise a family, and many other daily challenges of life all require local knowledge, which is often passed from one generation to the next through oral tradition and observation. When knowledge is accumulated in this manner, its growth is limited and haphazard. As long as environmental change is gradual, relative to the rate of knowledge accumulation, members of the society can manage to live a slow-paced life. But when environmental change is abrupt and rapid, leaving little time to adapt, the survival of the whole community is at risk.

In an agricultural society, the biological process is increasingly harnessed by human knowledge. Surpluses in food production make possible the rise of a nonagricultural population, including kings and lords, priests and scholars, knights and soldiers, as well as various craftsmen and merchants. After the rise of agriculture,

food supply becomes stable and human settlement becomes feasible. The market for ideas becomes institutionalized, at least for priests, scholars, craftsmen, and merchants, facilitating the accumulation of knowledge. In societies where (or during times when) knowledge is monopolized by political or religious power, the market for ideas is suppressed and the growth of knowledge is slow, if it occurs at all. This jeopardizes the survival of the whole society. In societies where (or during times when) trade is open and plays a significant role in economic life, the market for ideas has a much better chance to prosper and life is longer, healthier, richer, and easier.

Even though knowledge is crucial in hunter-gatherer and farming societies, its growth is often so slow that its critical role can hardly be recognized. The rise of manufacturing, however, puts knowledge at the center of production. Like craft production before it, manufacturing would simply be impossible without sufficient accumulation of knowledge. Manufacturing is subject to the division of labor, which in turn facilitates the growth of skills and knowledge and generates a positive feedback loop linking manufacturing and the growth of knowledge—a symbiotic relationship that hardly exists in more primitive societies.

Modern manufacturing transforms energy and material into consumer goods. Both inputs are subject to the law of preservation; the transforming force comes from knowledge, which can grow infinitely. The whole production process is directed by what consumers want and is constrained by what entrepreneurs know about the nature of their materials, the technology of production, and the art of organization. In a market economy, what consumers want changes constantly as innovation brings out waves of new products and makes the existing stock of goods obsolete. The pace and direction of this Schumpeterian process of "creative destruction" is largely determined by the working of the market for ideas.

Modern economics is divided into two separate fields, equilibrium theory and growth theory. The former is concerned with resource allocation; the latter focuses on determinants of economic progress. After Solow (1957), growth economics has gradually come to terms with the historical fact that sustainable economic growth has been driven primarily by the growth of knowledge—or technological innovation—rather that the accumulation of capital. This point was

made explicit by the recent endogenous theories of growth (e.g., Romer 1986 and Lucas 2000).[8]

But resource allocation is equally knowledge dependent. Unless economic actors hold different information and entertain different expectations about the future, and unless such information and expectations are constantly updated and revised, trade would be a one-time event, with the economy quickly moving into equilibrium. But without continuous trade, market-based efficient resource allocation would be impossible. Behind the shuffling and reshuffling of things in the marketplace, resource allocation provides a channel whereby private knowledge is revealed, business acumen is exercised, and idiosyncratic judgment is made. The frequency and intensity of trade, the characteristics of traded goods, the size of the trading network, and its diversity or heterogeneity—these are among the most important factors impacting the amount and nature of knowledge pooled together through trade.

As commonly presented in textbook economics, the seemingly innocuous assumptions of perfect information and equilibrium turn resource allocation into a mechanical problem of economizing, which simply means "shuffling around available resources in order to secure the most efficient utilization of *known* inputs in terms of a *given* hierarchy of ends" (Kirzner 1976: 79; emphasis added). Since inputs are all known, and ends are all given, all the knowledge required for resource allocation is already on the table, freely available to anyone with an interest. No effort is required on the part of economic actors to stay alert to any unexplored opportunities. Decisions are reduced to choices based on cost-benefit calculation, and there is no demand for judgment calls. In this world of zero transaction cost, there is no need for the firm (Coase 1937), the law (Coase 1960), or even the market (Cheung 1998). As this imaginary world is devoid of uncertainty, there is no room for entrepreneurship (Knight 1921). The Hayekian problem of knowledge (Hayek 1937, 1945) is completely assumed away.

[8]For a critical review of the new growth theory, see Nelson (1997). By emphasizing ideas as the engine of economic growth, the new growth theory takes an important step in recognizing the economy as an enterprise of knowledge. A common defect of the literature, however, is to assume ideas have a uniform impact across the economy.

Alfred Marshall was probably the first modern economist to foresee the growing importance of knowledge when he, as early as 1890, highlighted knowledge as the "most powerful engine of production" (Marshall [1890] 1920: 115). More than half a century later, Fritz Machlup (1962) published the first empirical investigation documenting the impact of knowledge in a modern economy.[9] Following Marshall, we accept knowledge as the most important factor of production. In *How China Became Capitalist* (Coase and Wang 2012), we take the market for ideas as a factor market for knowledge, like the capital market being a factor market for capital. Yet knowledge differs from other factors of production, and the market for ideas is unlike other factor markets. At the least, knowledge is nonrivalrous; it is this insight that has fired up the recent literature on endogenous growth. Moreover, knowledge is unfathomable, its depth endless, and its growth infinite (Bartley 1990).[10] As Popper (1985: 56) put it, "While differing widely in the various little bits we know, in our infinite ignorance we are all equal." Together with Hayek's (1937, 1945) insight on diffuse knowledge, the thesis of infinite knowledge and our endless ignorance implies that the market for ideas, as imperfect and fragile as it may be, is the most reliable solace available to save us from bigotry and idiocy.

That knowledge is unfathomable and its growth open, unpredictable, and infinite renders the knowledge-intensive modern economy non-ergodic, full of novelty and surprises. When Paul Samuelson (1969) made economics "scientific," moving it from "the realm of history" into "the realm of science," he was probably not aware that this tradeoff inevitably cut economics off from the real world economy.[11] In order to take seriously the idea that the human economy is an enterprise of knowledge, economics has no choice but to study man as he is and the economic system as it actually exists.

[9]According to Machlup's estimates, as early as 1958, the size of what he called "the knowledge economy" accounted for 29 percent of U.S. GDP. See Langlois (1985) for an account of Machlup's efforts to come to terms with the economics of knowledge.

[10]Zhuang-zi, a contemporary of Mencius, was one of the first philosophers to stress the infinity of knowledge, in contrast to the shortness of life. Among contemporary philosophers, Karl Popper and Michael Polanyi have taken great pains to develop the thesis and elaborate its profound implications. For an accessible and updated account, see Deutsch (2011).

[11]North (1999) and Davidson (2012) are among the minority who emphasize the non-ergodic nature of the modern economy.

Conclusion

China's market transformation in the past four decades is essentially an entrepreneurial revolution; it is a triumph of the market. Far from a state-led economic program, it resulted from "marginal revolutions"—initiatives undertaken by economic actors marginalized during Mao's radical socialism, such as farmers and unemployed city residents, as well as local officials. It was their ingenuity and tireless efforts to feed and clothe their families—steering away from starvation and poverty—that fortuitously transformed the Chinese economy. China's market transformation thus offers the latest example of what Hayek (1967: Ch. 6) called the "unintended consequences of human action." In this process, the Chinese government's role was mainly "to emancipate the mind" (*jiefang sixiang*), to free it from Mao's radical ideology, and allow it to approach economic development pragmatically. Deng Xiaoping and his comrades quickly realized that the state had to withdraw from economic planning and open the economy to the market and entrepreneurship. The rest is history.

As remarkable as it is, China's market transformation in the past four decades only consisted of the first stage of industrial revolution, a process mainly of catching up. In the decades to come, as China moves closer to the technological frontier, it will have to become far more innovative. Without a free market for ideas, China would certainly be ill positioned to face up to the challenge. Despite the many serious problems that China faces today with its much slower-growing economy, I remain cautiously optimistic that the Chinese people will choose to stand on the right side of history. The disaster of central planning is still too fresh to forget, and the fruit of economic freedom has not yet reached the most disadvantaged.

For the first time in the modern era, China has begun to come to terms with its history, as well as the outside world and its position in it. Chinese tourists now discover Mao's shadow in North Korea, seek the peace of Buddhism in Thailand and Bhutan, appreciate the past glory of Tang and Song in Kyoto and Nara, and go after the vagaries of modern fashion in Paris, London, and New York. At the same time, Arab merchants launch their enterprises in Yiwu, African traders seek their dreams in Guangzhou, western bankers and investors dig for their "first bucket of gold" in Shanghai and Shenzhen, and architects and designers from around the globe build

their architectural wonders all over China. In short, what Coase and I wrote a few years ago remains true today:

> China's embrace of both its history and globalization leads us to believe that Chinese capitalism, which just started its long journey, will be different. This is desirable not just for China, but for the West and everyone else as well. It is also desirable for the global market economy. Today, biodiversity is recognized as vital for sustaining our natural environment. Institutional diversity plays a similar role in keeping human society resilient. Capitalism will be much more robust if it's not a monopoly of the West, but flourishes in societies with different cultures, religions, histories, and political systems. While trade in the global market for goods makes war too expensive to fight, a global market for ideas can accommodate and thrive on the clash of ideas but steers us away from the clash of civilizations [Coase and Wang 2013: 10].

References

Bartley, W. W. III (1990) *Unfathomed Knowledge, Unmeasured Wealth: On Universities and the Wealth of Nations.* LaSalle, Ill.: Open Court.

Chatterjee, S., and Rossi-Hansberg, E. (2012) "Spinoffs and the Market for Ideas." *International Economic Review* 53 (1): 53–93.

Cheung, S. N. S. (1998) "The Transaction Cost Paradigm." *Economic Inquiry* 36 (4): 514–21.

Coase, R. (1937) "The Nature of the Firm." *Economica*, n.s. 4 (16): 386–405.

_____ (1960) "The Problem of Social Cost." *Journal of Law and Economics* 3 (October): 1–44.

_____ (1974) "The Market for Goods and the Market for Ideas." *American Economic Review* 64 (2): 384–91.

_____ (1992) "The Institutional Structure of Production." *American Economic Review* 82 (4): 713–19. (Coase's Nobel Lecture, 1991).

Coase, R., and Wang, N. (2012) *How China Became Capitalist*. London and New York: Palgrave Macmillan.

_____ (2013) "How China Became Capitalist." *Cato Policy Report* (January/February): 1, 8–10.

Dahl, R. (1998) *On Democracy*. New Haven: Yale University Press.
Davidson, P. (2012) "Is Economics a Science? Should Economics Be Rigorous?" *Real-World Economics Review* 59: 58–66.
Deutsch, D. (2011) *The Beginning of Infinity: Explanations that Transform the World*. New York: Viking Press.
Director, A. (1964) "The Parity of the Economic Market Place." *Journal of Law and Economics* 7: 1–10.
Friedman, M. (1962) *Capitalism and Freedom*. Chicago: University of Chicago Press.
Gans, J., and Stern, S. (2010) "Is There a Market for Ideas?" *Industrial and Corporate Change* 19 (3): 805–37.
Guo, D. (2015) "The Rule of Law in China and the Rule of Law in the World" (*Fazhi Zhongguo zai Fazhi Shijie zhong de Jiaose*). Available at: www.unirule.org.cn/index.php?c=article&id=3781.
Hayek, F.A. (1937) "Economics and Knowledge." *Economica,* n.s. 4: 33–54.
_____ (1945) "The Use of Knowledge in Society." *American Economic Review* 35 (4): 519–30.
_____ (1967) *Studies in Philosophy, Politics, and Economics*. Chicago: University of Chicago Press.
Kirzner, I. (1976) "Equilibrium vs. Market Processes." In E. Dolan (ed.), *The Foundations of Modern Austrian Economics*, 115–25. Kansas City: Sheed and Ward.
Knight, F. (1921) *Risk, Uncertainty and Profit*. Boston: Houghton Mifflin.
Langlois, R. (1985) "From the Economics of Knowledge to the Knowledge of Economics: Fritz Machlup on Methodology and on the 'Knowledge Society.'" *History of Economic Thought and Methodology* 3: 225–35.
Lucas, R. (2000) "Ideas and Growth." *Economica* 76 (301): 1–19.
Machlup, F. (1962) *The Production and Distribution of Knowledge in the United States*. Princeton: Princeton University Press.
Marshall, A. ([1890] 1920) *Principles of Economics*. London: Macmillan.
McCloskey, D. (2015) "Ideas, Not Interests or Institutions, Caused the Great Enrichment." *Man and the Economy* 2 (1): 7–24.
_____ (2016) *Bourgeois Equality: How Ideas, Not Capital or Institutions, Enriched the World*. Chicago: University of Chicago Press.

Nelson, R. (1997) "What is New in New Growth Theory?" *Challenge* 40 (5): 29–58.

North, D. C. (1999) "Dealing with a Non-Ergodic World: Institutional Economics, Property Rights, and the Global Environment." *Duke Environmental Law and Policy Forum* 10: 1–12.

Paixao-Cortes, V.; Henriques Viscardi, L.; Mauro Salzano, F.; and Hunemeier, T. (2013) "The Cognitive Ability of Extinct Hominins: Bringing Down the Hierarchy Using Genomic Evidence." *American Journal of Human Biology* 25 (5): 702–5.

Popper, K. (1985) "Knowledge without Authority." In D. Miller (ed.), *Popper Selections*, 46–57. Princeton: Princeton University Press.

Romer, P. (1986) "Increasing Returns and Long-Run Growth." *Journal of Political Economy* 94 (5): 1002–37.

Samuelson, P. (1969) "Classical and Neoclassical Theory." In R. W. Clower (ed.), *Monetary Theory*, 170–90. London: Penguin.

Solow, R. (1957) "Technical Change and the Aggregate Production Function." *The Review of Economics and Statistics* 39 (3): 312–20.

Tilly, C. (2007) *Democracy*. New York: Cambridge University Press.

Wu, J. (2016) "China's Economic Malaises Lie in Its Lack of a Market for Ideas" (*Zhongguo Zhuduo Jingji Bibing, Genyuan Zaiyu Quefa Sixiang Shichang*). Available at http://pit.ifeng.com/a/20161104/50204664_0.shtml.

Yu, K. (2009) *Democracy Is a Good Thing*. Washington: Brookings Institution Press.

Zhang, W. (2014) "The Market for Ideas Pushes China's Reforms" (*Sixian shichang tuidong zhongguo gaige*). *China News Weekly* (April 14).

_____ (2015) "The Power of Ideas and Leadership in China's Transition to a Liberal Society." *Cato Journal* 35 (1): 1–40.

Book Reviews

Monetary Regimes and Inflation: History, Economic and Political Relationships, 2nd ed.
Peter Bernholz
Cheltenham, UK: Edward Elgar, 2015, 240 pp.

An extensive academic literature exists on central bank independence and economic performance, especially inflation. The conclusions drawn from these studies are obtained from regressions utilizing fairly recent data from both developed and underdeveloped countries. They draw a similar conclusion: when central bank independence is defined as the ability to refuse to finance a government's budget deficit, it improves the inflation performance of the economy. However, almost all of these studies suffer from one major deficiency: the data they use commingle observations from both fixed and flexible exchanges rate regimes. (For a discussion of these studies and their methodologies, see Marc Labonte, *Central Bank Independence and Economic Performance: What Does the Evidence Show?* Congressional Research Service, Report RL31955.)

Peter Bernholz's *Monetary Regimes and Inflation: History, Economic and Political Relationships* is akin to this literature although, curiously, it is not mentioned in his study. This is the second edition of Bernholz's book. The first appeared in 2003 and much of its material is carried over to the second. There, are, however, several changes. First, Bernholz added one new hyperinflation to his discussion, the 2007–08 episode in Zimbabwe, and the book's

Cato Journal, Vol. 37, No. 1 (Winter 2017).

cover now contains a picture of the 100 trillion dollar note issued by its central bank. That episode receives scant mention, however, other than that it shares performance characteristics with similar episodes.

Also, 19 additional pages of text are now included that deal with two subjects. The first, for many, is a surprise: during the financial crisis beginning in 2007, several countries (especially the United States) experienced a large increase in their monetary base, of a magnitude often associated with high or hyperinflations, without such a calamity occurring (at least not yet). The second subject, which concludes the book, seeks to explain how monetary regimes that yield stable (or low) inflation arise and why they exist for long periods.

Bernholz's methodology consists of reviewing a large number of inflationary episodes, both ancient and modern, to see if they display consistent patterns from which conclusions can be drawn. Most of the episodes are well known and extensively researched. His discussion adds little to what is already known, but the episodes are of interest when studied collectively to see what they share in common. This is ably done. Only two pages of text out of 213 are given over to econometric estimates, and these are related to currency substitution during the Soviet hyperinflation of 1923–24. The remainder of the numerical presentation is confined to tables and graphs. Unfortunately, for some, this may be a shortcoming of the book. While the tables are in a standard format (numbered sequentially within each chapter with their titles appearing at the top and their notes and sources at the bottom), the graphs are not. Graph titles are presented on the bottom under the notes and sources, and many are so cluttered with data for the various countries distinguished by different shaped geometric symbols as to make them almost unreadable. Those places in the text where the discussion is mainly mathematical are marked by an * and can be omitted if desired without compromising the usefulness of the book.

Fortunately, each chapter ends with a series of numbered conclusions. I say this because I found the book difficult to read. It would have benefited from someone other than the author serving as translator. All one has to do is to compare Bernholz's ungraceful prose with the lucidity of Leland Yeager's in *Experiences with Stopping Inflation* (American Enterprise Institute, Studies in Economic Policy, 1981), a book that convers much the same ground.

The organizing framework of the book and its principal conclusions are set out in Chapter 2, "Inflation and Monetary Regimes." Using the equation of exchange, the proximate cause of inflation is identified as money growth in excess of the growth in real output or an increase in money per unit of output. Thus, nonmonetary causes of inflation are dismissed. The linkage of money to inflation leads to a distinction between metallic and discretionary paper money regimes. Inflation can occur in both, but the latter is the more likely culprit since it gives free rein to the inflationary predisposition of politicians, the villain in the piece. Constraining them requires independent central banks (or currency boards) and fixed exchange rates. Given the importance of central bank independence to the conclusions of this book, more discussion should have been devoted to defining what it means since a consistent definition is not to be found in the literature. Bernholz appears to have in mind freedom from the influence of political authorities, although what this freedom is remains unspecified.

Chapters 3, 4, and 5, comprising about half the book, deal sequentially with inflations in metallic regimes and those using fiat paper money. In the latter, moderate inflations are distinguished from hyperinflations with most of the discussion focused on hyperinflation. Many historical episodes are reviewed. Collectively they show that severe inflations display five regularities over their life cycles, and these dominate much of the discussion. First, depreciating money drives out of circulation stable-valued money (Gresham's Law). Second, budget deficits became larger over time since the real value of tax revenue falls between the time it is collected and disbursed (Tanzi's Law—this should be called Keynes's Law—see page 52 of his *Tract on Monetary Reform,* 1923). Third, real exchange rates become undervalued (Bernholz's Law). Fourth, the inflationary increases of money first stimulate output and employment and later the price level. As the inflation intensifies, the stimulatory effects of additional money issues diminish and output falls. Fifth, toward the end of each hyperinflation, as domestic money becomes worthless, transactions are increasingly made with stable foreign monies or, good money drives bad money out of circulation (Thiers's Law). (In Chapter 6, a rigorous mathematical model setting out this life cycle is developed.)

Interspersed in these three chapters is a substantial discussion of the costs inflation imposes upon a society in terms of output losses,

the damage it can do to the financial system, its ability to redistribute income, and the social strife it engenders. Let me note three omissions or problems with this discussion.

First, the discussion of inflation during the American War for Independence is incomplete. In Figure 4.5, the emission of the Continental currency ends in December 1779, yet the inflation continued at an accelerated pace for about another year and a half when the data end. Why this occurred is a mystery. Apparently it is associated with the successor currency to the Continental. Moreover, it would appear that a date is misplaced in this discussion. For the discussion on page 58 to make sense, the date at the top of the page should be 1780 and not 1790.

Second, the decline in output as inflation intensifies is attributed to the inefficiencies induced as barter is substituted for money exchange and the distorting effect of changes in relative prices. While this is true, another force is also at work—disintermediation of the banking system. Banks are a very important part of the financial system for most of the 30 countries noted and, during hyperinflations, the use of checks diminishes considerably and few hold saving or time deposits. The conversion of these deposits into other assets, especially currency, contracts the banking system and with it the credit base of the economy. Without credit, the wheels of production come to a grinding halt.

Third, the undervaluation of the inflating country's currency is attributed to currency substitution—Thiers's Law at work. Another reasonable alternative is available: capital flight. Not only do individuals in inflating countries get their capital out, but lending by foreigners dries up. If the goods and services being traded internationally are not perfect substitutes, the real exchange rate will have to depreciate to produce a net export of goods and services, the counterpart of the net import of stable-valued foreign financial paper.

Chapters 7 and 8, some 60 pages, explain how moderate and hyperinflations have been brought to an end. This distinction is necessary since, in hyperinflations, the real value of the national currency is reduced to near zero whereas this is not true in more moderate inflations—in these cases, it may actually rise.

Given Bernholz's belief in the inherent inflationary bias of politicians, the stabilization of moderate inflations requires both the reduction of money growth and the adoption of a monetary regime and constitution that bind the hands of politicians. His suggestions favor

fixing exchange rates to metal or stable-valued currencies, independent central banks, and currency boards. This, of course, rules out a monetarist option since it would require a central bank bound by a money growth rate rule operating in a flexible exchange rate regime.

The 30 hyperinflation cases, it is claimed, also reflect the inflationary biases of politicians, and all have occurred in discretionary paper money regimes and were (proximately) caused by large public-sector budget deficits that were largely financed by money creation. Thus, in addition to anchoring the monetary regime, stabilization requires eliminating the budget deficit through fiscal reform in which explicit taxes are substituted for the inflation tax and expenditures are brought into line with expected revenue. Also crucial to success: the public must believe that these changes are substantive and will be adhered to both in the short and long run. If this occurs, the demand for national money will increase and permit a one-time noninflationary increase in its supply. Not to accommodate this increase in demand will lead to rising real interest rates and appreciating exchange rates, both of which could damage the nascent recovery. Bernholz is not optimistic about the ability of the public to form correct expectations about stabilization policies.

While many elements in the stabilization strategy suggested by Bernholz would find wide agreement among economists, I have substantial doubt that hyperinflations are set in motion by an inflationary predisposition of politicians. Even a country with an independent central bank and a restrictive fiscal setup can fall prey to hyperinflation if the right circumstances occur. In the ones that I and my coauthors have studied (see *Studies in Hyperinflation and Stabilization*, Center for Financial Stability, 2014), the shocks that set them in motion would have done so even with independent central banks or constrained fiscal policy. Inflation in a number of these episodes was simply an alternative form of taxation forced on governments because they could impose no other. Resort to the inflation tax reflected the economic and political facts of life. History, however, teaches that the inflation tax, like other taxes, is subject to a Laffer curve and ultimately, as the rate of inflation intensifies, it produces declining amounts of revenue since the tax base vanishes as velocity rises. When revenue becomes very small, the government (or, as Bernholz notes, the government that succeeds it) is forced to enact a stabilization program since the only options remaining for survival are forced requisitions and foreign aid.

There is one subject not touched upon in this book: the longer-run legacy of the monetary regime put in place to stabilize inflating economies. Bernholz favors fixed exchange rates anchored to some metal, preferably gold. These regimes have some undesirable side effects: (1) real exchange rate adjustments to shocks require national price levels to change; (2) they can be deflationary over the long run; and (3) they are not cheap to operate.

Overall, this book provides the economist with a good picture of many countries over a long time period that were affected by inflation, how they dealt with it, and what measures contributed to success. Bernholz brings a wealth of experience and knowledge to this subject.

Gail E. Makinen
McCourt School of Public Policy
Georgetown University

The Power and Independence of the Federal Reserve
Peter Conti-Brown
Princeton: Princeton University Press, 2016, 368 pp.

Why should the president and Congress defer to the Federal Reserve on monetary policy? One of the typical justifications for central bank independence is that politicians are liable to artificially juice the economy, especially before elections, boosting economic activity through an expanded money supply in the short run but causing inflation in the medium to long run. This conflict of interests necessitates insulating central banks from political pressure. It is what Peter Conti-Brown calls the "Ulysses/punch-bowl" justification of Fed independence. Like Ulysses tied to the mast, the Fed should be shielded from the siren calls of Congress and other elected officials and left free to follow former Fed chairman William McChesney Martin's ideal of taking away "the punch bowl when the party is really heating up," that is to temper money growth in economic expansion—a move short-term-minded politicians might oppose.

Conti-Brown's *Power and Independence of the Federal Reserve* is at heart an attempt to challenge the Ulysses/punch-bowl justification for the Fed's independence. In Conti-Brown's conception, while the Ulysses/punch-bowl reasoning might accurately explain the general need for insularity from politics in the conduct of monetary policy, it

does not explain the extent and particularities of the Federal Reserve's specific power and independence. That's because the Fed, by practice and statute, does more than attempt to manage nominal income and inflation—Fed policymakers are also "recession fighters, bankers, financial regulators, bank supervisors, and protectors of financial stability."

It's worth asking at this point: Is Conti-Brown, in posing a justification of Fed independence that he himself knocks down, simply attacking a straw man? Even if the phrase "Ulysses/punch-bowl" is of Conti-Brown's own invention, it nevertheless captures a common perception of the Fed—namely, that its role in managing money specifically validates its unique autonomy compared to other government agencies.

Conti-Brown chose not to write an institutional or legal history of the Fed. Instead, he identifies a few key aspects of the Fed's structure and practical operation, which he then explores in detail. Conti-Brown attempts to use these selective inquiries to elucidate his main point: that the general need for central banks to conduct monetary policy independently from political pressure does not justify the degree to which the Fed is insulated from political oversight. He divides the book into four sections: (1) "The Federal Reserve Is a 'They' Not an 'It'"; (2) "The Five Hundred Hats of the Federal Reserve"; (3) "The Sirens of the Federal Reserve"; and (4) "The Democratic Demands of Fed Governance."

In the first section, Conti-Brown argues that none of the laws that codified the Fed's structure embodied any coherent logic regarding the need to separate monetary policy from politics. The Fed's structure is more a historical accident than intelligent design. According to Conti-Brown, the 1913 Federal Reserve Act (which established the Fed and remains the most important law shaping its structure) contained elements of influential German-American banker Paul Warburg's support for a powerful, privately managed central bank, Senator Carter Glass's preference for a decentralized set of regional banks, and President Woodrow Wilson's faith in technocratic expertise. The end result, a mix between a fully centralized and a regional system governed by a mix of public and private agents, represented this contingent melding of ideas.

In the second section, Conti-Brown discusses Fed activities that extend beyond its regulation of the money supply—the main thrust being that the Fed's oversight of the financial system occurs with

just as much independence as its monetary policymaking, even though the Ulysses/punch-bowl independence justification does not apply to it.

Financial stability has always been a large part of the Fed's activity; indeed, it was concern about financial crises that, on the surface at least, motivated the Fed's creation in 1913. Even as the Fed became primarily known for its macroeconomic stability role and the practice of what is often referred to as "conventional monetary policy," it not only maintained its original financial stability tools but also acquired new tools and regulatory functions. New Deal laws enacted the Fed's now infamous 13(3) emergency lending power, and while the simultaneous creation of deposit insurance and the FDIC is largely understood to have removed the Fed from the "front lines" of combating bank failure, the Fed has regularly provided emergency liquidity through its discount lending tools since the 1984 Continental Illinois bailout. Recently, the financial crisis and its aftermath brought the Fed's power to make loans in the name of stability in full view.

The Fed has also obtained significant financial regulatory obligations that are not part of its original function. The 1956 Bank Company Holding Act, which removed various prohibitions on both intra- and interstate branch banking, moved many bank oversight responsibilities from the Office of the Comptroller of the Currency to the Fed. The 2010 Dodd-Frank law added the monitoring of "systemic risk" to the Fed's plate by creating the Financial Stability Oversight Council (FSOC). While FSOC is a super-council composed of representatives from many financial regulatory agencies, Conti-Brown points out that the Fed's role on FSOC is that of a "first among equals." Considering the growth in both the Fed's financial regulatory duties and emergency lending, Conti-Brown asks, "Does the Ulysses/punch-bowl justification apply to these diverse, nonmonetary regulatory functions?" His answer is a resounding "no!" Yet, as he points out, Fed decisionmaking in these areas happens with the same degree of autonomy as in conventional monetary policy.

In the third section, Conti-Brown analyzes the Fed's governance structure and its relations with other branches of government. Administrative agencies are generally subject to oversight by the legislative and executive branches through the appointment and confirmation of senior-level officials and the budget appropriations process. As with other agencies, the president appoints and the

Senate confirms the Fed chair and other senior staff, namely the Board of Governors. However, some members of the Federal Open Market Committee (FOMC)—the body that oversees and votes on the Fed's buying and selling of securities—are not appointed or confirmed. The FOMC consists of 12 members, the 7 governors, of which one is the chair, and 5 of the 12 presidents of the regional Federal Reserve Banks. These presidents serve one-year terms on a rotating basis, and, despite their important role in crafting and approving monetary policy, they are appointed by each reserve bank's private board of directors (the president of the New York Fed is a permanent voting member of the FOMC). Every other official at a federal agency with equivalent responsibility is subject to presidential appointment and congressional confirmation. Also distinguishing the Fed's oversight is that, unlike most other agencies, the Fed does not receive funding through the annual congressional appropriations process. Instead, its budget comes from interest earned on the financial assets it holds for the purposes of conducting monetary policy.

There is much to commend in *The Power and Independence of the Federal Reserve.* Conti-Brown brings a sharp toolset for his institutional analysis approach: a law degree, a soon-to-be-completed doctorate in history, and, based on the bibliography, wide reading in the political science and legal literature on American administrative agencies and bureaucracy. Even knowledgeable readers who know the facts might find some of the analytical points novel and insightful.

There is more concrete value in his work besides originality and interestingness. Thinking of the Fed's function as historically contingent and seeing its structure through the lens of governance broadens the discussion beyond presentist and minimalist conversations about the level of interest rates. The book will hopefully provoke thought among readers who are frequently concerned about Fed policy but take its structure and power as an unalterable given. Last but not least, his main point, that the potentially corrosive influence of politics on monetary policy does not justify the Fed's particular insularity and discretion is true and worth noting.

Conti-Brown strays from solid ground in the fourth and final section, however, which includes reform proposals. That's because he desires, even after showing that Fed insularity as it currently exists is not justified, to "provide more accountability without compromising that crucial insulation of monetary policy from partisan politics." Maintaining day-to-day insularity while fostering democratic

accountability might sound nice on paper, but it falsely assumes a dichotomy between democracy and politics. It is ultimately a buzzword idea that does not substantively address the problems with Fed governance that Conti-Brown successfully explicates.

Conti-Brown proposes two sets of major reforms to ensure "a central bank that will protect the currency from the winds of electoral politics, without losing the benefits of democratic legitimacy." These are altering the term lengths of the Fed Chair and other members of the Board of Governors and increasing the number of Fed staff subject to the presidential appointment process.

The suggested changes to term lengths make sense, and there is a case to be made for increasing the number of key positions (the regional bank presidents especially) subject to appointment and review. But Conti-Brown does not show that increasing the scope and frequency of presidential appointments and congressional confirmations will have the effect of fostering greater democratic accountability. While it is true that the president and legislators with the respective duties of appointing and confirming bureaucratic officials are elected representatives, whether the politically controlled appointment process democratically justifies broad grants of discretion to unelected officials is very much a matter of open debate among administrative law scholars. Scholars who see agency discretion as democratically justified make plain that the appointment power is just one of a host of tools available to government's three main branches to control the bureaucratic "fourth branch"—tools like budget setting, judicial review, and the ability (of Congress) to call punitive hearings, mandate information disclosure, and even write legislation that revises a specific power delegation. Fed actions are notably subject to much less strict standards of judicial review than any other agency, and the Fed's ability to fund itself through monetary policy is of course unique as well.

Conti-Brown doesn't propose any changes to judicial review of the Fed or the Fed's funding mechanism, nor does he support any current legislative proposals designed to increase Fed transparency or limit its discretion. Moreover, Conti-Brown does not engage with the points made by skeptics of the political control justification. Scholars point out that the technical nature of the matters dealt with by agencies means that regulators often have a knowledge-based upper hand over congressional overseers and that the voting public puts little pressure on elected officials regarding administrative concerns. It is

presumptuous for Conti-Brown to slap the democratic accountability label on his reform proposals without considering these well-worn and important debates about the effectiveness of the appointment process and the other various levers of control possessed by the three branches in providing democratic legitimacy for the policies made by agencies with wide statutory discretion.[1]

Furthermore, even if the increased use of appointments led to a Fed with power that was better justified on democratic grounds, would that Fed still be as insulated from politics? There is no wide chasm between electoral politics and democracy, and any increase in democratic accountability would thereby correspond with an increase in the importance of "day-to-day politics" to Fed governance. Conti-Brown actually unwittingly demonstrates this when discussing unfilled spots on the Board of Governors during the Obama administration. He mentions the case of MIT economist Peter Diamond, a 2011 Obama appointee rejected by the Senate. "His rejection is inexcusable and may be the most egregious example of politics over substance in the history of Fed appointments," according to Conti-Brown. Yet, for better or worse, it is reasonable to assume that increasing the use of appointments would lead to more political fighting over appointees. And it is difficult to draw a line between reasonable vetting and political pettiness; seemingly low-brow confirmation debates and hearings often serve as an important site for hashing out ideological disputes and reaching pluralistic compromise.

This is not to say that increasing the use of appointments would end the Fed's ability to set monetary policy independent of day-to-day political pressure. But Conti-Brown's unwarranted distinction between politics and democratic accountability leads him to reject, in the name of protecting it from pernicious day-to-day political influence, a key reform that would increase the Fed's democratic accountability.

[1]See Jodi Short, "The Political Turn in American Administrative Law: Power, Rationality, and Reasons," *Duke Law Journal* 61 (2012): 1812–81 for an explanation of the political control model, and see Mark Seidenfeld, "The Role of Politics in a Deliberative Model of the Administrative State," *George Washington Law Review* 81 (2013): 1403–16 for a review of the literature supporting the model, and 1416–24 for a review of literature critiquing it.

That proposal is for continual GAO audits of the Fed. Iterations on this idea have been proposed in Congress since the 1970s. It seems that increasing the amount of information about the Fed's activities would foster accountability. Conti-Brown agrees that information disclosure is good in principle, but worries "that members of Congress will use cherry picked quotes and facts about the Fed's policies from the annual GAO audit to score political points and seek to influence the Fed by confusing the public and making the Fed appear more sinister or partisan than it is." Conti-Brown's mix of paeans to democratic accountability with skepticism of day-to-day politics shows his proposal's theoretical incoherence. An attempt to "have more accountability and more insulation" (as Conti-Brown puts it) is an attempt to have cake and eat it too.

I sympathize with Conti-Brown's train of thought though. The Ulysses/punch-bowl idea does not justify the Fed's structure, and human history is riddled with politicians and other sovereign authorities ruining their countries' currency. Is there a way to structure our monetary institutions in a manner congruent with our liberal democratic government that does not increase the risk of 1970s inflation? There is, and it is revealed by a better reading of the Homeric Ulysses and the sirens story, which Conti-Brown should have considered more closely. Ulysses had his men chain him to the mast of his ship when faced with the temptation of the siren calls. Ulysses could not escape the influence of the sirens, so he enabled a mechanism, a rule, to ensure that short-term temptations would not veer him off course. Likewise, strict, unalterable monetary rules (Ulysses instructed his men to kill him if he attempted to break his chains) are the best mechanism for preventing the Fed from being led astray by politics, other distracting influences, or simply the limits of its own knowledge.

Conti-Brown briefly mentions and rejects proposals to legislate or constitutionalize a monetary rule. His logic is that such rules are anti-democratic; if voters would like to have the famous Taylor rule used, for example, they should simply clamor for John Taylor to be appointed Fed chair. Aside from overlooking that the economic advantages of monetary rules relate to them being truly binding, Conti-Brown errs by using democratic legitimacy as the only grounds to judge the efficacy of Fed governance reform. The problem of unjustified Fed insularity is a problem of liberal principles, of rule of law, and constitutional governance. It would have behooved

Conti-Brown, given his critical discussion of independence, to engage with the literature on liberalism and sound money, on rule of law and monetary constitutions, and on the knowledge problem and the failures of central bank technocracy.

Ari Blask
Cato Institute

Politics as a Peculiar Business: Insights from a Theory of Entangled Political Economy
Richard E. Wagner
Cheltenham, UK: Edward Elgar Publishing, 2016, 234 pp.

Over the past few decades, insights from complexity theory and networking analysis have increasingly infused the social sciences. A complex economics perspective stresses the interactional processes between heterogeneous individuals, fallible yet capable of seeking exchange advantages and whose actions unfold in both time and space.

Despite the growing popularity of complexity approaches in economics, it is reasonable to suggest the subdisciplines of public finance and public-sector economics have been, and still remain, largely immune to process-oriented thinking.

The exclusion of complex systems approaches from public economics is illustrated by how fiscal policy action is treated in most textbooks. Convention paints the comparative static portrait of an omniscient, benevolent planner aloofly injecting tax and subsidy interventions into markets, which allegedly fail to achieve optimal resource allocations or desirable distributional outcomes.

The work of George Mason University economist Richard Wagner directly challenges that orthodoxy. He emphasizes an evolutionary political economy wherein commingled economic and political actors all interact on the same social plane. Wagner's new book, *Politics as a Peculiar Business*, is the latest in his longstanding attempts to illustrate an alternative approach to fiscal theorizing and political economy.

What is most refreshing about *Politics as a Peculiar Business* is the intellectual honesty with which Wagner depicts economic and political life in the contemporary age. Breaking down the conventional vision of bifurcated economic and political spheres, he sees a

remarkable similarity of observed behaviors regardless of institutional setting: "Political entities compete among one another just as do market entities, and with political and market entities also engaging in both competitive and collusive activities."

Although the pro-social Homo sapiens regularly demonstrates an inclination to compete and collaborate with one another, whether in a market or political setting, Wagner believes that the incentive structures within the economic and political spheres greatly differ.

Market participants try to create and realize economic value given the existence of private property, relative prices, and monetary profit-loss signaling. Politicians, on the other hand, seek prestige and fame, oftentimes by attempting to secure a policy legacy for themselves, and they do so in the absence of property rights, prices, and monetary profit and loss.

Furthermore, interaction between individuals in the market generally consists of mutually beneficial, value-added exchanges undertaken voluntarily. Wagner labels this "dyadic exchange"; they are the kinds of "win-win" economic interactions lauded by classical liberals as conducive to economic prosperity and social peace.

By contrast, there is a sense in which the politician and the beneficiary of a political act (say, a recipient of a tax break or a subsidy) engage in a somewhat loosely described "transaction," but affected nonbeneficiaries (typically a taxpayer forced into the given fiscal arrangement) lose out as a result. This "win-win-lose" configuration is denoted as a "triadic exchange," which tends to become more prevalent as the public sector expands.

The reality of extensive economic-political commingling enables Wagner to introduce the reader to his key concept of an "entangled political economy," in which "prudent commercial conduct cannot be determined independently of the desires expressed by political entities" and vice versa. From the vantage point of network theory, entangled political economy can be interpreted as a nonrandom and scale-free assortment of nodes and connections between economic and political actors.

Whether it is an entangled political economy or a political economy governed by a "constitution of liberty," Wagner insists the observed configuration of political economy represents a *spontaneously ordered* process, framed as the by-product of purposeful human action but not necessarily intentional design. Wagner perceives an entangled political economy as a framework that "accommodates

recognition that societies change only through individual action within those societies, and with those actions spreading within the society according to the receptivity of other members of that society to those changes."

The properties and implications of the entangled political economy are profound. Legislatures in entangled political economy can be perceived as "bazaars" through which the "peculiar business" of state takes place: "Parliaments don't produce goods and services for customers, but rather serve as intermediaries between people who are seeking support for their enterprises and people who have the means to support those enterprises. It is in this setting that parliamentary assemblies are peculiar forms of investment bank, in that a good deal of parliamentary intermediation is forced and not voluntary."

Keeping up with this metaphorical depiction of legislative bazaars, the political representatives deciding how fiscal resources are redistributed may be conceptualized as "stallholders" at the bazaar, whose positions are determined by periodic elections for political office. The stallholders' "stock," redistributed by the political stallholders, is originally procured by force from beleaguered taxpayers involved as third parties in unceasing triadic exchanges.

The supply and demand for political favoritism raises the question: How does "peculiar political business" affect the relative risk and rewards associated with undertaking productive activities? In practice, answering this question has taken on some urgency as the public discourse increasingly points to crony capitalism as a determinant of both excessive inequalities and suppressed opportunities within the United States and elsewhere.

Public choice theorists have long depicted processes in which market suppliers and other participants within civil society secure advantages by petitioning governments for fiscal and regulatory favors. Politically obliging rent-seeking petitioners with favorable deals ensures that "property rights are not absolute and invariant, but rather denote social relationships that are subject continually to margins of contestation and potential change."

Wagner suggests the tendency to seek particular benefits and favors through the political realm is magnified by the centripetal tendencies of power within modern Western polities, a concerning trend even in the United States given its much-venerated tradition of robust, competitive federalism.

From the perspective of network theory advanced in *Politics as a Peculiar Business*, policy centralization is depicted as an evolutionary transition from polycentric to increasingly monocentric arrangements of political offices. This transition, in turn, implies fundamental changes to the character of public governance: "Increases in the scale of governance bring oligarchic characteristics, including domination by entrenched interest groups and an electoral process characterized more by the continuing reelection of incumbents than by competition among political parties."

Holders of concentrated power seek to retain their advantages, thus the task to both decentralize and disentangle the economic and the political spheres will continue to be a most formidable one in practical terms.

Undeterred by the magnitude of such a task, Wagner outlines the key principles in the quest to reassert a liberal order, including constitutional limitations upon government and institutions to suppress the outbreak of special interests. In Wagner's words, "the task would seem to require both parchment and guns, that is, both knowledge pertinent to the task and rightly aligned desires and incentives to act consistently with that knowledge."

Dedication (or, in the contemporary case, a rededication) to the principles of limited government, as inscribed on constitutional "parchment," is necessary to frame widespread societal expectations about the functions of public-sector activities. The elucidation of constitutional controls also promotes the accountability of politicians to nonpolitical actors within the context of democratic forms of government.

Although people have the capacity to learn about the abuses and misuses of government, and accordingly guard against these through constitutional agreement, Wagner astutely notes that "while parchment paper is stronger than ordinary writing paper, it is not sufficiently strong to deter rapacious interest groups from using government as an instrument of predation."

In addition to a constitutional declaration of governmental limitations, institutional structures can help ensure politicized self-interests exhaust themselves into an impasse without harming voluntary economic interaction. Those self-interests can be harnessed in the generalized interest of promoting public well-being. This is where the reassertion of polycentric modes of public governance, such as competitive federalism, step into the picture, as well as the need to

revitalize institutions such as cross-checking legislatures, legislative-judicial separability, and so forth.

Wagner also emphasizes the importance of ethical norms that shape whether interacting agents endorse contract or status as the major operating principle within political economy. Recalling the sociological insights of Italian theorist Vilfredo Pareto, Wagner argues that even highly discriminatory political processes can acquire a begrudging tolerance among the populace, insofar as politicians successfully invoke nonlogical sentiments as persuasive rationales for their economically meddlesome actions. In this context, taxing, subsidizing, or regulating private economic interactions may be rationalized by the political class and special interests as "altruistic" acts of caring about redressing certain issues, even if evidence supporting the effectiveness of public interventions is often found wanting.

It is conceivable that, as the domain of collective action continues to gradually take precedence over noncollective action, the influence of beliefs, ideologies, sentiments, and values in political argumentation will keep growing in importance. Wagner cites "the language of welfare economics and economic policy . . . [treating] . . . all political entities as all-encompassing instances of 'we.' . . . Perhaps such ideology repeated again and again comes to be accepted. If so, the stage is surely set for an expansion of the reach of the political within society."

Wagner also suggests diminishing freedom in an entangled political economy is tied to an expedient rejection by larger numbers of people of exercising individual responsibility. As he explains, "participation in cooperative association . . . requires that people pull their own weight. Responsibility for one's conduct is a concomitant of liberty, and it's conceivable that some people will regard the cost of responsibility as too high relative to the value they place on liberty, and so opt for servility." If the propositions raised here are correct, the ongoing challenge for classical liberal advocates will be to also enunciate the case for economic-political disentanglement framed in new and attractive "nonlogical" ways that inspire people to support extended liberties and the responsibilities inextricably associated with them.

Wagner teaches us that the human capacity to truck, barter, and trade remains unquenchable; politicians are not alien interventionists designed to reign over the rest of us; and society is an emergent by-product of ongoing interaction between people situated in market

and political realms. These are important lessons for public policy practitioners and scholars alike, especially if they have been deeply versed in the idea of public finance as a mode of top-down social control.

Replete with original insights, and combined with mainline political economy themes from Adam Smith onward, *Politics as a Peculiar Business* stands as a monumental work for realigning our core understandings about complex human action in multifaceted social realms.

Mikayla Novak
Institute of Public Affairs
Melbourne, Australia

Who Cooked Adam Smith's Dinner? A Story of Women and Economics
Katrine Marçal
New York: Pegasus Books, 2016, 230 pp.

John Maynard Keynes once wrote of Hayek that one of his arguments was "an extraordinary example of how, starting with a mistake, a remorseless logician can end up in bedlam." Katrine Marçal's *Who Cooked Adam Smith's Dinner? A Story of Women and Economics* is a nice example of how an author can actually start with a correct observation, yet also end up in intellectual bedlam. The correct observation at the heart of Marçal's book is that economics has, historically, paid far less attention to the household and family than it has to the market, and that, when it has examined the family, it has frequently done so through a framework that is too narrow to really capture how the institution works. From that truth, she begins her path toward bedlam by laying blame on economics for constructing itself around the hyperrational, self-interested, and very masculine, "economic man," which she thinks has blinded the discipline to women's concerns and other forms of behavior more associated with women. As a result, economics fails at understanding a world containing men and women with a wider range of motivations and personalities, which, in turn, she believes, undermines the case for the market.

Her title gets at the core of her argument: Adam Smith invented modern economics by creating the idea of "economic man"—who is, in Veblen's terms, "a lightning calculator" of his self-interest—and

the idea of the invisible hand, by which greed and self-interest are always good in leading to economic growth. Furthermore, she argues that this vision of the economy, which came from Smith and has been mathematically refined by generations of thinkers, was an attempt to create a kind of Newtonian mechanics of the social world. Armed with this view of humans, which was masculine to its core, economists have been blind to the household and family, as well as ignoring all the ways other than self-interest in which humans interact. At one point, she describes economics as all about "conserving love." Economists believe love is scarce, but self-interest is not, so it makes more sense to run an economy on the latter rather than the former. Thus, Smith never discusses the question of who cooks his dinner (his mother), nor why such issues might be important.

Marçal then argues that once one understands that this is how economists see the world, we can understand the origins of the financial crisis and Great Recession that began in 2008. By focusing on and overestimating human rationality, by giving full license to greed, and by thinking that unhampered markets will always produce good outcomes, economists could not see the crisis coming and have not been able to offer meaningful regulatory and other government policies to prevent another one. This same logic also explains why the United States refuses to adopt "family friendly" policies like subsidized day care and stronger parental leave mandates, which, she believes, would help women.

There are numerous problems with this argument, despite it proceeding from a largely correct observation. The fundamental problem is that Marçal has created a straw economic man out of Adam Smith. At times, she suggests that her depiction of the humans that populate economic models—self-interested, greedy, and utterly unconnected to others and incapable of emotional reaction—is more the product of later economists. However, by continually returning to Smith as the source, she invites the question of whether her characterization of Smith is fair. It is not.

From all the evidence in her book, Marçal seems completely unfamiliar with Smith's book, *The Theory of Moral Sentiments* (TMS). (To be honest, I'm not so sure that she's really read *The Wealth of Nations* either.) In TMS, Smith offers his vision of moral philosophy, and his focus is on the faculty of sympathy. The famous first sentence of the book puts the lie to Marçal's straw economic man of Smith: "How selfish soever man may be supposed, there are evidently some

principles in his nature, which interest him in the fortune of others, and render their happiness necessary to him, though he derives nothing from it except the pleasure of seeing it." A search of the book reveals dozens of discussions of love, which for Smith was not something scarce to be conserved, but a core part of the human personality and key driver of our ethical behavior. Similarly, the book has numerous discussions of the importance of the family. Simply put, anyone remotely familiar with what Adam Smith actually wrote in his two major works would know that his vision of human nature bears no resemblance to the sociopathic one that Marçal attributes to him.

Recent scholarship on Smith explores the connections between the focus on sympathy in TMS with the focus on self-interest in *The Wealth of Nations*. The general conclusion is that Smith recognized that, in the more intimate worlds of direct interpersonal interaction, we are able to exercise sympathy and love more easily, whereas in the more anonymous world of the commercial society, it is harder to do so. That difficulty is not because love is scarce, but because we cannot know exactly what it would take to get others to cooperate with us when we know little or nothing about them. Just before the famous defense of the self-love of the butcher, brewer, or baker in *The Wealth of Nations*, Smith points out that we do not have the time or knowledge sufficient to elicit cooperation through sympathy and altruism from all of those we wish to cooperate with, so we rely on appeal to self-love in people we do not know well. This argument is best understood as a complement, not a substitute, for the emphasis on sympathy and love in TMS, especially because TMS was written first.

Smith's two books cover the range of human interaction, from the intimate worlds of friends and family to the larger world of the commercial society. Smithian man is not a sociopath. He is a man of love, sympathy, and has concern for others who also understand that, while those traits are necessary for moral behavior and an ethical society, they are not sufficient to generate social cooperation among anonymous actors. This whole framework for understanding Smith, which has a good deal of experimental evidence behind it, is much more sophisticated and nuanced than Marçal's straw man.

Marçal might respond that, even if she's wrong about Smith, she's not wrong about modern economics. She's on stronger ground here, but three points are worth making. First, even 20th-century economists understood that the notion of self-interest was not about the

self strictly considered. As Hayek wrote in a discussion of Smith and individualism: "The 'self' for which alone people were supposed to care, did as a matter of course include their family and friends, and it would have made no difference to the argument if it had included anything for which people did in fact care."

Second, even strong defenders of self-interest among modern economists do not think that greed is always good. What Marçal leaves out, and what economists have discussed at least since Smith, is that whether or not self-interest produces good social consequences (is "good") depends upon the institutional framework within which we act. If the law permits me to steal from others, then acting on my self-interest will not produce social good. If the law punishes theft, I will have to satisfy my self-interest by instead creating value for others through production and exchange, which will produce social benefits. Self-interest is not an unambiguous good. Its value is institutionally contingent. This well-understood point appears nowhere in Marçal's discussion of Smith or modern economics.

Finally, the second half of Marçal's argument is that the case for free markets rests upon what she sees as a demonstrably false conception of human behavior. By pointing out that humans are not the self-interested lightning calculator of modern economic models, she believes she has undermined the case for the market, with the Great Recession being Exhibit A. An economics that paid more attention to feminist concerns and whose model of human behavior was less sociopathic would never have created or justified a world in which something like the financial crisis would have taken place.

This is a variation of a long-standing argument from critics of markets. Such critics believe that the case for markets rests upon the rationality of individual actors, so if they can show that people aren't rationally self-interested, the case for markets collapses. But this is a major misreading of modern economics. As noted earlier, whether people's broadly self-interested actions produce good consequences depends on the institutional environment. Vernon Smith won a Nobel Prize for his work on "ecological rationality" that explores many of the issues Marçal overlooks. One would think that a book criticizing Adam Smith and modern economics would show some familiarity with a Nobel Prize winner who has written on these very topics, but, alas, it does not.

But had she just read both of Adam Smith's books, she would have seen that, from the start, the case for the system of natural liberty

hardly rested on a narrow conception of economic man. The case for the market is about the ways in which market institutions make it possible for lazy, ignorant, and self-interested humans to learn what they need to do to create value for themselves and others, and then to provide them with the incentives to do so. Markets don't depend on human rationality—it is markets that help us become more rational.

Marçal's confusion is culminated in her blaming the financial crisis and Great Recession on the economics profession's use of economic man and supposed love of free markets. The housing boom and bust was driven not by unregulated financial markets and "greed is good" capitalists, however, but by the Federal Reserve's attempts to manage market outcomes and political actors who claimed to be altruistically helping a wider range of people obtain houses through a variety of subsidies, mandates, and regulations. In other words, the Great Recession resulted from the same sort of rejection of free markets that characterizes Marçal's critique of modern economics. The Great Recession is what you get when you don't let markets work.

In addition, the very sorts of abstract models that Marçal sees as problematically gendered are what support the case for discretionary monetary policy and regulatory intervention. The tools of modern economics are far more often used to justify stifling markets than defending them. It was the partisans of intervention who thought they could model economies as Newtonian systems, not the defenders of markets. Blaming the Great Recession on a hypermasculinized model of human behavior that led economists and policymakers to unleash greed in a deregulated financial market is wrong both theoretically and historically.

Marçal is right to point out some of the ways in which economics has ignored the contributions of women. However, in addition to the flaws already noted, the book also overlooks the ways in which the very sort of muscular government regulation she wants has made it harder for women to compete effectively in the labor market—for example, occupational licensure, minimum wages, mandated benefits, and the secondary earner bias of the tax code. In her attempt to construct a more female-friendly economics, these very real issues are never discussed.

Marçal's book has been popular precisely because it appeals to the straw man of economics held by people who share her political priors. Unfortunately, her butchering of Adam Smith, her misunderstanding

of the role of self-interest and institutions, and her misreading of the causes of the financial crisis combine to leave little of value to people who really are familiar with modern economics, especially those who have actually read Smith. It is true that economics has not always done a good job in incorporating the work that women do, and it is also true that the discipline can and should do better. Marçal's book, however, does more damage than good to the cause of figuring out a path forward.

Steven Horwitz
St. Lawrence University and Ball State University

The Vanishing American Corporation: Navigating the Hazards of a New Economy
Gerald F. Davis
Oakland: Berrett-Koehler, 2016, 222 pp.

In his Introduction to *The Vanishing American Corporation*, Gerald F. Davis, the Wilbur K. Pierpont Collegiate Professor of Management at the Ross School of Business at the University of Michigan, succinctly expresses the core theme of his book: "Today, the compact between corporations and employees is increasingly under siege by low-cost alternatives that *make the traditional corporation unsustainable* [emphasis added]." Davis spends the first three of the four parts of his book explaining the rationale behind his conclusion, beginning with an evolutionary history of the corporation in America, proceeding to the "why" behind the impending "disappearance" of the American corporation, and concluding with the consequences resulting from the demise of this American legal and economic institution. The final part of the book is his take on a postcorporate America and how future Americans should educationally prepare to navigate this emerging economy.

In Part I, Davis reviews the legal and economic foundations of the modern American *public* corporation (the legal form that Davis focuses on) and its organizational variation among many developed economies. Throughout most of the 20th century, public corporations, such as General Motors, Exxon, and ITT, controlled the majority of economic activity in the United States. Yet public corporations (hereafter "corporations") differ globally by both board make-up—in Germany, for example, half of the supervisory board is elected by

employees—and by the development of stock exchanges—the Netherlands has only 130 publicly traded companies listed, while half the world's economies do not even have operating stock markets. By the early 20th century, the structure of the American corporation resembled what it would look like for the remainder of the century, yet it would grow larger through the continued application of mass production technology and the emergence of post-World War II conglomerates, such as ITT.

Davis notes that "taming the American corporation" was a primary goal of the American Progressive movement in the early 20th century. The federal government's enactment of the Clayton Act of 1914, the Glass-Steagall Act of 1933 (addressing the banking sector), and the National Labor Relations Act of 1935 were considered crucial for restricting the growing power and political influence of the American corporation. Subsequently, the "Treaty of Detroit"—negotiated between General Motors and the United Auto Workers in 1949—was considered an apex point for U.S. labor management relations. That new "social contract" established an employer responsibility to provide health and guaranteed retirement benefits for *all* employees (unlike in other countries, such as Canada and Great Britain, which operated government-run nationalized health care systems). Moreover, President Richard Nixon continued the Progressive vision (begun by Teddy Roosevelt) when he established such federal regulatory agencies as the Environmental Protection Agency and the Occupational Safety and Health Administration in 1970, and the Equal Employment Opportunity Commission in 1972, all preceded by the Tax Reform Act of 1969, the purpose of which was to create a more progressive tax system.

So why does Davis believe the American corporation is disappearing? In Part II, he lays out his case. Importantly, argues Davis, the Reagan presidency allowed for American shareholders to act upon their claim that "bloated, incoherent, and undervalued" conglomerates of the 1960s and 1970s were of diminished financial benefit to them. Federal antitrust authorities loosened restrictions on horizontal mergers (a result of the "Chicago School" predominating in antitrust policy circles) and state laws were largely eliminated that had made hostile takeovers difficult, resulting in hostile takeovers being the rage in the 1980s. By the end of the 1980s, nearly one-third of the Fortune 500 had merged or were acquired. Moreover, from the early 1980s onward, an increasing number of companies were

moving employees from defined pension plans to portable 401(k) plans invested in stock mutual funds, with more than half of all U.S. households invested in the stock market by 2000. Thus, by the early 1990s, as shareholders became increasingly organized and influential on corporate boards, the obligation of the corporation was clearly to create shareholder value. Today, even "blue chip" corporations can be targeted by hedge fund activists.

The 1990s also saw the rise of the "virtual corporation," whereby American corporations outsource segments of their supply chain, usually manufacturing and distribution functions. Nike became the "poster child" for this new business model by focusing its efforts on design and brand management (its "core competencies") while contracting out manufacturing to East Asian suppliers. The advent of the World Wide Web has accelerated the spread of the virtual corporation, and manufacturing outsourcing has spread to other industries, including computer technology, pet food, pharmaceuticals, and government services. However, there are downsides to the virtual corporation. For example, this business model has resulted in the loss of significant numbers of good-paying manufacturing jobs across America. Furthermore, "Nikefication" means that companies have little understanding of their own supply chains, leading to what Davis calls the "responsibility paradox": "Companies are increasingly dispersed and 'virtual,' yet we ask them to be responsible for the actions of their suppliers, employees, and even the countries where they do business."

The 21st century, says Davis, has unveiled the "pop-up" company (i.e., needing only "a web connection and a credit card" to start up), as the economies of scale responsible for the birth of the modern corporation have disappeared in many business sectors. These so-called lightweight entrants can scale up (or down) rapidly by renting rather than buying capacity, and their low costs translate to superior choices for consumers—in 2007, for example, Vizio (with 200 employees) sold as many televisions in the United States as Sony (with 150,000 employees). The decline of the public corporation, with all its regulatory costs, has resulted in the increasing use of a legal hybrid, the limited liability corporation (LLC), by entrepreneurs and established corporations. The LLC is inexpensive to establish, flexible in design, its tax returns are submitted by owners, and it has regulatory advantages because Congress often passes legislation that applies only to corporations listed on the stock exchanges (which LLCs are not).

Davis also notes that the market for initial public offerings (IPOs) hit its high-water mark in 2000, and through 2014, there have been approximately 1,600 IPOs in the United States. Many 21st-century companies simply do not need the capital to build capacity. Moreover, Davis found the number of full-time jobs created by these IPO companies is quite modest, as the median IPO firm grew its employment by 51 jobs—certainly not qualifying them as an "engine" of American employment growth.

In Part III, Davis evaluates the consequences of the collapse of the American public corporation. First up, he discusses the public policy consequences of the disappearing social safety net. As mentioned earlier, corporate employers began shifting employees from defined retirement benefits to "defined contribution" 401(k) pension plans and further streamlined their health benefits. As Davis notes, "the 401k plan has been . . . a miserable failure that is destined to leave many retirees in poverty," making them dependent on Social Security and Medicare. Davis quotes Teresa Ghilarducci, professor of economics at the New School for Social Research: "Seventy-five percent of Americans nearing retirement age in 2010 had less than $30,000 in their retirement accounts. Almost half of middle-class workers, 49 percent, will be poor in retirement, living on a food budget of about $5 a day."

When Davis approaches the topic of income inequality in the United States, he offers a novel observation: it is not the increasing *power* of corporations but their *collapse* that is causing this inequality. He notes that the link between corporate size—that is, employment concentration—and income inequality (using the Gini coefficient) in the United States for the 10 largest companies since 1950 (through 2006) reveals a remarkable correlation of almost −0.9. This correlation means that when American companies grew larger (say during the 1960s), income inequality went down; when companies reduced in size (as many did in the merger wave of the 1980s), income inequality went up. The resulting paradox is that while larger corporations are more unequal internally than smaller ones, those economies with a higher volume of larger corporations tend to have more national income equality.

Furthermore, Davis argues that the flattening of corporate hierarchies in recent years has eliminated the once-stable pathway for career advancement and upward mobility for Americans. As he puts it, "Careers were replaced by jobs, and now jobs are being replaced

by tasks." He concludes that we now are in a "chutes-and-ladders" economy where the "connection between effort and outcomes is often obscure." But there are "silver linings" to corporate decline. Davis sees many entrepreneurial opportunities in service industries utilizing what he calls "platform capitalism" (descriptive phraseology that Davis characterizes as "more accurate" than the popular "sharing economy"), where there are a variety of organizational forms enabled by new technologies (such as the World Wide Web) and based in networks of collaboration. Moreover, in manufacturing, where computer numerical control (CNC) machine tools have continued to drop in price over the last 20 years, less costly CNC technology allows for smaller, low-cost, on-demand production facilities to now be re-shored to the United States.

Yet the public corporation will not vanish completely, says Davis, as they will continue to exist where there are greater risks, economies of scale continue to predominate, and there is a reasonably long time horizon (usually involving infrastructure capital budgeting decisions). With private sources of funding emerging ("alternative platforms for intermediating capital") for new ventures, Davis sees Wall Street's role in the economy receding, perhaps returning to its more traditional, pre-IPO functions.

In Part IV, Davis offers his forecast of two possible scenarios of the American "post-corporate future": one dystopian and the other localized and dynamic. The dystopian scenario is an American economy based on the "uberization" of labor, with tasks and jobs digitally bid on demand, thus eliminating much of the daily interactions of traditional employment. A less onerous alternative scenario has the promise of new technologies favoring distributed manufacturing—small plants with low-cost CNC equipment or 3-D printing can be operated efficiently at the city or neighborhood level. Lastly, Davis recommends that anyone preparing for this postcorporate future would be best prepared by undertaking a strong liberal arts education, as a liberal education creates enduring value beyond specific subject content. It is very difficult to predict which firms or industries will be the winners a decade in the future, so a strong foundation in logical thinking, communication skills, and an understanding of human cultures can go a long way to prepare one for future employment.

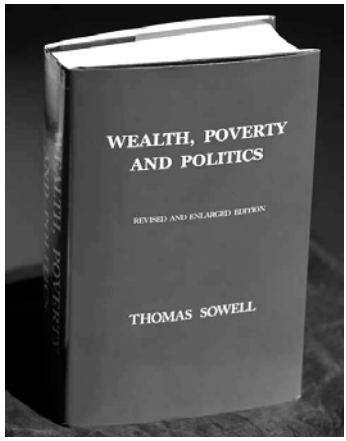
In *The Vanishing American Corporation*, Davis offers a compelling look at the past and present and offers a glimpse of what the American public corporation might look like in the future. The

present status of public corporations in America should be an "eye opener" to the reader. This is where Davis shines. Technological innovation, reduced restrictions on international trade, and the increasingly "heavy hand" of the regulatory state can lead to the emergence of alternative organizational forms. In the case of the American public corporation, the last quarter century has led to "virtual corporations" focused on core competencies and the increasing use of the LLC legal form to reduce public scrutiny. I am an agnostic as to how a business should organize (as management has the advantage of superior contextual insight), but the public policy consequences of the decline of the American public corporation are chilling for American workers.

American society will have to come to grips with the results of what Davis sagaciously identified as the fallout from the "collapse" of the traditional American public corporation: the failure of 401(k) plans to ensure a humane retirement for many retirees; the elimination of stable opportunity and upward mobility offered to new college graduates by the "tall" hierarchies of the traditional public corporation; and a future for American workers that offers increased employment unpredictability and its resulting social manifestations (which Davis does not fully explore, but touches upon in his reference to long-term labor force participation and social security disability trends). With new opportunities come new hazards, and Davis offers sufficient grist for Americans to actively discuss in thinking about their economic future.

Thomas A. Hemphill
University of Michigan–Flint

WEALTH, POVERTY



Wealth, Poverty and Politics is a new approach to understanding age-old issues about economic disparities among nations and within nations. These disparities are examined in the light of history, economics, geography, demography and culture.

Wealth, Poverty and Politics is also a challenge to much that is being said today about income distribution and wealth concentration— a challenge to the underlying assumptions and to the ambiguous words and misleading statistics in which those assumptions are embedded, often even by leading economists. This includes statistics about the much-discussed "top one percent."

This revised and enlarged edition should be especially valuable to those who teach, and who want to confront their students with more than one way of looking at issues that are too important to be settled by whatever the prevailing orthodoxy happens to be.

*****

A true gem in terms of exposing the demagoguery and sheer ignorance of politicians and intellectuals in their claims about wealth and poverty… Dr. Sowell's new book tosses a monkey wrench into most of the things said about income by politicians, intellectuals and assorted hustlers, plus it's a fun read.

(Professor Walter E. Williams, George Mason University)

At a time when many politicians, academics and media commentators are focusing on income inequality, Thomas Sowell's *Wealth, Poverty and Politics: An International Perspective* offers a refreshing and stimulating view.

(Professor John B. Taylor, Stanford University)

Sowell… draws from this well of research to do what he has done so well for so long: question basic assumptions behind public policy and follow the facts where they lead him. (Jason Riley, *Wall Street Journal*)

and POLITICS

It's a scandal that economist Thomas Sowell has not been awarded the Nobel Prize. No one alive has turned out so many insightful, richly researched books. His latest is another triumph of crackling observations that underscore the ignorance of our economists and policymakers. His take on how culture, geography, politics and social factors affect how societies progress— or don't— will rile those addicted to political correctness but leave everyone else wiser.

(Steve Forbes, *Forbes* magazine)

CONTENTS

Had such an approach been available in this reviewer's student days, his understanding of the world would be that much better. (*Library Journal*)

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

8^{00}

THE COMMANDING HEIGHTS OF DODD-FRANK

I fear Dodd-Frank's ultimate purpose is to eventually render effective control of our capital markets to the state; to turn large money-center banks into functional utilities, so that the state can allocate credit within our economy to politically favored classes. In other words, to take over the commanding heights of our economy.

—Rep. Jeb Hensarling

Remarks at the Economic Club of New York, June 7, 2016.

CATO INSTITUTE

www.cato.org

the CATO JOURNAL

VOLUME 37 NUMBER 1 ● PUBLISHED BY THE CATO INSTITUTE ● WINTER 2017