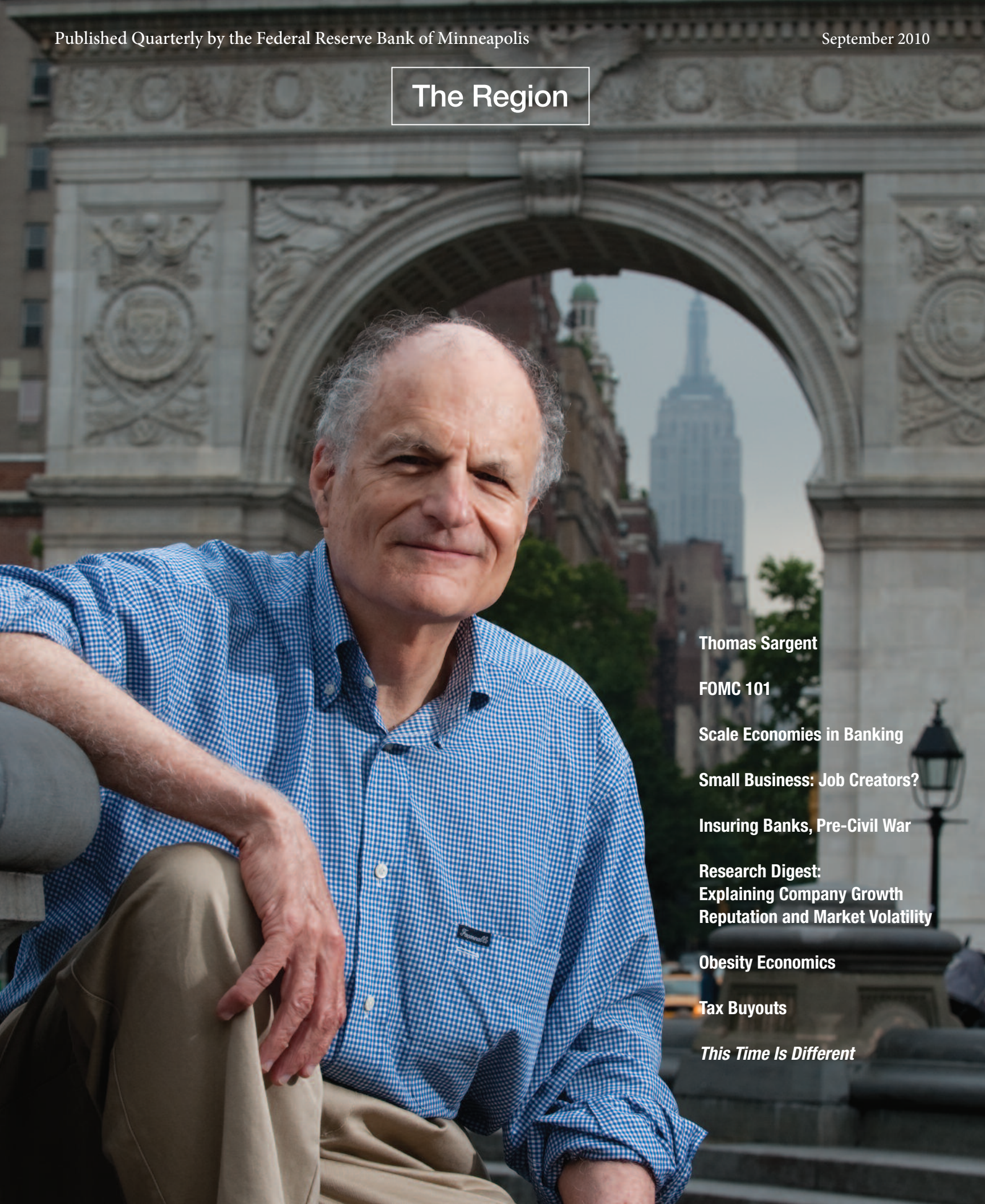
Published Quarterly by the Federal Reserve Bank of Minneapolis September 2010

The Region



Volume 24 Number 3
September 2010
ISSN 1045-3369

The Region

The Region
Federal Reserve Bank of Minneapolis
P.O. Box 291
Minneapolis, MN 55480-0291

E-mail: letters@mpls.frb.org
Web: minneapolisfed.org

Cover photograph by Peter Tenzer

An Introduction to the FOMC

How the Fed's central decision-making body sets monetary policy

Narayana Kocherlakota

President
Federal Reserve Bank of Minneapolis

It has been nearly a year since I became president of the Federal Reserve Bank of Minneapolis, and during that time one of my primary responsibilities has been participating in meetings of the FOMC. I trust that *Region* readers are far more familiar with that acronym than was the government official referred to in a classic 1998 speech by former Fed Governor Laurence Meyer. "What did he believe it stood for?" asked Meyer. The reply: "Fruit of the Month Club."[1]

"Federal Open Market Committee" is the right answer, of course. But while that name may be known to the *Region*'s audience, the actual activities of the FOMC no doubt continue to be something of a mystery. I'd like to take this opportunity to dispel, at least partially, whatever obscurity and confusion might remain.

The FOMC is the Federal Reserve's principal decision-making body with regard to monetary policy, and its name reflects the fact that the Fed influences the nation's interest rates and thereby its economic activity, primarily by buying and selling U.S. government securities through the open market. The term "open market" refers to the securities markets where the FOMC's decisions are implemented through the purchase or sale of U.S. Treasury and federal agency securities in order to influence short-term interest rates;[2] these markets are "open" in the sense that dealers compete with one another on the basis of price alone.

Of course, the Federal Reserve System has other monetary instruments at its disposal, including traditional tools like the discount rate and reserve requirements. The Board of Governors is responsible for those tools. The Fed also has used less conventional innovations, such as the Term Auction Facility that was employed to great effect during the recent financial crisis.[3] Nonetheless, open market operations remain our primary tool for influencing economic activity; therefore, the FOMC has central responsibility for setting the Fed's monetary policy.

Rather than discuss the Fed's open market *procedures*—a rather technical process explained well elsewhere,[4] I'll do my best to describe the FOMC's *composition* and the *deliberations* it goes through at each of its meetings. To make this a bit more concrete, I'll offer examples from the FOMC's most recent meeting held on August 10. (For those interested in a high level of detail, minutes of each meeting are released three weeks after the meeting itself.[5])

I should begin by pointing out that the FOMC was created by Congress 75 years ago, in the Banking Act of 1935. Thus, it did not exist at the 1913 creation of the Federal Reserve, but was born of the recognition that while open market operations should be conducted centrally (by the Domestic Trading Desk of the New York Fed), information-gathering about the nation's economy and decision-making about the future of monetary policy should have a quintessentially American structure.

A federalist Fed

What do I mean by an American structure? Unlike the central banks of other countries, ours is specifically designed to draw upon the diverse insights of small-town businesses, farmers and ranchers, and large manufacturers, among others, to formulate policy. And to achieve that goal, our "central" bank has a structure that is, in fact, highly decentralized. The Federal Reserve Bank of Minneapolis is one of 12 regional Reserve banks that, along with the Board of Governors in Washington, D.C., make up the Federal Reserve System. Our bank represents the ninth of the 12 Federal Reserve districts. The Ninth District is, by area, the second largest. It includes Montana, the Dakotas, Minnesota, northwestern Wisconsin and the Upper Peninsula of Michigan.

Eight times per year, the FOMC meets to set the path of short-term interest rates over the next six to seven weeks. (Other meetings are held as necessary—either in person or by conference call. During 2008, at the peak of the financial crisis, the FOMC met 14 times. In 2010, we've held six meetings to date, with three more scheduled.) All 12 presidents of the various regional Federal Reserve banks—including me—and the seven governors of the Federal Reserve Board contribute to these deliberations. Right now, there are only four governors—three positions are unfilled—but the White House has nominated excellent candidates for these vacancies. However, the committee itself consists only of the governors, the president of the Federal Reserve Bank of New York and a rotating group of four other presidents (currently Cleveland, St. Louis, Kansas City and Boston). I'll be on the committee in 2011.

In this way, the structure of the FOMC mirrors the federalist structure of the U.S. government. Just as people from around the nation deliberate in the U.S. House and Senate, in the FOMC the district bank presidents from different regions of the country provide input into Fed policy deliberations. The input from the presidents relies critically on information from their districts about local economic performance. We obtain this information through the work of our research staffs—but we also obtain it from business leaders in industries and towns, in my case, across the Upper Midwest. The Federal Reserve System is deliberately designed so that the residents of Main Street are able to have a voice in monetary policy.

Go-rounds

So how, exactly, do the FOMC meetings work? The typical meeting features two so-called go-rounds, in which every president and every governor has a chance to speak without interruption. The first is the *economics* go-round. Participants describe their views on current economic conditions and their outlook for future conditions. Bank presidents' remarks will typically include references to their own local economies as well as the national and global situation.

As part of my contribution to the economics go-round at FOMC meetings, I typically discuss my outlook for gross domestic product (GDP), inflation and unemployment. So, at last month's meeting, for example, my input about the national economy in the economics go-round was, in essence: GDP is growing, but more slowly than we would like. Inflation is a little low, but only temporarily. The behavior of unemployment is deeply troubling; I see current and future problems in labor markets that are likely to continue to prove resistant to the tools of monetary policy.

After the economics go-round, the FOMC meeting moves to its second phase, the *policy* go-round. Again, the meeting participants have a chance to speak in turn about what they perceive to be the appropriate policy choices for the committee. We are all committed to achieving the Fed's dual mandate to attain both price stability and maximum employment—objectives set by the Full Employment and Balanced Growth Act of 1978, generally referred to as the Humphrey-Hawkins Act.

The former objective is generally understood as keeping inflation in a tight range around 2 percent. The second part of the mandate is much more of a moving target. Employment is shaped by many determinants beyond the Fed's control: demographics, social custom, taxes and so on. The Fed's job is to keep employment as high as possible, given these other factors.

Interest rates

Right now, to accomplish its dual mandate, the FOMC has to think about two quite distinct policy tools: short-term interest rates and balance sheet management. (I should stress that each of these policy tools is directed at both mandates, not one tool for one mandate and the other for the other.) I'll talk about each in turn.

Setting the federal funds rate—that key short-term interest rate targeted by the FOMC—is, again, the FOMC's central and traditional tool. For over 18 months, the FOMC has set a target of 0 to 1/4 percent. In terms of its future level, the FOMC's statement in August contains the following key sentence:[6]

"The Committee will maintain the target range for the federal funds rate at 0 to 1/4 percent and continues to anticipate that economic conditions, including low rates of resource utilization, subdued inflation trends, and stable inflation expectations, are likely to warrant exceptionally low levels of the federal funds rate for an extended period."

What do we learn from this rather long sentence? The unemployment rate is 9.6 percent. Market and survey measures of expected inflation are also low (also below 2 percent). In its August statement, the FOMC is essentially saying: We're going to conduct open market operations to keep interest rates low in order to prevent unemployment from going any higher, and we feel safe in doing so because there seems to be little threat of inflation.

Asset management

Then there is the issue of the Fed's balance sheet, the management of which has been a central concern to the FOMC in recent years. As a result of its actions to improve the health of credit and funding markets, the Fed's assets and liabilities have grown dramatically since 2008. Currently, the Federal Reserve has $2.3 trillion of assets—over 2.5 times what it owned in September 2008—and changes in these balances may have a real impact on the national economy.

So, at its current meetings, the FOMC typically discusses recent and potential shifts in Fed assets and liabilities, and sets policy accordingly. At our August meeting, for example, the FOMC deliberated about trends in the over $2 trillion of Fed assets currently in Treasuries, debt issued by Fannie Mae and Freddie Mac or mortgage-backed securities issued by Fannie Mae and Freddie Mac. These MBSs are not "toxic" assets in any sense of the word—they are backed by the U.S. government, and so the Federal Reserve faces no credit risk in holding them. But the MBSs do face so-called prepayment risk. If long-term interest rates are low, many people are likely to prepay the mortgages in the MBS. The owners of the MBS—in this case, the Fed—will then get a large coupon payment, and the MBS's principal falls.

That is precisely what has happened in recent months. Long-term interest rates declined surprisingly fast, leading more people to prepay their mortgages. As a result, the Fed's MBS principal balances have fallen. That fluctuation led the FOMC to make another decision at its August meeting, again spelled out in the statement released—as is standard practice—at about 2:15 p.m. on the final day of the meeting:[7]

"To help support the economic recovery in a context of price stability, the Committee will keep constant the Federal Reserve's holdings of securities at their current level by reinvesting principal payments … in longer-term Treasury securities."

What's behind this somewhat arcane statement? With the prepayment of mortgages and resulting decline in Fed MBS principal balances, the Fed's holdings of long-term assets were shrinking. That left a larger share of the economy's long-term risk in the hands of the private sector. The FOMC concluded that this extra risk in private hands could force up risk premiums on long-term bonds and create a drag on the real economy. To achieve its dual mandate of price stability with maximum employment, then, the FOMC decided to arrest the decline in its holdings of long-term assets by reinvesting the principal payments from the MBSs into long-term Treasuries.

The importance of independence

So, I've taken you through a typical FOMC meeting and the monetary policy situation in the United States. My discussion may strike you as rather techy and wonkish—maybe even verging on the nerdy. I'm sure that my colleagues will forgive me for saying that this nerdy quality mirrors the tone of the discussion within the meeting itself. There is no inflated political rhetoric. We are unabashed technocrats, seeking to solve an unabashedly technical problem: How do we manage monetary policy so as to ensure lower unemployment and maintain inflation at an appropriate rate? We certainly disagree with one another on occasion. But our disagreements ultimately stem from different assessments of the complicated economic situation and not from political differences.

I believe that the apolitical nature of the FOMC's work hinges critically on another aspect of central bank structure, and that has to do with the Federal Reserve's relationship with the U.S. Congress. On the one hand, the Federal Reserve is a creation of Congress. It has the power to amend the Fed's responsibilities, as the recent financial reform legislation certainly attests. The Senate approves the presidential appointments to the Board of Governors. Both chambers receive regular reports from the Board of Governors on the conduct of monetary policy, financial supervision and the payments system. In addition, the Federal Reserve undergoes regular audits of its finances and various operations.

On the other hand, Congress has intentionally removed itself from the direct conduct of monetary policy by granting the Federal Reserve the independence to perform this function on its own. In effect, Congress has said that it does not want monetary policy unduly affected by political considerations. This independence not only is a hallmark of this country's central bank, but is also a characteristic of developed economies worldwide.

Speaking on my own behalf, as I have throughout, I believe that the Fed has a responsibility to sustain the trust inherent in that independence by maintaining a high level of transparency and openness. And it can do so best through clear and frequent communication about how it seeks to carry out its designated functions. I hope that this essay contributes to that goal in some small degree. R

Endnotes

[1] Meyer, Laurence. 1998. "Come with Me to the FOMC." April 2. http://www.federalreserve.gov/boarddocs/speeches/1998/199804022.htm.

[2] Specifically, the FOMC sets a target for the federal funds rate, the interest rate at which depository institutions make overnight loans of their balances held as reserves at the Federal Reserve to other depository institutions.

[3] For a description of these programs and review of their effectiveness, see Willardson, Niel, and LuAnne Pederson, "Federal Reserve Liquidity Programs: An Update," June 2010, *The Region,* http://www.minneapolisfed.org/publications_papers/pub_display.cfm?id=4451.

[4] See Open Market Operations, http://newyorkfed.org/markets/openmarket.html, and Davies, Phil, "Right on Target," December 2004, *The Region,* http://www.minneapolisfed.org/publications_papers/pub_display.cfm?id=3310.

[5] See Meeting calendars, statements, and minutes (2005-2011), http://www.federalreserve.gov/monetarypolicy/fomccalendars.htm.

[6] See the Aug. 10, 2010, press release, http://www.federalreserve.gov/newsevents/press/monetary/20100810a.htm.

[7] See the Aug. 10, 2010, press release, http://www.federalreserve.gov/newsevents/press/monetary/20100810a.htm.


THE UNITED STATES OF AMERICA
E13806638L
Alan E. Cober

Scale Economies in Banking

Economists have studied the presence of economies of scale in banks for some time. But until recently, this body of research has remained the province of the cognoscenti. The financial crisis and resulting efforts to reform financial regulation have given the topic increased attention among economists and others. Arguing that some banks have grown too big and that size brings with it substantial costs to society—including government bailouts—many prominent observers have advocated breaking up the largest banks. These breakup proponents contend that the economic literature does not find that "large" means more efficient for banks. In other words, they argue that the research shows that in the financial industry as a whole, significant economies of scale do not exist.

Perspective on the bank economies-of-scale literature, in general—and its use in recent regulatory reform, in particular—would therefore be of great value, so I am very pleased that two experts in this area agreed to present their views in a *Region* "symposium." Loretta Mester of the Federal Reserve Bank of Philadelphia and the Wharton School at the University of Pennsylvania discusses the most recent findings from the literature. Robert DeYoung of the University of Kansas, a permanent visiting scholar at the Federal Reserve Bank of Kansas City, highlights some specific limitations on what economists and policymakers can actually know about economies of scale from the current literature. In my essay, I've sought to provide additional context to the overall debate and for the other two essays.

Mester, DeYoung and I share two common, overriding conclusions: First, there remain important unanswered questions about economies of scale in banking. Research that provides answers to such questions will have a high return. Second, even if research shows the presence of economies of scale for large banks, government could potentially improve outcomes by limiting the activities, and size, of these firms.

—Ron Feldman

Size and Regulatory Reform in Finance: Important but Difficult Questions

Ron J. Feldman
Senior Vice President

Many developed countries have experienced financial crises from 2007 to the present, and their governments have responded by, among other steps, protecting creditors of banks from loss. Creditors of large banks have been among the most prominent recipients of government support. Policymakers argue that protecting large bank creditors limited the reduction in economic output that would have resulted otherwise; losses from large banks would have "spilled over" to the broader economy had bailouts not occurred.

Reforms aimed at reducing the likelihood that creditors will receive future bailouts—that is, addressing the too-big-to-fail (TBTF) problem—naturally look to bank size as a potential culprit. Proposed and enacted reforms include putting size caps on banks, limiting bank ability to engage in specific activities, subjecting bank mergers and acquisitions to additional scrutiny and requiring government to proactively break up select banks. Are such reforms good ideas? I am skeptical that reforms focused on size per se will achieve their stated purpose of addressing TBTF; I have more confidence in reforms that identify and address features that produce spillovers in the first place.[1]

Moreover, even if they could address TBTF, reforms that take aim at bank size directly might be bad policy because their costs could exceed their benefits. The size of banks might be positively related to other benefits—that is, big banks could offer cost advantages that would ultimately benefit society. In particular, some banking production processes might benefit from *economies of scale,* wherein the average total cost declines as the quantity of output increases.

Supporters of size-focused reforms generally dismiss the potential for economies of scale in finance. They point to an economics literature that has found scale economies only at firms much smaller than those at the epicenter of the financial crisis. I am sympathetic to this tactic. Indeed, I have used it myself in the past![2] But more recently, I have become dubious of this response, for three reasons.

First, some of the recent econometric work on economies of scale for banking finds such benefits at *all* sizes of banks. Loretta Mester nicely summarizes this extensive research in her *Region* essay. From her review, it's clear that blanket assertions that the "literature" supports one position or another are hard to justify.

Second, and more importantly, we may simply not yet know very much about the presence of scale economies for today's unprecedentedly large banks. Robert DeYoung makes this point in his essay. He argues that the unique nature of today's large banks makes it difficult to apply statistical techniques to historical data to divine the extent of scale economies.

And the limits of our knowledge may go still deeper.

In the first place, it is not entirely clear why the financial sector grew as large as it did in recent years.[3] Banks contribute to economic output through intermediation—that is, by taking in cash from savers and using it to finance projects of households and firms. Banks have performed this economically useful function in many countries, for hundreds of years. Such widespread persistence suggests that banks are particularly adept intermediaries, relative to alternatives.

But value-added intermediation does not justify an infinitely large banking sector. There are reasons to think the sector can be too big in the sense that too many of society's resources are allocated to it.[4] Perceptions by creditors of banks that the government will protect them can lead the sector to grow inefficiently large as TBTF guarantees attract excessive funding to banks. These creditors understand that their bank investments are implicitly subsidized by the assurance of government bailouts should the bank begin to fail.

The market share of banks relative to alternatives like capital markets also varies a great deal over time and place, suggesting that the advantage of banks is not absolute. But this does not mean that alternative markets or institutions could provide intermediation without the potential TBTF downside of banks: The ability of markets and nonbank financial firms to generate potential systemic risk has been clearly demonstrated by the recent financial crisis.

In sum, we do not know if society should continue to rely on banks as much as it does given the potential cost and the alternatives available. This question deserves deep consideration. Calling for more research is a cliché. In this case, the cliché is

apt. Research to better understand the optimal size of the banking sector *could* have high returns.

I emphasize "could" because—and this is a second point about the limits of our knowledge on scale economies—analysts face real challenges in measuring the "output" of banks. Economies of scale relate production size to cost. But what exactly do banks produce? Loans? Deposits? "Liquidity"? Economies-of-scale analysis requires cross-firm comparison. Making sure the comparison is apples to apples is a tremendous challenge. All deposits are not made alike, let alone loans and other products banks offer. Economists working in this area know these and other methodological hurdles well and seek to address them, but the barriers are inherently steep.

Third, the debate about TBTF and scale economies presents the two in contradiction, when in fact they may complement one another. Some activities of a bank—for instance, bank production that relies heavily on automation—may both benefit from scale economies *and* enhance that bank's TBTF status. Banks have automated some types of lending, such as certain credit card and mortgage lending, to a significant degree. Processing of payments, trust and custody services, and provision of Treasury services to firms also depend heavily on automated systems, as would certain types of asset management.

These bank services and products require large investments in automated systems. Once the bank incurs the fixed costs of the systems, it can drive down its total average costs by increasing the volume of goods and services produced. Such automation-dependent products and services can generate a material portion of the revenue banks earn. A superficial guesstimate puts the annual revenue from economies-of-scale services at around 30 percent of the total for one of the largest bank holding companies in the United States.[5]

Many of these automation-based services also enhance TBTF status. Payments processing offers an obvious example. If another bank could not quickly take over or substitute for an important, failing bank provider of payments, important capital markets may not function effectively and some commercial firm payments—perhaps even payroll—would not go through. Even the threat of a payments collapse would lead policymakers to seriously consider all available means to keep the payment train running. Greater scale activity, therefore, could come with higher TBTF cost. The presence of economies of scale, from this perspective, suggests that policymakers sharpen their focus on fixing TBTF. More research on the relationship between larger scale and a more severe TBTF problem therefore seems necessary.

Some bottom lines

Smart people seeking to reduce TBTF have justified policies that would make large banks smaller in part on the basis of published research that does not find significant economies of scale in the financial industry. There are (at least) two reasons that conclusion may not hold. It may not reflect the current state of the literature and, more importantly, it may overstate what we actually do know about such scale economies. Indeed, it may be that banks become more TBTF precisely because they are taking advantage of significant scale economies. More generally, policymakers should focus on addressing the potential for spillovers from failing financial institutions even if scale economies exist. R

Endnotes

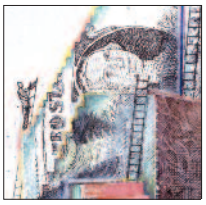
[1] Feldman, Ron J. 2010. "Forcing Financial Institution Change Through Credible Recovery/Resolution Plans." Economic Policy Paper 10-2, Federal Reserve Bank of Minneapolis; Stern, Gary H., and Ron J. Feldman. 2009. "Addressing TBTF by Shrinking Financial Institutions." *Region* 23 (June), Federal Reserve Bank of Minneapolis.

[2] Stern, Gary H., and Ron J. Feldman. 2004. *Too Big To Fail*. Washington, D.C.: Brookings Institution, p. 66.

[3] Philippon, Thomas. 2008. "The Evolution of the U.S. Financial Industry from 1860 to 2007." http://pages.stern.nyu.edu/~tphilipp/papers/finsize.pdf.

[4] There are additional reasons that some observers may view the banking sector as being too big. Some researchers have argued that consumers make systematic errors in their purchases of financial products, including some offered by banks. These errors could lead consumers to consume too many financial products or perhaps the "wrong" financial products. Other observers consider certain activities—such as the trading of financial assets, which sometimes is conducted within a banking organization—as inherently "wasteful." For an example of the first point, see the discussion in John Y. Campbell et al. in "The Regulation of Consumer Financial Products," Social Science Research Network, July 27, 2010, online at papers.ssrn. com. For an example of the second point, see the discussion of a financial transaction tax in "A Fair and Substantial Contribution by the Financial Sector," June 2010, online at imf.org.

[5] To make this extremely rough estimate, we review the fourth quarter 2009 earnings release financial supplement of JPMorgan Chase & Co. In the spirit of Lawrence Radecki (FRBNY *Economic Policy Review,* July 1999), we make the estimate by identifying certain business lines as benefiting from scale and then tallying financial data for these business lines. In particular, we assume that mortgage and credit card lending benefits from scale as do asset management and principal transactions (which include trading activities, among others). The net revenue in 2009 from these operations is $33 billion out of a total of $100 billion. We provide this crude estimate primarily to encourage interested parties to more seriously review bank-specific data to determine the potential importance of scale.



Scale Economies in Banking and Financial Regulatory Reform[1]

Loretta J. Mester
Federal Reserve Bank of Philadelphia
and The Wharton School, University of Pennsylvania

Global financial markets will be shaped for years to come by the regulatory reforms being implemented in response to the recent financial crisis. In my view, two key principles should guide reform efforts. First, reforms should take into account the incentives they create and their longer-run consequences. Second, reforms should harness market forces, not work against them.

U.S. policymakers have sought to foster stability by lowering the probability of a crisis and by reducing costs imposed on the rest of the economy when a shock hits the financial system. An important part of their deliberations has concerned financial firms deemed too big to fail or too interconnected to fail. I believe that, ironically, the United States will have a more stable financial system if failing firms are permitted to fail instead of being rescued.

Policymakers therefore need a way to allow a financial firm—of any size—to fail without precipitating a crisis. For this, a realistic "resolution" mechanism—a means of restructuring or dissolving a firm's assets and liabilities—must be created. A credible mechanism must impose losses on creditors as well as shareholders and do it in a consistent manner so that stakeholders expect this imposition and have incentive to take adequate precautions against failure. The mechanism should be transparent and rule-based, giving regulators less discretion, not more.[2]

A related issue is how to deal with large or interconnected financial firms *before* they get into financial trouble. There has been a striking amount of consolidation in the banking industry in the United States and abroad over the past 30 years, and it has led to some very large banks. In the United States, the number of commercial banks has fallen from about 14,000 in 1980 to fewer than 7,000 today.[3] Even as new banks have entered the industry, there have been over 12,000 bank mergers since 1980, and today, each of the three largest bank holding companies (BHCs)—Bank of America, JPMorgan Chase and Citigroup—has over $2 trillion in assets. Size is not the only indicator of systemic importance: Some institutions are small but important because of interconnections with other financial firms; others are organizationally very complex.[4]

Some argue that the best way to handle banks that are too big to fail is to break them up.[5] To evaluate such a solution, it is important to know why banks have gotten so large. Research suggests that some institutions have gotten large, not to game the system, but for reasons of efficiency. The systemic risks posed by large, complex institutions might still outweigh the efficiencies gained by scale, but without estimating these efficiencies, it is impossible to compare costs against benefits. Moreover, the effectiveness of size limits depends on knowing the market pressures on banks that encourage growth. The literature on scale economies in banking, including my own studies, suggests that imposing a strict size limit would have unintended consequences and work against market forces—contrary to both of my guiding principles for regulatory reform.

To my mind, a better solution than legislative limits on bank size is to develop a credible resolution mechanism coupled with other reforms, including revised capital requirements that involve contingent capital and capital charges based on the firm's contribution to systemic risk, increased disclosures from financial firms, consolidated supervision of large nonbank financial firms, and systemic-risk-focused supervision.

Insights from the literature on scale economies

What has motivated the consolidation of the banking industry?[6] A growing body of research supports the view that there are significant scale economies in banking. Scale economies are usually measured

with respect to costs and refer to how scale of production (size) is related to costs. A firm is said to be operating with constant returns to scale if, for a given mix of products, a small proportionate increase in all outputs would increase costs by the same proportion. A single-product firm operating with scale economies can lower average cost of production by increasing its scale.

Some cite older research that used data from the 1980s and which did not find scale economies in banking.[7] The consensus of these earlier studies was that only small banks had the potential for significant scale efficiency gains and the gains were usually small, on the order of 5 percent of costs or less. But more recent studies, using data from the 1990s and 2000s and models of bank production that incorporate risk management aspects of banking, find significant scale economies at even the largest banks in the sample.

Part of the difference appears to reflect improvements in methods used for measuring scale economies,[8] but it also likely reflects real changes in banking technology, such as computing and telecommunications, and environmental factors, such as a relaxation of governmental restrictions on geographic and product expansion, that have led to a larger efficient scale. The global nature of banking consolidation and increase in scale suggests that U.S. deregulation has not been the only driver. The finding of significant scale economies at banks that are large, but not considered too big to fail, suggests that policy toward the largest institutions is not the only factor.

By their nature, the empirical studies on scale economies derive estimates based on a sample. Constructing samples to include banks that use similar production techniques is important for deriving sound estimates. Newer statistical techniques can overcome some of the drawbacks of earlier studies by fitting the data at the more extreme parts of the sample and not just the sample's average bank. However, only a few existing studies use the most recent data, and bank size has increased significantly over the past 10 years. So, further work needs to be done. Also, the typical estimation techniques do not address whether any particular bank is operating efficiently; other techniques, such as case studies, are more applicable for this type of question. Still, even with these caveats, the studies of scale economies are persuasive that the efficient scale of commercial banking has risen over the past 20 years.

Results of some of the studies

Berger and Mester (1997) estimated the efficiency of almost 6,000 U.S. commercial banks in continuous existence, with complete and accurate data, from 1990 to 1995, and found that about 20 percent of banking costs were lost due to scale inefficiencies, similar to estimates of the loss due to so-called X-inefficiencies (or waste). In every bank size class from less than $50 million in assets to well over $10 billion, we found scale economies for more than 90 percent of firms in the size class. In each class, the typical bank would have to be two to three times larger to maximize scale efficiency for its product mix and input prices.[9] We also found that a simple measure, costs per dollar of gross total assets, displayed scale economies up to $25 billion in assets, but we concluded that "serious estimates of scale economies for U.S. banks over $25 billion will likely have to wait for the consolidation of the industry to create enough of these large banks to yield reasonable estimates." That time has come.

At its heart, banking is about handling risk, and the amount of risk to take on is a management choice. The standard analysis used in earlier studies might not have detected scale economies that actually exist because standard analysis does not account for the risk or capital structure that a bank chooses. A series of papers incorporate managerial preferences over the risk-return trade-off into models of bank production.[10] These studies find that risk management and revenue effects are, indeed, correlated with bank size.

There are two opposing effects on the costs of risk management as banks grow in size. Larger scale may mean better diversification, which could reduce liquidity risk and credit risk. So, there is a diversification effect: Larger scale can lead to reduced marginal cost of risk-taking and reduced marginal cost of risk management, all else equal.

But all else is not necessarily equal because risk-taking is endogenous—a management choice. If banks respond to the lower cost of risk management by taking on more risk in return for greater profits, then we would see another effect of increased scale of operations—a risk-taking effect, which can raise costs, all else equal, if banks have to spend more to manage increased risk or more time dealing with nonperforming assets. Therefore, unless risk is incorporated into the analysis, the increase in costs due to increased risk-taking may

mask scale economies due to diversification.

Hughes, Mester and Moon (2001) found constant returns to scale in a sample of large BHCs using data from 1994 when we used the standard cost-function model from the earlier literature. However, using our more general model incorporating bank managers' preferences about risk and capital structure, we found that BHCs of all sizes were operating with significant returns to scale.[11] We also found that large BHCs were operating with less capital than would have minimized their costs and that small banks were operating at more than the cost-minimizing level of capital. And we found evidence of both a diversification effect and a risk-taking effect. Better diversification is associated with larger-scale economies, and increased risk-taking is associated with smaller-scale economies.[12] So the results support the conclusion that scale economies exist, but the usual method cannot find them because it ignores the fact that banks choose their level of risk and their capital structure. Larger scale means lower cost per unit of risk—a scale economy—but it also means banks have the capacity to take on more risk.

Studies that use more recent data are scarce, but those that do exist find significant scale economies in U.S. banking. Using a large data set covering all U.S. commercial banks from 1984 to 2006, Wheelock and Wilson (2009) find that banks had increasing returns to scale throughout the entire distribution of banks—even in 2006, when the largest banks had nearly $1 trillion in assets. They conclude that "industry consolidation has been driven, at least in part, by scale economies" and that this would imply some cost to limiting bank size. Feng and Serletis (2010), using data from 2000 to 2005 on 293 U.S. banks with over $1 billion in assets, also find scale economies at the largest banks.

Note that none of the research suggests that regulators should stop considering market power when deciding whether to approve a merger. Indeed, the results are based on banks operating under current regulations and Justice Department guidelines. Nor does the literature suggest that all consolidation and growth is beneficial for society. Too-big-to-fail considerations may be a source of some gains—although not the entire source, since scale economies have been found at banks smaller than those most consider to be too big to fail. Also, other research indicates that managerial entrenchment—that is, the ability of managers to resist market discipline—can lead to inefficient consolidation strategies.[13]

Implications for financial reform

Significant scale economies in banking suggest that economic forces have been an important driver of banks' increasing size. This does not mean that the benefits necessarily outweigh the potential costs that larger size may impose on the financial system and broader economy if size is accompanied by higher risk of systemic problems. But if policymakers do conclude that the costs of size outweigh the benefits, the existence of scale economies suggests that a strict size limit on banks is not likely to be an effective solution. Such limits work against market forces and do not align incentives. Given the potential benefits of size, strict limits would create incentives for firms to avoid these restrictions, and could thereby push risk-taking outside of the regulated financial sector, without necessarily reducing systemic risk.

A better tack would be to increase the costs of becoming too complex or too large commensurate with the risks that these types of institutions impose, for example, imposing a capital charge for contribution to systemic risk, while at the same time trying to close the gaps in supervision. Better understanding of the incentives that financial firms have to avoid supervision and regulation and a focus on macro-prudential supervision of the financial system as a whole will be beneficial in helping to foster financial stability. R

Endnotes

[1] The views expressed here are those of the author and do not necessarily represent those of the Federal Reserve Bank of Philadelphia or the Federal Reserve System. An expanded version of this article can be found at http://www.philadelphiafed.org/research-and-data/economists/mester/.

[2] This is not just a theory. Empirical research by Barth, Caprio and Levine (2006) supports this view. They study banking regulatory structures in more than 150 countries and find that transparency and public accountability lead to better banking sector performance than reliance on supervisory discretion.

[3] See the Federal Deposit Insurance Corp.'s *Historical Statistics on Banking,* at http://www2.fdic.gov/hsob/hsobRpt.asp.

[4] Rajan (2009) discusses factors other than size that are related to systemic importance.

[5] See, for example, Johnson and Kwak (2010).

[6] Mester (2008) provides an overview of methods of measuring productive efficiency in banking and a review of the literature.

[7] See, for example, Greenspan (2010), p. 32: "For years the Federal Reserve had been concerned about the ever larger size of our financial institutions. Federal Reserve research had been unable to find economies of scale in banking beyond a modest-sized institution."

[8] These improvements include using more flexible functional forms to capture the relationship between costs, input prices and output levels; taking into account the bank's risk and financial capital structure in empirical models; and incorporating banks' off-balance-sheet activities.

[9] That both small and large banks operate below efficient scale is not a contradiction; each bank's level of scale economies is measured based on its own product mix and input prices. Small and large banks choose different product mixes, each suitable to its own scale of operations (see Berger and Mester, 1997). We grouped banks with assets over $10 billion into a single class because there were too few banks to form credible size classes within this largest category.

[10] See Hughes, Mester and Moon (2001); Hughes, Lang, Mester and Moon (1996, 1999); and Hughes, Lang, Mester, Moon and Pagano (2003). Also, see the summaries in Mester (2008) and Hughes and Mester (2010).

[11] Hughes and I are currently working on a study using data from 2007 and 2008.

[12] Diversification referred to the degree of macroeconomic diversification in a BHC's geographic scope of operations. It was measured by the correlation in unemployment rates over states in which a BHC operates.

[13] See Hughes, Lang, Mester, Moon and Pagano (2003).

References

Barth, James R., Gerard Caprio, Jr., and Ross Levine. 2006. *Rethinking Bank Regulation: Till Angels Govern*. New York and Cambridge: Cambridge University Press.

Berger, Allen N., and Loretta J. Mester. 1997. "Inside the Black Box: What Explains Differences in the Efficiencies of Financial Institutions?" *Journal of Banking and Finance* 21(July), pp. 895–947.

Feng, Guohua, and Apostolos Serletis. 2010. "Efficiency, Technical Change, and Returns to Scale in Large U.S. Banks: Panel Data Evidence from an Output Distance Function Satisfying Theoretical Regularity." *Journal of Banking and Finance* 34(1), pp. 127–38.

Greenspan, Alan. 2010. "The Crisis." Manuscript.

Hughes, Joseph P., William Lang, Loretta J. Mester and Choon-Geol Moon. 1999. "The Dollars and Sense of Bank Consolidation." *Journal of Banking and Finance* 23(2/4), pp. 291–324.

Hughes, Joseph P., Willim Lang, Loretta J. Mester and Choon-Geol Moon. 1996. "Efficient Banking Under Interstate Branching." *Journal of Money, Credit, and Banking* 28(4), pp. 1045–71.

Hughes, Joseph P., William Lang, Loretta J. Mester, Choon-Geol Moon and Michael Pagano. 2003. "Do Bankers Sacrifice Value to Build Empires? Managerial Incentives, Industry Consolidation, and Financial Performance." *Journal of Banking and Finance* 27(3), pp. 417–47.

Hughes, Joseph P., and Loretta J. Mester. 2010. "Efficiency in Banking: Theory and Evidence," in *Oxford Handbook of Banking,* Allen N. Berger, Philip Molyneux, and John O. S. Wilson, eds. Oxford: Oxford University Press, pp. 463–85.

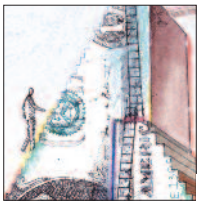
Hughes, Joseph P., Loretta J. Mester and Choon-Geol Moon. 2001. "Are Scale Economies in Banking Elusive or Illusive? Evidence Obtained by Incorporating Capital Structure and Risk-Taking into Models of Bank Production." *Journal of Banking and Finance* 25(12), pp. 2169–2208.

Johnson, Simon, and James Kwak. 2010. 13 Bankers: *The Wall Street Takeover and the Next Financial Meltdown*. New York: Pantheon Books.

Mester, Loretta J. 2008. "Optimal Industrial Structure in Banking," in *Handbook of Financial Intermediation,* Arnoud Boot and Anjan Thakor, eds. Amsterdam: North-Holland, pp. 133–62.

Rajan, Raghuram. 2009. "Too Systemic to Fail: Consequences, Causes, and Potential Remedies." Testimony before the Senate Banking Committee, U.S. Senate, May 6.

Wheelock, David C., and Paul W. Wilson. 2009. "Are U.S. Banks too Large?" Working Paper 2009-054B, Federal Reserve Bank of St. Louis. Revised December 2009.



Scale Economies Are a Distraction

The real issue for policy is credible resolution of failing financial firms, not bank size

Robert DeYoung
Federal Reserve Bank of Kansas City
and University of Kansas

A small cadre of banking economists (including, for a time, me) has studied banking companies for nearly half a century in an effort to answer the following question: Can banks become more efficient by growing larger? Or, in the technical vernacular, do banks exhibit *scale economies*? This question has garnered fresh attention today as policymakers consider steps to regulate bank size in light of too-big-to-fail concerns.

Possible scale economies in the banking industry were also a crucial question for bank regulatory policy during the 1980s and 1990s. Existing regulations kept banks small by prohibiting their expansion across state lines; bankers argued that these rules made the U.S. banking system inefficient. Removing these constraints, they said, would enable them to expand their geographic footprints and capture scale economies. And because banking services are sold in competitive markets, much of the resulting cost savings would be passed along to customers and not simply accrue to bank shareholders.

The question of scale economies was important for banks of all sizes. If two small banks from neighboring states merged, would running the resulting medium-sized bank be cheaper than running the two small banks separately? What if two medium-sized banks merged to create a regional bank? Or if two regional banks merged to create a bank with national presence?

According to the earliest statistical studies, scale economies "ran out" once a bank had accumulated assets of $100 million or $200 million—that is, only small banks could hope to capture scale economies by growing larger. But as my research colleagues developed new and better analytical tools, their conclusions evolved. Subsequent studies found available scale economies up to $500 million in assets … then $1 billion … then $10 billion to $25 billion—that is, all but a handful of U.S. banks at the time had access to scale economies. By the mid-1990s, some of the more innovative studies were reporting that, under certain circumstances, even the largest banks had access to scale economies.[1]

In retrospect, those scale economy studies were the right tool for the job. They provided objective evidence on an argument being made by the (perhaps less than objective) financial services industry. In a significant way, those studies helped pave the way for deregulation and the mix of local, regional and national banks in existence today.

Scale economies redux

The last of the major restrictions on banking geography were removed in 1997 when the Riegle-Neal Act was implemented. In the wave of industry consolidation that ensued, banks of all sizes grew larger by acquiring banks in other states.

At the upper end, the merger wave created banking companies far larger than the banks examined in the scale economy studies of the 1980s and 1990s. For example, today the three largest U.S. banking firms (Bank of America, JPMorgan Chase and CitiGroup) all exceed $2 trillion in assets, while the three next largest (Wells Fargo, Goldman Sachs and Morgan Stanley) all have assets in excess of $800 billion, well above the range covered by academic researchers.

In 2008 and 2009, some of these banking giants suffered huge financial losses that, by virtue of their size alone, threatened the stability of financial markets and the macroeconomy. Government policymakers judged that the risks of allowing those firms to fail were too great; famously, financially troubled banking firms received hundreds of billions of dollars in capital injections and other forms of taxpayer-backed bailouts.

Preventing such an episode from happening again was the focus of long congressional debates this year over legislation to reregulate financial institutions. However, the new law that emerged

leaves important questions related to bank size unanswered: Should the public continue to live with these large banks and the risks they impose? Should regulators break up these firms? Or should policy give these firms incentives to downsize, such as imposing size-based taxes or higher capital requirements?

Clearly, understanding the existence and/or limits of bank scale economies is once again important for forming public policy. But the nature of this inquiry is different from the deregulatory questions of the 1980s and 1990s. First, policymakers and researchers are now interested only in scale economies at the very largest banks, not at banks of all sizes. Second, policymakers now need to know whether any resulting efficiencies are substantial enough to justify living with the social costs and macroeconomic risks posed by these newly enormous firms.

Despite the hard and often ingenious work of my colleagues in the bank scale economy field, I am not optimistic that this line of research will generate the answers needed this time around. Why not? The standard approaches to measuring scale economies are the least accurate for precisely those firms most relevant to the question at hand: the very largest banking companies.

The wrong tool for the job

It is well-known that the statistical techniques employed to measure scale economies in any industry deliver the most accurate estimates for "average" companies in that industry; for firms that are substantially smaller or larger than average, estimates grow increasingly less precise. This characteristic is especially problematic for the banking industry, due to the drastically skewed size distribution of its firms. As of March 2010, the three largest banking companies (mentioned above) each had assets of over $2 trillion, *10 times* larger than the 13th-largest banking company, Bank of New York Mellon, with assets of $220 billion. They were *100 times large*r than the 43rd-largest bank, BOK Financial of Tulsa, Okla., with assets of $23 billion. Because of these dramatic size differences, statistical estimates of scale economies among large banks can be quite sensitive to the good or bad financial fortunes of just one or two of these largest banks.

A second problem arises because the largest banks operate quite differently than small and medium-sized banks; that is, they differ in kind, not just size. But because most of the available data come from the thousands of small and medium banks, bank scale economy models are based on the business processes most often used by these banks. This segment of the industry relies predominantly on traditional banking approaches: holding illiquid loans, issuing liquid deposits to finance those loans and earning profits chiefly from the resulting interest margin. But the very largest banking companies produce financial services quite differently. They rely less on deposits and more on short-term market financing, they sell many of their loans rather than hold them, and they earn a substantial portion of their profits from customer fees rather than interest margins. Using models built around smaller bank production processes to describe the relative efficiency of large banking companies can be misleading.

These methodological deficiencies did not prevent scale economy studies from usefully informing the deregulation debate of the 1980s and 1990s. Geographic deregulation was relevant for banks of all sizes and, at that time, bank production processes were still pretty similar for large and small banks. But these issues may be debilitating in today's debate over reregulating the largest banking companies—while scale economies *might* exist for these banking giants, we cannot be sure because measuring these phenomena stretches our analytic tools to, and perhaps beyond, their limits.

What about market forces?

Perhaps there is a simpler way. Rather than estimating complex models of bank scale economies, could we simply depend on the market to reveal the best size for banks?

The argument goes like this: The fact that banks have grown increasingly large over time is *prima facie* evidence that scale economies exist for even the largest banks. If this were not the case, managers of large banks would be operating inefficiently large firms, and their ill-served shareholders would attribute lower profits to *dis*economies of scale and sell their shares. Investors would purchase, pull apart and reallocate the assets of these firms.[2] Thus, market discipline would ensure that banks would exhibit the most profitable range of sizes and other attributes.

While I generally embrace this line of reasoning, the argument fails for the very largest banking com-

panies in the United States today. Even if these banks are too large to operate efficiently, shareholders are unlikely to recognize or act on this, because the performance-detracting effects of scale diseconomies are masked by the performance-enhancing effects of the too-big-too-fail subsidies enjoyed by these banks. Given the government bailouts of 2008 and 2009, there is no longer any doubt that the largest U.S. financial companies are considered too big to fail. Because these firms can perform poorly and still remain in business, shareholders and creditors benefit from upside success without suffering the full downside losses, which gives the largest banking companies a cost-of-capital advantage over their smaller rivals. In other words, there may be the appearance of scale economies for these firms where none really exists.

Focus on resolution policy, not bank size

If we cannot confidently measure scale economies at the very largest banking companies—and indeed, although researchers have attempted methodological "fixes" of the deficiencies I've mentioned above, I am not sure that we can—then are we forced to make uninformed regulatory policies for these firms? Must we make decisions about whether to break up, downsize or somehow limit the growth of these institutions without reasonable certainty as to the consequences of such actions for the future efficiency of the banking sector?

My sense is that the question of scale economies in banking, while of real interest, is something of a distraction to the primary issue. The chief concern should be not how big banks must be to achieve optimal efficiency, but rather, how policymakers can establish a credible strategy for resolving banks when they fail—regardless of their size, complexity and inter-connectedness. The public needs policies and policymakers that impose harsh discipline on the managers, shareholders and junior debt holders of large failed banks—while simultaneously using bridge banks, other available resolution techniques and expanded resolution authority to preserve the liquidity of borrowers, depositors and other counterparties of these banks.

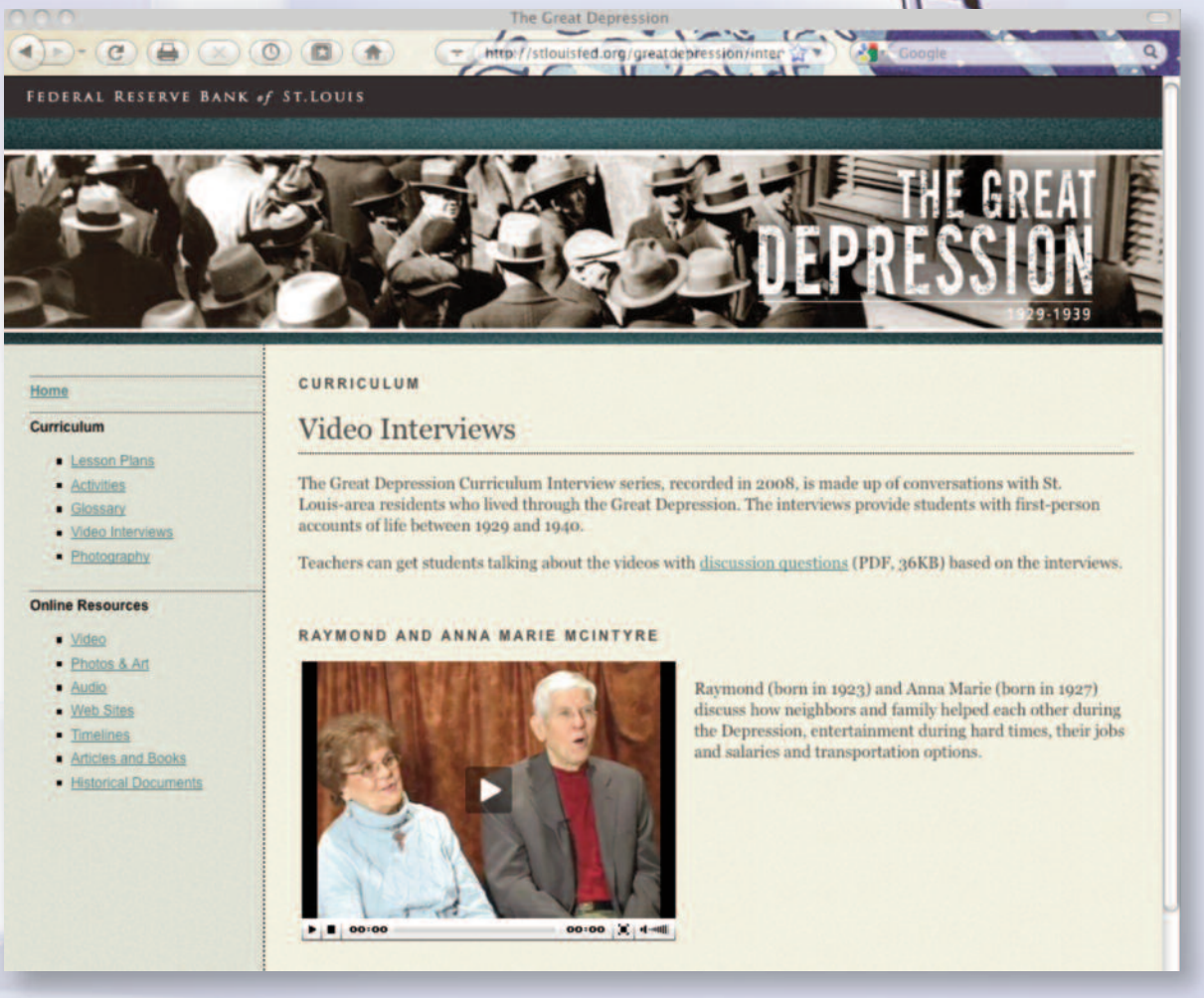
Of course, this is a tall order. But the current inability to do this is the root cause of the too-big-to-fail problem often attributed to bank size. And by addressing this root cause—rather than placing regulatory limits on bank assets or some other measure of size, an ad hoc policy that will surely result in unintended consequences—we will generate a number of benefits. Chief among them: The primary justification for too-big-to-fail subsidies would disappear. Large banks might continue to pose a problem for competitive efficiency (a concern of antitrust policy), but no longer for macroeconomic stability. And we could then rely on the marketplace—no longer handicapped by poorly designed policy—to reveal the optimal size for banks. R

Endnotes

[1] An article by Allen Berger, Rebecca Demsetz and Philip Strahan in the February 1999 *Journal of Banking and Finance* discusses this literature in more detail (see pages 157-60). While the advancing research has found increasing access to scale economies for banks, no similar consensus has emerged regarding the dollar magnitudes of these savings or whether managers running large banks are able to fully exploit the potential for savings.

[2] Because changes in ownership of banks require regulatory approvals, this "market for corporate control" mechanism would likely work more slowly in the banking industry than in other industries.

Virtual Fed



The Great Cache: 1929

The Great Depression is all the rage nowadays. The "great recession" of 2007–2009 was rife with comparisons to the 1930s, while the academic debate on what exactly caused the Great Depression continues, with new theories still coming forth (see greatdepressionsbook.com for more on the latter). These comparisons usually take the form of charts of employment, output and the like, while the academic debate mostly revolves around data and economic theories about what makes an economy grow or decline.

For a more personal take on the Great Depression, you can't do much better than the St. Louis Fed's Great Depression Web site. As part of its educational curriculum on the period intended for economics and history teachers, the site naturally includes a good deal on the data and economics. But it also holds a treasure trove of links to audio and video files—news reels, music, photo archives and much more. In perhaps the most personal look of all, the site also features a collection of recently added video interviews with St. Louis-area residents who lived through the depression.

For more, visit stlouisfed.org/greatdepression. Video interviews are at stlouisfed.org/greatdepression/interviews.html.

—Joe Mahon

Sizing Up Job Creation

Are small businesses truly the engine of job growth? It depends on how you look at it

Phil Davies
Senior Writer

As the nation struggles to recover from the Great Recession, many policymakers view small businesses as the best hope for increased hiring and renewed prosperity. Again and again, in speeches and in media interviews, public officials have declared that small businesses are the major generators of new jobs, more so than large firms. By providing assistance to small firms, the thinking goes, government can help them grow and spark an economic resurgence.

"We know that small businesses are the engine of growth in the economy," said Christina Romer, chair of the White House Council of Economic Advisers, on the TV program "Meet the Press" last year. "We absolutely want to do things to help them." Romer was voicing support for the Obama administration's plans to give small businesses greater access to credit. The president himself, in a 2009 speech, referred to small businesses as "job generators" and "the heart of the American economy."

This conventional wisdom about the job-creating powers of small firms is nothing new, nor is it a partisan argument; Presidents Ronald Reagan, Bill Clinton and George W. Bush also praised small enterprise, quoting statistics on the large share of new jobs created by small businesses. For decades, public policy has favored small businesses with tax breaks, regulatory relief, low-interest loans from the U.S. Small Business Administration (SBA) and other support programs. Government has striven even harder to lend a hand to small firms during the economic trauma of the past two years, focusing much of its aid on increasing the flow of capital to small firms.

Concerned that small businesses are having trouble borrowing in a tight credit market, federal lawmakers in 2009 authorized the use of economic stimulus funds to waive SBA fees and increase guarantees for bank loans to small businesses. In June of this year, the U.S. House passed a bill that would establish a $30 billion federal capital fund for community banks to encourage them to lend to small businesses. The Federal Reserve System has also tried to help small businesses gain access to credit by supporting secondary lending markets through its Term Asset-Backed Securities Loan Facility and by encouraging banks to lend to creditworthy small firms.

There's no question that small businesses are an important source of jobs; firms with fewer than 50 workers employ roughly one-third of all Americans. And, in today's difficult credit environment, small firms may need a leg up in obtaining the capital they need to expand and hire. "Making credit accessible to sound small businesses is crucial to our economic recovery and so should be front and center among our current policy challenges," Federal Reserve Chairman Ben Bernanke said in July at a Federal Reserve forum on addressing the financing needs of small businesses.[1] The forum was the capstone of a nationwide series of meetings in which small-business owners, lenders,

* WARNING
The workings of this engine are complicated, ambiguous and not fully understood.

government officials and other stakeholders shared ideas about the challenges facing small firms.

But aside from other merits small businesses may have—and regardless of whatever policies are promulgated to help them—the question remains: Are they actually the fountainhead of job creation, as advocacy and support for small businesses over the years imply? Economists have sharply debated the issue for 30 years. Some investigators conclude that small firms do indeed punch above their weight class in generating job gains. Others, looking at similar data, find scant evidence to support this conventional wisdom.

At first blush, settling the question seems straightforward, a matter of dividing businesses into small and large categories, then calculating which group creates more net (gains minus losses) jobs in proportion to its share of total employment. But in fact, the issue is far from simple. Ambiguity and statistical pitfalls abound; much depends on the methods researchers use to analyze data on job flows. Matters that economists struggle with in assessing the job-creating capacity of small firms include the reliability of long-run data on hiring and firing by businesses and how to allocate changes in employment based on firm size (different counting methods can produce strikingly divergent results).

This complexity often gets lost in translation when public officials talk about jobs, said John Haltiwanger, an economist at the University of Maryland who has done extensive research on job creation. "Unfortunately, there remains persistent confusion about the role of small businesses in job growth," he said in an e-mail. One example is a failure to distinguish between gross and net job creation (more on that later).

To trace the origins of the ongoing debate about the role of small businesses in job creation, you have to go back to the 1970s, when a researcher at the Massachusetts Institute of Technology (MIT) overturned the then-conventional wisdom that *large* firms were mainly responsible for new job growth.

In Brief

Who creates jobs?

- Small businesses have long received government support on the assumption that they are the primary engine of job creation. However, the question of whether they actually generate more net job growth than large businesses is far from settled.
- Methodological issues, including different ways of allocating job gains and losses based on firm size, have frustrated economists in their attempts to either confirm or debunk the conventional wisdom.
- Some recent analysis suggests that *young* businesses generate a disproportionate fraction of new net jobs. Future research may shed more light on which types of businesses best stimulate job growth.

Does (firm) size matter?

David Birch was one of the first economic researchers to provide hard evidence for the idea that small businesses are the wellspring of job growth. In a groundbreaking study at MIT, he analyzed data on over 5 million business establishments compiled by the Dun & Bradstreet (D&B) credit rating company, looking for patterns in job growth by firm size and age. He relied on longitudinal data—records that follow the same firms over a number of years—to analyze employment trends. Previously, labor economists had studied aggregate statistics to gauge employment growth, simply counting annual increases or decreases in jobs in each size class.

Birch's findings were startling; in contrast to the aggregate studies, which had consistently found that big firms account for most net employment growth, his analysis identified small firms as the economy's primary generators of new jobs. "The results tell a clear story," Birch wrote in a seminal 1979 report. "On the average about 60 percent of all jobs in the U.S. are generated by firms with 20 or fewer employees."[2] Birch also estimated that firms with 100 or fewer employees created 82 percent of all net new jobs from 1969 to 1976. In contrast, large firms with over 500 workers accounted for less than 15 percent of net job growth.

Birch's conclusions, restated in subsequent papers and a popular 1987 book he wrote, lent credence to government policies that treated small

businesses as vital job generators. Congress created the SBA in 1953 to aid small businesses by providing them with ready access to credit. Since the 1970s, a raft of federal and state laws has granted small businesses tax incentives; exemptions from environmental rules, insurance requirements and other regulations; and other forms of government support.

Even as Birch's findings gained wide currency in policy circles, some economists took him to task, questioning his results and the economic impact of small-business jobs creation. A 1982 study reexamined the D&B data and found that small businesses accounted for much lower percentages of net job growth—roughly proportional to their share of the labor force—than those reported by Birch. A 1990 critique argued that even if more net new jobs emanate from small firms than large ones, they're less desirable because they pay lower wages and don't last as long as positions at large firms.

One of the strongest retorts to Birch came in a widely cited study by Haltiwanger, University of Chicago economist Steven Davis and Scott Schuh, an economist for the Federal Reserve Board at the time. In a 1993 paper and later book, the economists criticized Birch's methods, impugned the quality of his data and drew a different conclusion about the role of small businesses in creating jobs.

Davis, Haltiwanger and Schuh claimed that Birch had fallen victim to a "regression fallacy" that exaggerates the impact of small firms on job growth due to transitory movements among size classes. Adopting different statistical methods in an attempt to correct for the fallacy, they conducted their own longitudinal study of employment at U.S. manufacturing plants—and found little difference between the net job growth rates of small and large firms. "In a nutshell, net job creation behavior in the U.S. manufacturing sector exhibits no strong or simple relationship to employer size," the researchers wrote.[3]

Argument over the job-creating potency of small businesses has continued, with considerable attention devoted to methodological matters such as longitudinal links, class-size boundaries and the regression fallacy (also called regression-to-the-mean bias).

A 2008 study[4] funded by the Kauffman Foundation, an organization devoted to entrepreneurship, supports the small-business job engine hypothesis. Analyzing longitudinal data for virtually every employer in the country, the researchers found that "net job creation is in fact tilted towards smaller businesses," said study co-author David Neumark, an economics professor at the University of California, Irvine, in a telephone interview.

Another recent paper that examined the question through multiple lenses, including U.S. Census data and employment figures from Denmark, France and Brazil, concluded that the balance of job creation shifts between large and small firms according to the business cycle; small businesses are powerful job generators during recessions and in the early stages of recovery. However, in an expanding economy, large firms take over the lead in creating new net jobs.[5] (Whether that pattern applies to the current economic recovery has yet to be determined.)

Statistically speaking

"There are three kinds of lies: lies, damned lies, and statistics," Mark Twain famously said, long before anyone kept data on job growth by size of business. Labor statistics don't lie in the sense that Twain implied; they distill the activities of myriad firms and workers into a form that allows economists to see patterns that might otherwise remain hidden. But statistics on jobs created by different-sized firms tend to induce head-scratching, both by members of the public and by economists.

On a basic level, figures quoted by public officials can give a skewed picture of how jobs are created, and by whom. For example, the SBA defines a small business as a firm with fewer than 500 employees. But that definition encompasses 99.7 percent of U.S. employers and roughly half of all workers in the private sector. Many people would consider a company with 300 or 400 employees fairly big—hardly a mom and pop operation. And most economists studying job creation use a lower cutoff for small businesses—20, 50 or 100 employees. SBA tables that included a medium-sized category for larger firms with fewer than 500 workers

would clarify the contribution of small firms to overall job growth.

In addition, figures used to measure job creation can obscure the vast amount of job creation and destruction that goes on in the economy; that "churn," as economists call it, means that it's important to distinguish between "gross" and "net" job creation. *Gross* job creation data count the number of new hires by firms in a given period, before "separations"—layoffs, retirements and voluntary quits—are subtracted. *Net* job creation is the number of jobs that remain after separations are accounted for—how much the workforce either grew or shrank overall. An oft-quoted SBA statistic states that small businesses have accounted for almost two-thirds of net new jobs over the past 15 years.

In public statements, officials often use one term when they should use the other, omitting the crucial qualifier "net," for example, when referring to small-business employment growth. And net figures give no inkling of the total number of jobs created by firms across the size spectrum. As Davis, Haltiwanger and Schuh noted in their 1993 paper, "a common confusion between net and gross job creation distorts the overall job creation picture and hides the enormous number of new jobs created by large employers."[6]

In 2007, before the recession took hold, U.S. firms (of all sizes) increased their net hires by 1.5 million over the previous year, according to Bureau of Labor Statistics (BLS) data. The gross number of jobs created during that period was 13.4 million, while at the same time, 11.9 million jobs were eliminated.

Let us count the ways

On a more abstruse level, trouble with statistics goes a long way toward explaining why economists who have studied job creation for years can come to starkly opposing conclusions about the contribution of small businesses to net employment growth.

The great conundrum for economists trying to prove or disprove the conventional wisdom about small-business job creation is how to accurately measure net job growth by class size; compared with large employers, do small firms generate more jobs than they destroy, proportional to their share of the workforce? Answering this question has proven difficult because of imperfect data on job flows and statistical effects that change the outcome depending on the method used to count job gains and losses.

The source of statistics on job creation can be critical. Researchers have used a variety of longitudinal databases, including refined versions of D&B files, BLS data and Census Bureau records. Each database tracking employment over time at firms or establishments (individual firm locations such as stores or branches) has its strengths, but also weaknesses that may influence the results.

Birch was criticized for using D&B data that didn't square with Census or BLS figures and underreported firm births. Davis, Haltiwanger and Schuh mined Census data on employment at U.S. manufacturing plants, but Haltiwanger now acknowledges that their analysis was "not definitive" because the manufacturing data are arguably not representative of job creation in the economy as a whole. More recent databases developed by the Census Bureau, BLS and private firms are more comprehensive, but have their own limitations. For example, the BLS's Business Employment Dynamics (BED) program tracks firms and establishments only back to 1992.

A particularly vexing problem for researchers lies in the arithmetic of allocating changes in employment to different firm-class sizes. Typically, economists measure job growth (or loss) on an annual basis, counting the number of new hires or layoffs at businesses compared with staffing levels in the previous year. Job gains and losses are tabulated according to various class sizes—say, for example, to firms with fewer than 10 workers, those with 10 to 19 workers, those with 20 to 50 and so on, up to large corporations with over 1,000 employees.

There's nothing complicated about this process. But the math gets tricky when businesses change size classes as they add or lose jobs, moving up or down the scale from one year to the next. If a firm is initially classified as "small," then hires more workers and moves up to a larger size class during

Supersize Me

How base-year sizing increases apparent net job growth by small firms

Small firms = 1–49 workers, Large firms = over 50 workers

	Firm 1	Firm 2	Firm 3
Year 1	10	40	56
Year 2	14	67	44
Net change	**4**	**27**	**–12**

Net job creation of small firms = **31**
Net job creation of large firms = **–12**

Small firms appear to produce more net job growth than large firms when changes in employment are allocated according to a firm's size in the *base* year (Year 1). When Firm 2 hires more workers and shifts into the large-firm category, all of its job gains count as job growth by small firms. When Firm 3 shrinks and becomes a small firm, all of its job losses count as losses by large firms. So small firms account for 31 new jobs (= 4 + 27) using this method. Large firms lose 12 jobs.

If *end*-year sizing were used (job gains or losses allocated according to the firm's size in Year 2), large firms would appear to grow faster than small firms: a loss of 8 net jobs (4 – 12) by small firms versus 27 new jobs by large firms.

the next 12 months, should the additional jobs be credited to the small-firm category or to the large one? Conversely, if a large firm shrinks and becomes small, are those losses laid at the door of a big firm or a small firm?

One counting method attributes employment changes to whatever size class the firm occupied in the initial or "base" year, before the firm grew or shrank. Another approach allocates the jobs gained or lost by a given firm into different firm-size categories according to the size of the firm in the current or "end" year.

One accounting method isn't more "correct" than the other, but the choice makes a significant difference to the researcher's ultimate findings. Birch in his D&B study used base-year sizing. This is also the method used by the SBA to compute annual job creation and destruction in many of its statistical reports and tables. In and of itself, base-year sizing increases the apparent contribution of small firms to job growth, because an increase in employment that lifts a small firm into the large category is credited to the small size class. Year-over-year job losses by large firms are debited to that size class (see accompanying table for a more detailed explanation).

The regression fallacy

This effect is magnified by another, subtler statistical phenomenon that causes consternation for researchers on a variety of phenomena, including job creation. This is the regression-to-the-mean bias that Davis, Haltiwanger and Schuh claimed

skewed Birch's results, a type of distortion that renowned economist Milton Friedman called the "most common fallacy in the statistical analysis of economic data."[7] When base-year accounting is used, the regression fallacy systematically allocates job growth to smaller size classes while allocating job losses to larger size classes. The result: consistently higher employment growth rates for small firms.

Think of the regression bias as random "noise" caused by measurement errors or momentary fluctuations in employment at individual firms. Suppose that in the year the researchers designated as the "base year," a manufacturer that has been large for a decade becomes "small" because of a temporary shock—a product recall, for example—that forces it to lay off workers. The next year the company hires back those workers and "regresses"—or returns—to its customary long-run size. But because the firm was small in the base year of the study, the restored jobs are counted as job gains by firms in the small category.

The fallacy works in reverse for a small manufacturer that enjoys a surge in sales and is temporarily classified as large; when in the following year it reverts to its regular size, the drop in head count goes into the large-firm column as a job loss.

Many economists and some employment databases have tried to address the regression fallacy—and the inherent tendency of base-year sizing to inflate job creation by small businesses—by using various statistical techniques that smooth the distribution of job gains and losses among size categories. For example, "dynamic sizing," used by the BLS to compute BED figures, divides up quarterly changes in employment by a given firm, assigning incremental gains or losses during the quarter as closely as possible to the firm-size class in which they actually occurred.

However, these attempted statistical fixes haven't eliminated concern that statistical effects cloud our understanding of the relationship between firm size and job creation.

Still looking for answers

Small businesses are struggling in the wake of the recession, not hiring as readily as they did in past recoveries. A U.S. Treasury department analysis of unpublished BLS figures shows that between July of 2009 and last February, firms with fewer than 50 employees lost over 150,000 net jobs in an average month, while firms with at least 250 employees slightly increased their hiring. Possible explanations for this lingering joblessness at small enterprises include slack demand for goods and services, an uncertain economic outlook and restricted access to credit—factors that may disproportionately affect small firms.

In response, policymakers have redoubled their efforts to help small businesses. In September, the U.S. Senate was considering giving small businesses $18 billion in capital-equipment write-offs and other tax breaks, in addition to $30 billion in federal loan funds approved earlier by the House.

Yet there's no consensus among economists that this government assistance is going where it can do the most good to alleviate unemployment. After three decades of investigation, the question of whether small firms do indeed create proportionally more jobs than large firms resists resolution. Uncertainties about the reliability of employment data and thorny statistical problems such as the regression fallacy continue to bedevil researchers. Some economists have even suggested that while these statistical issues are real, they don't have a major quantitative impact on the final results.[8]

It's also quite possible that economists have been asking the wrong question about the agents of job creation all along. New research in the field points to *young* firms as the true dynamos of employment growth.

In the 1990s, Haltiwanger took a stand against the conventional wisdom that small businesses outperform big ones in job creation; today he believes that it's not a matter of small versus large, but young versus old. In a recent study, Haltiwanger teamed with researchers at the Census Bureau to analyze 13 years of Census data on U.S. business establishments, controlling for the effect of firm age on net job generation. They found no systematic link between net growth rates and firm size. But the contribution of firms less than 10 years old, particularly startups, to job creation was substantial. Startups less than a year old account for only 3 percent of U.S. employment but almost 20 percent of new gross jobs.[9]

Many new businesses fail, destroying jobs, the researchers note in a working paper. But young firms that survive add employees quickly, outpacing more mature businesses in net job growth. Of course, most new businesses are small, so "one might view this is as a more nuanced view of the contribution of small businesses," Haltiwanger said via e-mail. When presenting his findings to government officials, he often jokes that the SBA should be renamed the Young Business Administration.

Future research may shed more light on which types of businesses—small or large, young or old—stimulate job growth the most. The question takes on extra significance during recessions, when policymakers are looking for ways to jump-start economic recovery. As Bernanke and other Federal Reserve officials have noted in recent months, small businesses are central to job creation in this country. But for now, it seems, the conventional wisdom that small businesses are the *primary* source of job creation remains a matter for continuing debate. R

Endnotes

[1] Bernanke, Ben S. 2010. Remarks at "Addressing the Financing Needs of Small Businesses," a forum organized by the Board of Governors of the Federal Reserve System, July 12, Washington, D.C. http://www.federalreserve.gov/newsevents/speech/bernanke20100712a.htm

[2] Birch, David L. 1979. *The Job Generation Process*, Unpublished report prepared by the MIT Program on Neighborhood and Regional Change for the Economic Development Administration, U.S. Department of Commerce, Washington, D.C., p. 29.

[3] Davis, Steven J., John Haltiwanger, and Scott Schuh. 1993. "Small Business and Job Creation: Dissecting the Myth and Reassessing the Facts." NBER Working Paper 4492, p. 10.

[4] Neumark, David, Brandon Wall, and Junfu Zhang. 2008. "Do Small Businesses Create More Jobs? New Evidence for the United States from the National Establishment Time Series." Institute for the Study of Labor (Bonn, Germany) Discussion Paper 3888, (forthcoming in *Review of Economics and Statistics*).

[5] Moscarini, Giuseppe, and Fabien Postel-Vinay. 2009. "Large Employers Are More Cyclically Sensitive." NBER Working Paper 14740.

[6] Davis, et al., p. 2.

[7] Friedman, Milton. 1992. "Do Old Fallacies Ever Die?" *Journal of Economic Literature*, 30(4), p. 2131.

[8] Davidsson, Per, Leif Lindmark, and Christer Olofsson. 1998. "The Extent of Overestimation of Small Firm Job Creation—An Empirical Examination of the Regression Bias." *Small Business Economics* 11, pp. 87–100.

[9] Haltiwanger, John, Ron S. Jarmin, and Javier Miranda. 2010. "Who Creates Jobs? Small vs. Large vs. Young." Working paper, pp. 33–34.

Thomas Sargent

All scholars strive to make important contributions to their discipline. Thomas Sargent irrevocably transformed his.

In the early 1970s, inspired by the groundbreaking work of Robert Lucas, Sargent and colleagues at the University of Minnesota rebuilt macroeconomic theory from its basic assumptions and micro-level foundations to its broadest predictions and policy prescriptions.

This "rational expectations revolution," as it was later termed, fundamentally changed the theory and practice of macroeconomics. Prior models had assumed that people respond passively to changes in fiscal and monetary policy; in rational expectations models, people behave strategically, not robotically. The new theory recognized that people look to the future, anticipate how governments and markets will act, and then behave accordingly in ways they believe will improve their lives.

Therefore, the theory showed, policymakers can't manipulate the economy by systematically "tricking" people with policy surprises. Central banks, for example, can't permanently lower unemployment by easing monetary policy, as Sargent demonstrated with Neil Wallace, because people will (rationally) anticipate higher future inflation and will (strategically) insist on higher wages for their labor and higher interest rates for their capital.

This perspective of a dynamic, random macroeconomy demanded deeper analysis and more sophisticated mathematics. Sargent pioneered the development and application of new techniques, creating precise econometric methods to test and refine rational expectations theory.

But by no means has Sargent limited himself to rational expectations. Among his dozen books and profusion of research articles are key contributions to learning theory (the study of the foundations and limits of rationality) and to economic history, including influential work on monetary standards and international episodes of inflation.

Interviewed here by now-retired Research Director Art Rolnick, a colleague since the 1970s at the University of Minnesota and Minneapolis Fed, Sargent explores issues ranging from polar models of banking regulation and crisis to causes of persistently high unemployment to a compelling defense of modern macro. Underlying the entire conversation is the "vocabulary of rational expectations," observes Sargent. "In our dynamic and uncertain world, our beliefs about what other people and institutions will do play big roles in shaping our behavior."

Photography by Peter Tenzer

MODERN MACROECONOMICS UNDER ATTACK

Rolnick: You have devoted your professional life to helping construct and teach modern macroeconomics. After the financial crisis that started in 2007, modern macro has been widely attacked as deficient and wrongheaded.

Sargent: Oh. By whom?

Rolnick: For example, by Paul Krugman in the *New York Times* and Lord Robert Skidelsky in the *Economist* and elsewhere. You were a visiting professor at Princeton in the spring of 2009. Along with Alan Blinder, Nobuhiro Kiyotaki and Chris Sims, you must have discussed these criticisms with Krugman at the Princeton macro seminar.

Sargent: Yes, I was at Princeton then and attended the macro seminar every week. Nobu, Chris, Alan and others also attended. There were interesting discussions of many aspects of the financial crisis. But the sense was surely not that modern macro needed to be reconstructed. On the contrary, seminar participants were in the business of using the tools of modern macro, especially rational expectations theorizing, to shed light on the financial crisis.

Rolnick: What was Paul Krugman's opinion about those Princeton macro seminar presentations that advocated modern macro?

Sargent: He did not attend the macro seminar at Princeton when I was there.

Rolnick: Oh.

Sargent: I know that I'm the one who is supposed to be answering questions, but perhaps you can tell me what popular criticisms of modern macro you have in mind.

Rolnick: OK, here goes. Examples of such criticisms are that modern macroeconomics makes too much use of sophisticated mathematics to model people and markets; that it incorrectly relies on the assumption that asset markets are efficient in the sense that asset prices aggregate information of all individuals; that the faith in good outcomes always emerging from competitive markets is misplaced; that the assumption of "rational expectations" is wrongheaded because it attributes too much knowledge and forecasting ability to people; that the modern macro mainstay "real business cycle model" is deficient because it ignores so many frictions and imperfections and is useless as a guide to policy for dealing with financial crises; that modern macroeconomics has either assumed away or shortchanged the analysis of unemployment; that the recent financial crisis took modern macro by surprise; and that macroeconomics should be based less on formal decision theory and more on the findings of "behavioral economics." Shouldn't these be taken seriously?

It is true that modern macroeconomics uses mathematics and statistics to understand behavior in situations where there is uncertainty about how the future will unfold from the past. But a rule of thumb is that the more dynamic, uncertain and ambiguous is the economic environment that you seek to model, the more you are going to have to roll up your sleeves, and learn and use some math.

Sargent: Sorry, Art, but aside from the foolish and intellectually lazy remark about mathematics, all of the criticisms that you have listed reflect either woeful ignorance or intentional disregard for what much of modern macroeconomics is about and what it has accomplished. That said, it is true that modern macroeconomics uses mathematics and statistics to understand behavior in situations where there is uncertainty about how the future will unfold from the past. But a rule of thumb is that the more dynamic, uncertain and ambiguous is the economic environment that you seek to model, the more you are going to have to roll up your sleeves, and learn and use some math. That's life.

Rolnick: Putting aside fear and ignorance of math, please say more about the other criticisms.

Sargent: Sure. As for the efficient markets hypothesis of the 1960s, please remember the enormous amount of good work that responded to Hansen and Singleton's ruinous 1983 *JPE* [*Journal of Political Economy*] finding that standard rational expectations asset pricing theories fail to fit key features of the U.S. data.[1] Far from taking the "efficient markets" outcomes for granted, important parts of modern macro are about understanding a large and interesting suite of asset pricing puzzles, brought to us by Hansen and Singleton and their followers—puzzles about empirical failures of simple versions of efficient markets theories. Here I have in mind papers on the "equity premium puzzle," the "risk-free rate puzzle," the "Backus-Smith" puzzle, and on and on.[2]

These papers have put interesting new forces on the table that can help explain these puzzles, including missing markets, enforcement and information problems that impede trades, difficult estimation and inference problems confronting agents, preference specifications with novel attitudes toward the timing

and persistence of risk, and pessimism created by ambiguity and fears of model misspecification.

Rolnick: Tom, let me interrupt. Why should we at central banks care about whether and how those rational expectations asset pricing theories can be repaired to fit the data?

Sargent: Well, there are several important reasons. One is that these theories provide the foundation of our ways of modeling the main channels through which monetary policy's interest rate decisions affect asset prices and the real economy. To put it technically, the "new Keynesian IS [investment-savings] curve" is an asset pricing equation, one of a form very close to those exposed as empirically deficient by Hansen and Singleton. Efforts to repair the asset pricing theory are part and parcel of the important project of building an econometric model suitable for providing quantitative guidance to monetary and fiscal policymakers.

Another important reason for caring is that monetary policymakers have often been urged to arrest bubbles in asset markets. Easier said than done. Before you can do that, you need a quantitatively reliable theory of asset prices that you can use to identify and measure bubbles.

Rolnick: Before I interrupted, you had begun responding to those criticisms of modern macro. Please continue.

Sargent: I have two responses to your citation of criticisms of "rational expectations." First, note that rational expectations continues to be a workhorse assumption for policy analysis by macroeconomists of all political persuasions. To take one good example, in the spring of 2009, Joseph Stiglitz and Jeffrey Sachs independently wrote op-ed pieces incisively criticizing the Obama administration's proposed PPIP (Public-Private Investment Program) for jump-starting private sector purchases of toxic assets.[3] Both Stiglitz and Sachs executed a rational expectations calculation to compute the rewards to prospective buyers. Those calculations vividly showed that the administration's proposal represented a large transfer of taxpayer funds to owners of toxic assets. That analysis threw a floodlight onto the PPIP that some of its authors did not welcome.

And second, economists have been working hard to refine rational expectations theory. For instance, macroeconomists have done creative work that modifies and extends rational expectations in ways that allow us to understand bubbles and crashes in terms of optimism and pessimism that emerge from small deviations from rational expectations. An influential example of such work is the 1978 *QJE* [*Quarterly Journal of Economics*] paper by Harrison and Kreps.[4] You should also look at a fascinating paper that builds on Harrison and Kreps, written by José Scheinkman and Wei Xiong in the 2003 *JPE*.[5] As I mentioned earlier, for policymakers to know whether and how they can moderate bubbles, we need to have well-confirmed quantitative versions of such models up and running. We don't yet, but we are working on it.

Rolnick: And the other criticisms?

Sargent: OK. The criticism of real business cycle models and their close cousins, the so-called New Keynesian models, is misdirected and reflects a misunderstanding of the purpose for which those models were devised.[6] These models were designed to describe aggregate economic fluctuations during normal times when markets can bring borrowers and lenders together in orderly ways, not during financial crises and market breakdowns.

By the way, participants within both the real business cycle and new Keynesian traditions have been stern and constructive critics of their own works and have done valuable creative work pushing forward the ability of these models to match important properties of aggregate fluctuations. The

It is just wrong to say that this financial crisis caught modern macroeconomists by surprise. ... Researchers have systematically organized empirical evidence about past financial and exchange crises in the United States and abroad. Enlightened by those data, researchers have constructed first-rate dynamic models of the causes of financial crises and government policies that can arrest them or ignite them.

authors of papers in this literature usually have made it clear what the models are designed to do and what they are not. Again, they are not designed to be theories of financial crises.

Rolnick: What about the most serious criticism—that the recent financial crisis caught modern macroeconomics by surprise?

Sargent: Art, it is just wrong to say that this financial crisis caught modern macroeconomists by surprise. That statement does a disservice to an important body of research to which responsible economists ought to be directing public attention. Researchers have systematically organized empirical evidence about past financial and exchange crises in the United States and abroad. Enlightened by those data, researchers have constructed first-rate dynamic

models of the causes of financial crises and government policies that can arrest them or ignite them. The evidence and some of the models are well summarized and extended, for example, in Franklin Allen and Douglas Gale's 2007 book *Understanding Financial Crises*.[7] Please note that this work was available well before the U.S. financial crisis that began in 2007.

Rolnick: I'll come back to that in a second, but you haven't said anything yet about what is to be gained in terms of understanding financial crises from importing insights of behavioral economics into macroeconomics.

Sargent: No, I haven't.

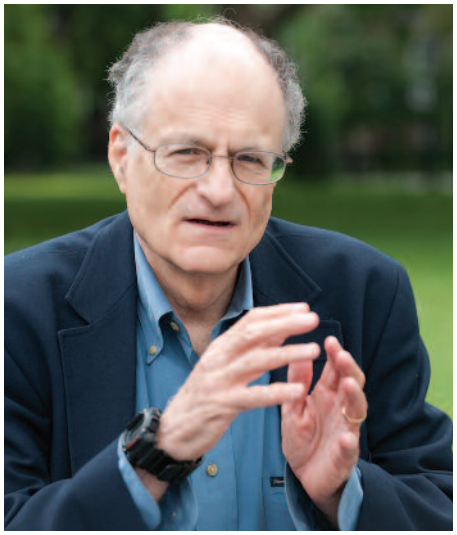
FINANCIAL CRISES

Rolnick: OK then. Well, what useful things does macroeconomics have to say about financial crises, what causes them, how to manage them after they start and what can be done to prevent them?

Sargent: A lot. In addition to the formal literature summarized in the Allen and Gale book, I want to mention the example of the 2004 book by Gary Stern and Ron Feldman, *Too Big to Fail*.[8] That book doesn't have an equation in it, but it wisely uses insights gleaned from the formal literature to frame warnings about the time bomb for a financial crisis set by government regulations and promises. Indeed, one of the focuses of Gary Stern's long tenure as president of the Minneapolis Fed was steadily to draw attention to financial fragility issues and what the government does either to arrest crises or, unfortunately as an unintended consequence, to incubate them.

Rolnick: Thanks for the nice words about Gary, but please elaborate further on macro scholarship and financial crises.

Sargent: I like to think about two polar models of bank crises and what government lender-of-last-resort and deposit insurance do to arrest them or promote them. Both models had origins in papers written at the Federal Reserve Bank of Minneapolis, one authored by John Kareken and Neil Wallace in 1978 and the other by John Bryant in 1980, then extended by Diamond and Dybvig in 1983.[9] I call them polar models because in the Diamond-Dybvig and Bryant model, deposit insurance is purely a good thing, while in the Kareken and Wallace model, it is purely bad. These differences occur because of what the two models include and what they omit.



One of the focuses of Gary Stern's long tenure as president of the Minneapolis Fed was steadily to draw attention to financial fragility issues and what the government does either to arrest crises or, unfortunately as an unintended consequence, to incubate them.

Without deposit insurance, the economy is vulnerable to bank runs. ... The good news in the Diamond-Dybvig and Bryant model, however, is that if you put in government-supplied deposit insurance ... people don't initiate bank runs because they trust that their deposits are safely insured.

The Bryant and Diamond-Dybvig model starts with an environment in which banks can do things that are very worthwhile socially; namely, they provide maturity transformation and liquidity transformation activities that improve the efficiency of the economy. They enable coalitions of people, namely, the banks' depositors, to make long-term investments—loans, mortgages and the like—while at the same time the bank's depositors hold demand deposits, bank liabilities that are short term in duration, because they can withdraw them at any time. Banks thereby facilitate risk-sharing among people with uncertain future liquidity needs. These are all good things.

But there is a potential problem here because for the long-term investments to come to fruition, enough patient depositors must leave their funds in the bank to avoid premature liquidation of a bank's long-term investments. Without deposit insurance, situations can arise that induce even patient depositors to want to withdraw their funds early, causing the banks prematurely to liquidate the long-term investments, with adverse affects on the realized returns.

What triggers a bank run is patient depositors' private incentive to withdraw early when they think that other patient investors are also choosing to withdraw early. Technically speaking, that amounts to multiple Nash equilibria. There are situations in which I run (i.e., withdraw from the bank early) because I expect you to run, and when you also run because you expect me to run. But there are other situations in which we both trust that the other person isn't going to run and we don't run. Which equilibrium prevails is anyone's guess, or something resolved only by an extraneous random device for correlating behavior, a device that economists sometimes call a "sunspot."

So without deposit insurance, the economy is vulnerable to bank runs. The situations where depositors don't run lead to good outcomes, but when there are bank runs, outcomes are bad. The good news in the Diamond-Dybvig and Bryant model, however, is that if you put in government-supplied deposit

insurance, that knocks out the bad equilibrium. People don't initiate bank runs because they trust that their deposits are safely insured. And a great thing is that it ends up not costing the government anything to offer the deposit insurance! It's just good all the way around.

Rolnick: Do you think that an abstract model like this ever influences policymakers?

Sargent: I believe that the Bryant-Diamond-Dybvig model has been very influential generally, and in particular that it was very influential in 2008 among policymakers. A perhaps oversimplified but I think largely accurate way of characterizing the vision of many policy authorities in 2008 was that they correctly noticed that a Bryant-Diamond-Dybvig bank is not just something that has "B A N K" written on its stationary and front door. It's any institution that executes liquidity transformation and maturity transformation, thereby offering a kind of intertemporal risk-sharing.

So in 2008, there were all sorts of institutions that were really banks in the economic sense of the Bryant-Diamond-Dybvig model but that did not have access to explicit deposit insurance, institutions like money market mutual funds, shadow banks, even hedge funds that were doing exactly those maturity-transforming and risk-transforming activities.

When monetary policy authorities, deposit insurance authorities and others looked out their windows in the fall of 2008, they saw Bryant-Diamond-Dybvig bank runs all over the place. And the logic of the Bryant-Diamond-Dybvig model persuaded them that if they could arrest the runs by effectively convincing creditors that their loans—that is, their short-term deposits—to these "banks" were insured, that could be done at little or no eventual cost to the taxpayers. You could nip the run in the bud and really prevent the next Great Depression. This is a very optimistic view of those 2008 interventions enlightened by the Bryant and Diamond-Dybvig model.

But Diamond and Dybvig themselves were cautious about promoting such optimism. In the last part of their 1983 *JPE* paper, Diamond and Dybvig recommend that their readers take seriously the message of a 1978 paper (written at the Minneapolis Fed, as I mentioned earlier) by Kareken and Wallace. That paper includes something important that Diamond and Dybvig recognize that they left out: moral hazard.

Rolnick: And the Kareken-Wallace story?

Sargent: The main idea is that when a government is in the business of being a lender of last resort or a deposit insurer, depending on how it regulates banks, it affects the risk that banks take and the probability that the government is actually going to be required to exercise lender-of-last-resort and bail out facilities. Neil and Jack call it the "moral hazard" problem, which is the idea that when you insure a bank, you alter its incentives to undertake risks.

In the Kareken-Wallace model, deposit insurance is purely a bad thing. Kareken-Wallace envisions a different economic setting than Bryant and Diamond-Dybvig. Of course, like all models, it's an abstraction; it simplifies things in order to isolate key forces. The Kareken-Wallace setting has complete markets. There are markets in all possible risky claims. There are also some people who wanted to hold risk-free deposits.

Kareken and Wallace compare two different situations. In one, there is no deposit insurance; depositors are on their own and know that their deposits are uninsured. If they want to hold risk-free deposits, they'd better hold them in banks that are holding risk-free portfolios. Some very conservative banks emerge that can issue safe deposits because the bank portfolio managers themselves hold assets that allow these banks to pay depositors in all possible states of the world.

The Kareken and Wallace model's prediction is that if a government sets up deposit insurance and doesn't regulate bank portfolios to prevent them from taking too much risk, the government is setting the stage for a financial crisis. ... So, of those two models, the Kareken-Wallace model makes you very cautious about lender-of-last-resort facilities and very sensitive to the risk-taking activities of banks. The Diamond-Dybvig and Bryant model makes you very sensitive to runs and very optimistic about the ability of insurance to cure them. Both models leave something out, and I think in the real world we're in a situation where we have to worry about runs and we also have to worry about moral hazard.

Kareken and Wallace compare that no-deposit-insurance situation to another situation in which a government agency provides deposit insurance that is either free or is priced too cheaply, meaning that it's not priced with a proper risk-loading. Kareken and Wallace show that in that situation, banks have an incentive to become as risky as possible, and as large as possible. Therefore, with a positive probability, banks will fail and taxpayers will have to compensate banks' depositors. It is in banks' shareholders' interest that

the banks organize themselves this way. This lets them gamble with the insurers' and depositors' money.

The Kareken and Wallace model's prediction is that if a government sets up deposit insurance and doesn't regulate bank portfolios to prevent them from taking too much risk, the government is setting the stage for a financial crisis. On the basis of the Kareken-Wallace model, Jack Kareken wrote a paper in the *Federal Reserve Bank of Minneapolis Quarterly Review* referring to the "cart before the horse."[10] He pointed out that if you're going to deregulate financial institutions, which we in the United States did in the late '70s and early '80s (deregulation is the cart), you'd better reform deposit insurance first (that's the horse). You'd better make it clear that financial institutions that take these risks are not allowed to have access to lender-of-last-resort facilities. But the U.S. government didn't do that.

So, of those two models, the Kareken-Wallace model makes you very cautious about lender-of-last-resort facilities and very sensitive to the risk-taking activities of banks. The Diamond-Dybvig and Bryant model makes you very sensitive to runs and very optimistic about the ability of insurance to cure them. Both models leave something out, and I think in the real world we're in a situation where we have to worry about runs and we also have to worry about moral hazard. As you know, an important theme of research for macroeconomics in general and at the Minneapolis Fed in particular has been about how to strike a good balance.

Rolnick: Jack and Neil concluded their 1978 paper with a proposal for dealing with this tension, and that was to require much more capital than was required at the time. Now the government actually requires even less capital than it did when Jack and Neil wrote. If you go back prior to FDIC insurance, turn-of-the-century banks were holding, by some estimates, 20 percent, maybe 30 percent, capital. Capital-equity ratios were that high.

What would *you* recommend? You just observed that if deposit insurance isn't priced properly, that leads you in one direction. And Jack and Neil had this idea of making sure there's a lot more skin in the game, meaning much closer to what banks used to hold when there was no deposit insurance, no too-big-to-fail.

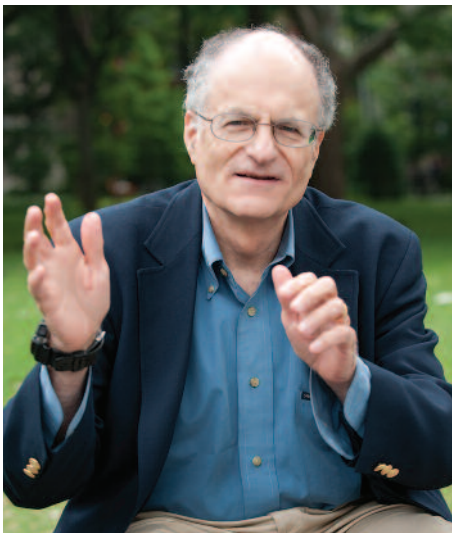
Sargent: The function of capital is exactly to protect against making risky loans. Another proposal is the narrow banking proposal of Milton Friedman and [other economists at the University of] Chicago, which is a proposal to force deposit banks to hold safe portfolios.

Rolnick: Well, with large banks, too-big-to-fail concerns and deposit insurance, I would make the case to tier it based on size. Jack and Neil made the point, I believe, that shareholders of large banks can diversify, but shareholders of smaller banks find it harder to diversify, so they tend to be more risk-averse. Their prediction would therefore have been, I think, that moral hazard is more likely to manifest itself in larger banks—and I think that's what we saw in the 2007-09 financial crisis. How seriously would you take the relevance of the historical evidence that I cited?

Sargent: I would take it very seriously. I recommend a very interesting paper by Warren Weber presented at the Minneapolis Fed conference in honor of Gary Stern this past April in which Warren compared different private insurance arrangements for managing banks' risk-taking before the U.S. Civil War.[11]

THE 2009 FISCAL STIMULUS

Rolnick: A January 2009 article quotes you as saying, "The calculations that I have seen supporting the stimulus package are back-of-the-envelope ones that ignore what we have learned in the last 60 years of macroeconomic research."[12] What calculations had you seen?



In early 2009, I recall President Obama as having said that while there was ample disagreement among economists about the appropriate monetary policy and regulatory responses to the financial crisis, there was widespread agreement in favor of a big fiscal stimulus among the vast majority of informed economists. His advisers surely knew that was not an accurate description of the full range of professional opinion.

Sargent: I said something like that to a reporter. I had just read an Obama administration's Council of Economic Advisers document e-mailed to me by my friend [Stanford University economist] John Taylor.[13] I agreed with John that the CEA calculations were surprisingly naive for 2009. They were not informed by what we learned after 1945.

But I suspect that the council was asked to do something quickly, and they did what they thought was "good enough for government work," as some of us said during my days at the Pentagon in 1968 and 1969. Back-of-envelope work can be a useful starting point or benchmark. But it does mischief when it is oversold.

In early 2009, President Obama's economic advisers seem to have understated the substantial professional uncertainty and disagreement about the

wisdom of implementing a large fiscal stimulus. In early 2009, I recall President Obama as having said that while there was ample disagreement among economists about the appropriate monetary policy and regulatory responses to the financial crisis, there was widespread agreement in favor of a big fiscal stimulus among the vast majority of informed economists. His advisers surely knew that was not an accurate description of the full range of professional opinion. President Obama should have been told that there are respectable reasons for doubting that fiscal stimulus packages promote prosperity, and that there are serious economic researchers who remain unconvinced.

Rolnick: Do any New Keynesian models provide any support for the CEA numbers?

Sargent: Some do; some don't. I recommend looking at calculations by John Taylor and his pals.[14] Based on that work, John remains very skeptical of the 2009 CEA calculations. But Christiano, Eichenbaum and Rebelo have used variants of a New Keynesian model together with particular assumptions about paths of shocks to create quantitative examples of situations in which fiscal multipliers can be as big as those assumed by the CEA.[15]

PERSISTENT UNEMPLOYMENT IN EUROPE (AND NOW THE UNITED STATES?)

Rolnick: Let me go on to another set of questions that I have struggled to answer. Is U.S. unemployment in this recession special? Is it different from the previous 10 recessions? If so, do you have any explanation for why that might be the case? Why it went so high and why it's staying there as long as it is, relative to the pattern of other recoveries?

I haven't heard many economists expound on this, but clearly the labor markets are behaving much differently than they did in previous recoveries, and it's not obvious to me why. I'm curious what you might say about that.

Sargent: May I talk about this by linking to some of my work with Lars Ljungqvist on European unemployment?

Rolnick: By all means.

Sargent: I have little new to say about the details of the big rise in U.S. unemployment since 2008, although the financial crisis was a huge adverse shock to the labor market, so I suspect that we'll be able to explain the rise. But the main thing that concerns me is the threat of *persistent* high unemployment, and here the European experience of the last three decades fills me with dread.

Let me begin by explaining what motivated Lars Ljungqvist and me to study European unemployment, why we have been obsessed by it for 15 years. To Lars and me, Europe's high unemployment rate during the last three decades represents an enormous waste of human resources and individuals' well-being, what we think is a tragedy in the lives of the people who have not been able to participate in the labor market.

Early explanations in the 1980s for Europe's high unemployment were that it was due to insufficient demand and wage rigidities. But soon those explanations came to be regarded as unsatisfactory because they couldn't explain the persistence of unemployment. Some theories blamed Europe's labor market institutions with their generous government-supplied unemployment insurance and strong government-mandated job protection.

But those theories were decisively criticized by Paul Krugman and others who pointed out that the European institutions that liberally subsidized unemployment and disability retirement were there also in the 1950s and '60s, periods when Europe had lower unemployment rates than the United States. Therefore, Krugman and others concluded that you can't blame those generous European social safety nets for the high unemployment rates that Europe has experienced since 1980.

Here's how Lars and I have attacked the problem. We believe that despite Krugman's observation, Europe's generous unemployment compensation system has made an important contribution to sustained high European unemployment, but that those adverse effects came to life only after there occurred what seem to have been permanent changes in the microeconomic environment confronting individual workers. So the culprit was the interactions of those altered microeconomic conditions with those generous European social safety nets.

Rolnick: What changes in microeconomic conditions do you have in mind?

Sargent: Empirical microeconomists have documented that, despite what macroeconomists called the "Great Moderation" in macroeconomic volatility before 2007, individual workers have experienced more turbulent labor market outcomes since the late 1970s and early '80s. Empirical studies have documented increased volatility of both the transient and permanent components of individuals' labor earnings. Peter Gottschalk and Robert Moffitt, Costas Meghir and Luigi Pistaferri, and others have documented that.[16] David Autor and Larry Katz have assembled a convincing catalogue and critical summary of the evidence.[17] So if you look at instances when a job separation causes an individual's earnings to suffer a big reduction, usually that individual must live with a substantial reduction for a long time.

Lars and I use the shorthand "increased turbulence" to refer to this increased volatility and magnitude of adverse earnings shocks at the time of job loss. In the context of several rational expectations models with human capital dynamics and labor market frictions that impede the ability of displaced workers to find new jobs, we have found that an increase in economic turbulence generates persistently high unemploy-

ment when combined with a generous welfare system.

Furthermore, the same government-financed social safety net could actually produce *lower* unemployment in a low-turbulence environment like the 1950s and 1960s. It could do this through strong government-mandated job protection. But when the microeconomic turbulence increases to the high-turbulence post-1980 environment, that same safety net can unleash persistently higher unemployment. An important element of our analysis is the view that a worker's human capital tends to grow when he or she is employed, but deteriorates when he or she is not employed. We analyzed these mechanisms in detail in two papers, one in the *JPE* in 1998, another in *Econometrica* in 2008.[18]

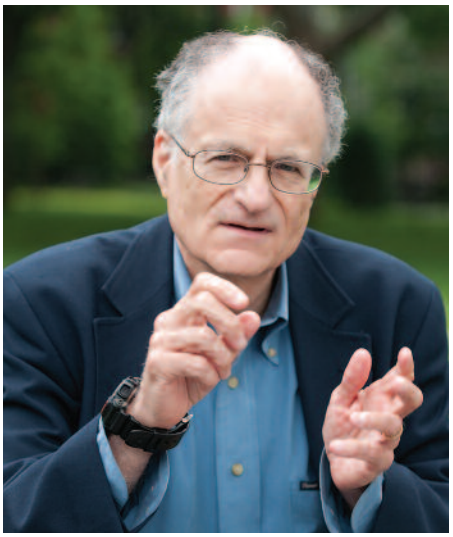
That's our explanation for the higher unemployment rate observed in Europe from 1980 to 2007. Our vision is that an increase in microeconomic turbulence of individual earnings processes occurred in both Europe and the United States. Displaced American workers faced stingier unemployment compensation systems, stingier in both their more limited durations and their lower monthly payments.

Rolnick: When did the microeconomic turbulence begin?

Sargent: The empirical evidence is that it increased substantially sometime in the late 1970s. It happens that it increased just about when high and persistent unemployment broke out in Europe. This is what attracted us to it as a key part of the explanation for the persistent jump in unemployment in Europe relative to the United States.

Rolnick: So turbulence broke out in Europe, OK, but you get the impression that the Great Moderation—a decline in economic volatility—was taking place here in the United States.

Sargent: Well, the so-called Great Moderation really refers to a decrease in *macro*economic volatility. That's why I



After 1980, people in Western economies started suffering bigger drops in their human capital at the moment that they suffer a job displacement. Some of the forces leading to this outcome come from various technological changes going under the umbrella name of "globalization."

Thomas Friedman's 2005 book *The World Is Flat* has many stories testifying to such forces.

stress the difference between individual and aggregate volatility by emphasizing the term *micro*economic. The Great Moderation is indeed there in the aggregate data. An econometrician would think about running a simple autoregressive process for aggregate data and then looking at the error variance. For aggregate data (until 2007), that error variance decreased. But for the micro or individual-level data, just the opposite happened: For individual workers, the error variance—or less technically, unpredictable volatility in earnings—increased.

Rolnick: And why did microeconomic earnings volatility increase?

Sargent: Lars and I believe that when people now become unemployed, they're taking a more or less permanent hit to their level of human capital, a larger one than they might have received before 1980. We have a theory that people build up human capital while they're working on a job, but lose human capital when they're displaced from a job. We think that after 1980, people in Western economies started suffering bigger drops in their human capital at the moment that they suffer a job displacement. Some of the forces leading to this outcome come from various technological changes going under the umbrella name of "globalization."

Thomas Friedman's 2005 book *The World Is Flat* has many stories testifying to such forces.[19] By positing increased turbulence in this sense at the microeconomic level, Lars and I have been able both to come to grips with the observations on aggregate unemployment across Europe and the United States and also to explain some of the micro observations collected by Gottschalk and Moffitt and others. So the Great Moderation seems not to have been occurring at the individual level. Just the opposite.

Our theory goes beyond the aggregate unemployment rate and focuses on individuals. Our models have cohorts of aging heterogeneous workers. Our models imply that people in Europe, especially older workers, are suffering from long-term unemployment because of the adverse incentives brought about by a generous social safety net when it interacts with these human capital dynamics. Unfortunately, the data bear this out. In Europe, there has been a long-term unemployment problem especially affecting older workers.

Rolnick: In your model, what type of labor market frictions impede people who want to work from immediately finding a job?

Sargent: The models that we like best for our purposes view unemployment as an "activity" distinct from "work" and "leisure." We've cast the heart of our theory in several contexts, including, for example, search models in the spirit of George Stigler and John McCall,[20] where finding a job requires a time-

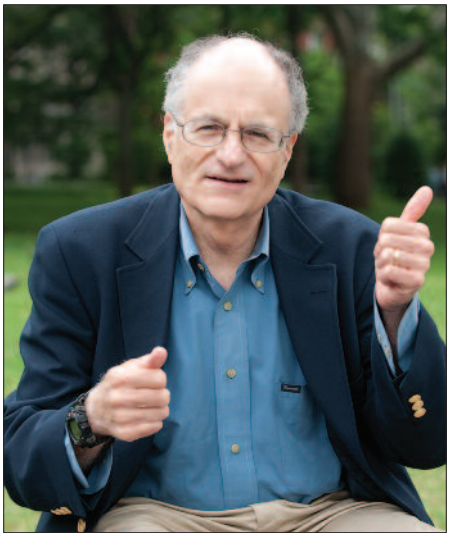
consuming activity of sorting through offers for jobs with various levels of pay and compensating differences; and also Diamond-Mortensen-Pissarides[21] matching models where an aggregate matching function imposes a congestion externality on workers' activity of waiting for a match and firms' activity of waiting for vacancies to be filled. The same forces come through across a variety of structures, so we think there's a lot of robustness to our basic story.

Rolnick: OK, so part of your story for lower U.S. unemployment in the past has to be that in the United States, especially for the older workers, the safety net wasn't as generous. They had to go back and get retrained or whatever; therefore, they chose to be more active in the labor market than their European cousins did.

Sargent: Yes. In a 2003 paper in a volume to honor Edmund S. Phelps, Lars and I exhibited simulations of our model illustrating this.[22] What would a typical, say, 50-year-old worker do if he or she loses his or her job and then immediately gets hit by a human capital loss? What differences in behavior would be exhibited by otherwise similar workers, one facing European benefits versus another facing U.S. benefits?

Our simulations exhibit a force that traps the European worker in unemployment. Unemployment compensation systems typically award you compensation that's linked to your earnings on your last job; those past earnings reflect your *past* human capital, not your current opportunities or current human capital. That can make collecting unemployment compensation at rates reflecting your past (and now obsolete) human capital more desirable than accepting a job whose earnings reflect a return on your current depreciated level of human capital. This mechanism sets an incentive trap that induces the European worker to withdraw from active labor market participation.

Rolnick: Earlier, you said that the European experience with persistently high unemployment over the last three decades fills you with dread about the prospects for the United States.



Collecting unemployment compensation at rates reflecting your past (and now obsolete) human capital [can be] more desirable than accepting a job whose earnings reflect a return on your current depreciated level of human capital. This mechanism sets an incentive trap that induces the ... worker to withdraw from active labor market participation.

Low unemployment rates enabled the United States politically to sustain a modest unemployment compensation system. But the politics of the current situation can imply that so long as unemployment is high, we're going to extend the duration and generosity of benefits. And that extension, done out of the best of motives, is exactly what can lead to the trap of persistently high unemployment.

Sargent: The prospect that concerns me might sound like I'm hardhearted, but that's just the opposite of my feelings. What you've seen in the recent recession—and it's quite natural because it's been so severe—is a tendency of Congress to expand unemployment benefits, over and over again. What Lars and my theory tells us is that if, in the United States, we create a system where unemployment and disability benefits are permanently extended in their generosity and their duration, we will inadvertently put ourselves into the situation that much of Europe has suffered for three decades.

I don't know enough about politics to predict whether that's likely to happen. The unfortunate thing is you can see a multiple equilibrium trap here. Low unemployment rates enabled the United States politically to sustain a modest unemployment compensation system. But the politics of the current situation can imply that so long as unemployment is high, we're going to extend the duration and generosity of benefits. And that extension, done out of the best of motives, is exactly what can lead to the trap of persistently high unemployment. An intriguing thing is that some European countries like Sweden and Denmark are now moving exactly in the opposite direction.

EUROPE AND "UNPLEASANT ARITHMETIC"

Rolnick: Let me ask another question about events in Europe. Some people believe there's a serious conflict between fiscal and monetary policy, that it's the result of the Europeans having asked monetary policy to do things it can't without real fiscal discipline. And as you and Neil pointed out 30 years ago—was it that long ago?!—in "Some Unpleasant Monetarist Arithmetic," you'd better worry about those links. Is that the way you would interpret what's going on in Greece, or Europe in general, and concern over Europe's ability to maintain the euro, that they face some unpleasant arithmetic that could undermine the euro?

Sargent: The people who set up the euro clearly knew about the unpleasant arithmetic and they strove to set things up to protect the euro from any adverse con-

sequences of that arithmetic. Indeed, the whole system was designed to force governments to balance their budgets in a present value sense, adjusting appropriately for growth. Indeed, the Maastricht Treaty actually put in fiscal rules that amounted to overkill in the interests of creating a fail-safe system.

What I mean is that it put in place more restrictive rules on fiscal policy than were needed to express the requirement that a government's budget had to be balanced in the present-value sense with little or no contributions coming from seigniorage revenues from the inflation tax. The treaty built redundancy into the rules by restricting both debt-to-GDP ratios and deficit-to-GDP ratios.

Remember that under the gold standard, there was no law that restricted your debt-GDP ratio or deficit-GDP ratio. Feasibility and credit markets did the job. If a country wanted to be on the gold standard, it had to balance its budget in a present-value sense. If you didn't run a balanced budget in the present-value sense, you were going to have a run on your currency sooner or later, and probably sooner. So, what induced one major Western country after another to run a more-or-less balanced budget in the 19th century and early 20th century before World War I was their decision to adhere to the gold standard.

Rolnick: What does the gold standard have to do with the euro in 2010?

Sargent: The euro is basically an artificial gold standard. The fiscal rules in the Maastricht Treaty were designed to make explicit the present-value budget balance that was unspoken under the gold standard. In terms of the monetarist arithmetic, the rules made sense.

Rolnick: So what's the problem now?

Sargent: Here is what went haywire. In the 2000s, France and Germany, the two key countries at the center of the Union, violated the fiscal rules year after year. Of course, an intriguing thing about the unpleasant arithmetic is that it's about *present values* of government primary deficits, and not just deficits for one, two or three years. And remember that the overkill Maastricht Treaty rules are sufficient but not necessary to sustain present-value budget balance, adjusted for real economic growth, so maybe there was no cause for alarm at that time.

In the 2000s, France and Germany, the two key countries at the center of the Union, violated the fiscal rules year after year. ... They lost the moral high ground to hold smaller countries to the fiscal rules intended to protect monetary policy from the need to monetize government debt.

But in hindsight, there *was* cause for alarm. The reason is that France and Germany lost the moral authority to say that they were leading by example. They lost the moral high ground to hold smaller countries to the fiscal rules intended to protect monetary policy from the need to monetize government debt.

Rolnick: And so …

Sargent: So, a number of countries at the European Union economic periphery—Greece, in particular—violated the rules convincingly enough to unleash the threat of unpleasant arithmetic in those countries. The telltale signs were persistently rising debt-GDP ratios in those countries. Of course, the unpleasant arithmetic allows them to go up for a while, but if that goes on too long, eventually you're going to get a sovereign debt crisis.

Rolnick: What could the European Central Bank do then?

Sargent: Well, here is one thing that you can imagine the ECB doing (which it hasn't). It could take the stance, "If the government of Greece wants to try to issue euro-denominated bonds, let them do it, or try to do it. And if investors want to hold euro-denominated bonds that are understood to be liabilities of the Greek government, and not of the ECB, let them do it. It's not any of the ECB's business. If those bonds threaten to go bad, if Greece just isn't a good risk, that's the bondholders' problem. Let the investors bear that risk. And if Greece defaults or renegotiates, that's the investors' problem, not the ECB's problem."

Rolnick: Of course, the ECB hasn't said that, or at least not yet!

Sargent: Well, one reason the ECB hasn't said that yet is that after the financial crisis of 2008, what seemed to some European banks to be a promising source of higher-yielding instruments was sovereign debt in the form of euro-denominated bonds issued by countries like Greece. The banks located in the center of the euro area, France and Germany, hold Greek-denominated debt, so a threat of default on Greek government debt threatens the portfolios of those banks in other European countries. Because it is the lender of last resort, now it *is* the ECB's business.

Rolnick: Tom, this reminds me of an example of a breakdown in one of the lines between monetary and fiscal policy that you wrote about in your paper "Where to Draw Lines" that you presented at the Stern conference we held in April.[23]

Sargent: Yes, this is a big breakdown in a line between fiscal and monetary policy intended to be set by the Maastricht Treaty in order to enforce that artificial gold standard, as I view the euro to be. Once the monetary authority starts assisting the fiscal authorities of these countries, you've drifted from the original conception of the euro.

Rolnick: Would you argue that Jack and Neil's analysis comes back into play here, the one about too-big-to-fail and moral hazard?

Sargent: Unfortunately, yes, that's what I was trying to suggest.

Rolnick: Did things have to get to this point?

Sargent: Ultimately, that's a question about politics, about which I know too little. But in purely economic terms, things could have gone differently. Here's a "virtual history" of what could have happened:

France and Germany stay "holier than thou" from beginning to end, and always respect the fiscal limits imposed by the Maastricht Treaty. They thereby acquire the moral authority to lead by example, and the central core of euro-area countries are running budgets that without doubt are balanced in a present-value sense. Therefore, the euro is strong. The banks of the core countries (France and Germany again) are well regulated (the message of Kareken and Wallace has been heard), so the banks in France and Germany are not holding any dodgy bonds issued by governments of dubious peripheral countries that have adopted the euro but that flirt with violating the Maastricht Treaty rules.

In this virtual history, the ECB could play tough and let the Greek government default on its creditors by renegotiating terms of the debt. For the euro, letting the Greek bondholders suffer would actually be therapeutic; it would strengthen the euro by teaching peripheral countries that the ECB means business.

More About Thomas J. Sargent

Current Positions

William R. Berkley Professor of Economics and Business, New York University, since 2002

Senior Fellow, Hoover Institution, Stanford University, since 1987

Research Associate, National Bureau of Economic Research, 1970–73 and since 1979

Previous Positions

Donald Lucas Professor of Economics, Stanford University, 1998–2002

David Rockefeller Professor of Economics, University of Chicago, 1991–98

Visiting Scholar, Hoover Institution, Stanford University, 1985–87

Visiting Professor of Economics, Harvard University, 1981–82

Ford Foundation Visiting Research Professor of Economics, University of Chicago, 1976–77

Adviser, Federal Reserve Bank of Minneapolis, 1971–87

Associate Professor of Economics, University of Minnesota, 1971–87; Professor from 1975

Associate Professor of Economics, University of Pennsylvania, 1970–71

First Lieutenant and Captain, U.S. Army; served as Staff Member and Acting Director, Economics Division, Office of the Assistant Secretary of Defense, 1968–69

Professional Affiliations

President, American Economic Association, 2007; President-elect, 2006; Vice President, 2000–01; Executive Committee, 1986–88

President, Econometric Society, 2005; First Vice President, 2004; Second Vice President, 2003; Council, 1995–99, 1987–92; Fellow, 1976

President, Society for Economic Dynamics and Control, 1989–92

Member, Brookings Panel on Economic Activity, 1973

Honors and Awards

Moore Distinguished Scholar, California Institute of Technology, 2000–01

Marshall Lecturer, Cambridge, England, 1996

Erwin Plein Nemmers Prize in Economics, Northwestern University, 1996–97

Fellow, American Academy of Arts and Sciences, 1983

Fellow, National Academy of Sciences, 1983

Mary Elizabeth Morgan Prize for Excellence in Economics, University of Chicago, 1979

Most Distinguished Scholar, University of California, Berkeley, Class of 1964

Phi Beta Kappa, 1963

Publications

Author of a dozen books, including *Robustness* (with Lars Peter Hansen, 2008; *Recursive Macroeconomic Theory* (with Lars Ljungqvist), 2d ed., 2004; *The Big Problem of Small Change* (with François Velde), 2002; *The Conquest of American Inflation,* 1999; and *Bounded Rationality in Macroeconomics,* 1993. Author of more than 170 research papers focusing on macroeconomic theory, time-series econometrics, learning theory, fiscal and monetary policy, and economic history.

Education

Harvard University, Ph.D., 1968

University of California, Berkeley, B.A., 1964

Rolnick: Right. Although if that scenario had been foreseen, Greece might not have been able to issue that debt in the first place.

Sargent: Aha! The plot thickens. So then we confront again the issue of how separate can monetary and fiscal policy be? In the spirit of your observation, remember that there were huge capital gains on Italian debt after it became clear that it would be allowed to join the euro area. So, what really was the reason for those capital gains? Were they based on expectations of a reformed and more disciplined fiscal policy in Italy? Or was it rather an expectation that by joining the euro, Italy had gained access to bailouts from other euro-zone countries?

Note that a related point pertains to the 2009 stress tests in the United States. What did it truly mean when a bank passed the stress test? Did it mean that the bank's balance sheet was solid? Or did it mean that since the Fed said that bank had passed the stress test, the Fed would make sure that henceforth that bank would have access to lender-of-last-resort facilities?

It's difficult to sort these things out. But notice that throughout our discussion, Art, we've been using the vocabulary of rational expectations. In our dynamic and uncertain world, our beliefs about what other people and institutions will do play big roles in shaping our behavior.

Rolnick: Indeed. Thank you again, Tom.

—Art Rolnick
June 15, 2010

Endnotes

[1] Hansen, Lars Peter, and Kenneth J. Singleton. 1983. "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns." *Journal of Political Economy* 91(2), pp. 249–65.

[2] Mehra, Rajnish, and Edward C. Prescott. 1985. "The Equity Premium: A Puzzle." *Journal of Monetary Economics* 15(2), pp. 145–61.

Weil, Philippe. 1989. "The Equity Premium Puzzle and the Risk-Free Rate Puzzle." *Journal of Monetary Economics* 24(3), pp. 401–21.

Backus, David K., and Gregor W. Smith. 1993. "Consumption and Real Exchange Rates in Dynamic Economies with Non-Traded Goods." *Journal of International Economics* 35(3/4), pp. 297–316.

[3] Stiglitz, Joseph E. 2009. "Obama's Ersatz Capitalism." *New York Times,* April 1.

Sachs, Jeffrey. 2009. "Obama's Bank Plan Could Rob the Taxpayer." *Financial Times,* March 25.

[4] Harrison, J. Michael, and David M. Kreps. 1978. "Speculative Investor Behavior in a Stock Market with Heterogeneous Expectations." *Quarterly Journal of Economics* 92(2), pp. 323–36.

[5] Scheinkman, José A., and Wei Xiong. 2003. "Overconfidence and Speculative Bubbles." *Journal of Political Economy* 111(6), pp. 1183–1219.

[6] "New Keynesian economics" refers to the school of thought that has refined Keynes' original theories in response to the Lucas critique and rational expectations. Advocates of New Keynesian theory rely on the idea that prices and wages change slowly (referred to as "sticky" prices and wages) to explain why unemployment persists and why monetary policy can influence economic activity. But their models are adapted from the dynamic general equilibrium models developed by new classical economists such as Sargent and Lucas. See http://www.econlib.org/library/Enc/NewKeynesianEconomics.html for elaboration.

[7] Allen, Franklin, and Douglas Gale. 2007. *Understanding Financial Crises*. Oxford and New York: Oxford University Press.

[8] Stern, Gary H., and Ron J. Feldman. 2004. *Too Big to Fail: The Hazards of Bank Bailouts*. Washington, D.C.: Brookings Institution Press.

[9] Kareken, John H., and Neil Wallace. 1978. "Deposit Insurance and Bank Regulation: A Partial-Equilibrium Exposition." *Journal of Business* 51(July), pp. 413–38. (Also at http://minneapolisfed.org/research/sr/SR16.pdf.)

Bryant, John. 1980. "A Model of Reserves, Bank Runs, and Deposit Insurance. *Journal of Banking & Finance* 4(4), pp. 335–44.

Diamond, Douglas W., and Philip H. Dybvig. 1983. "Bank Runs, Deposit Insurance, and Liquidity." *Journal of Political Economy* 91(3), pp. 401–19. (Also at http://minneapolisfed.org/research/QR/QR2412.pdf.)

[10] Kareken, John H. 1983. "Deposit Insurance Reform or Deregulation Is the Cart, Not the Horse." *Federal Reserve Bank of Minneapolis Quarterly Review* 7 (Spring), pp. 1–9.

[11] Weber, Warren E. 2010. "Bank Liability Insurance Schemes Before 1865." Research Department Working Paper 679, Federal Reserve Bank of Minneapolis. (See also "An Antebellum Lesson" in this issue of *The Region*.)

[12] *Chicago Tribune*. 2009."The Stimulus Rush." Jan. 13. (Also in Brannon, Ike, and Chris Edwards. 2009. "The Troubling Return of Keynesianism." *Tax & Budget Bulletin* 52, Cato Institute, http://www.cato.org/pubs/tbb/tbb_0109-52.pdf.)

[13] Romer, Christina, and Jared Bernstein. 2009. "The Job Impact of the American Recovery and Reinvestment Plan," http://otrans.3cdn.net/ee40602f9a7d8172b8_ozm6bt5oi.pdf.

[14] Cogan, John F., Tobias Cwik, John B. Taylor and Volker Wieland. 2009. "New Keynesian versus Old Keynesian Government Spending Multipliers." Working Paper 47, Stanford University, http://www.hoover.org/news/daily-report/15586; for a newer version of this paper (Jan. 8, 2010), see http://www.stanford.edu/~johntayl/CCTW_100108.pdf.

[15] Christiano, Lawrence, Martin Eichenbaum and Sergio Rebelo. 2009. "When Is the Government Spending Multiplier Large?" Working Paper, Northwestern University, http://www.kellogg.northwestern.edu/faculty/rebelo/htm/multiplier.pdf.

[16] Gottschalk, Peter, and Robert Moffitt. 1994. "The Growth of Earnings Instability in the U.S. Labor Market." *Brookings Papers on Economic Activity* 2, pp. 217–72.

Meghir, Costas, and Luigi Pistaferri. 2004. "Income Variance Dynamics and Heterogeneity." *Econometrica* 72(1), pp. 1–32.

[17] Katz, Lawrence F., and David H. Autor. 1999. "Changes in the Wage Structure and Earnings Inequality," in *Handbook of Labor Economics,* vol. 5. Oxford and New York:

Elsevier Science, North-Holland, pp. 1463–1555.

[18] Ljungqvist, Lars, and Thomas J. Sargent. 1998. "The European Unemployment Dilemma." *Journal of Political Economy* 106(3), pp. 514–50.

Ljungqvist, Lars, and Thomas J. Sargent. 2008. "Two Questions about European Unemployment." *Econometrica* 76(1), pp. 1–29.

[19] Friedman, Thomas L. 2005. *The World Is Flat: A Brief History of the Twenty-First Century*. New York: Farrar, Straus and Giroux.

[20] Stigler, George J. 1962. "Information in the Labor Market." *Journal of Political Economy* 70(2), pp. 94–105.

McCall, John J. 1970. "Economics of Information and Job Search." *Quarterly Journal of Economics* 84(1), pp. 113–26.

[21] Diamond, Peter A. 1982. "Wage Determination and Efficiency in Search Equilibrium." *Review of Economic Studies* 49(2), pp. 217–27.

Mortensen, Dale T. 1982. "Property Rights and Efficiency in Mating, Racing, and Related Games." *American Economic Review* 72(5), pp. 968–79.

Pissarides, Christopher A. 1992. "Loss of Skill During Unemployment and the Persistence of Employment Shocks." *Quarterly Journal of Economics* 107(4), pp. 1371–91.

Mortensen, Dale T., and Christopher A. Pissarides. 1999. "Unemployment Responses to 'Skill-Biased' Technology Shocks: The Role of Labor Market Policy." *Economic Journal* 109(455), pp. 242–65.

[22] Ljungqvist, Lars, and Thomas J. Sargent. 2003. "European Unemployment: From a Worker's Perspective," in *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*. Philippe Aghion, Roman Frydman, Joseph Stiglitz and Michael Woodford, eds. Princeton, N.J.: Princeton University Press, pp. 326–50.

[23] Sargent, Thomas J. 2010. "Where to Draw Lines: Stability versus Efficiency," http://homepages.nyu.edu/~ts43/research/phillips_ver_9.pdf. (See also http://www.minneapolisfed.org/research/events/2010_04-23/index.cfm.)

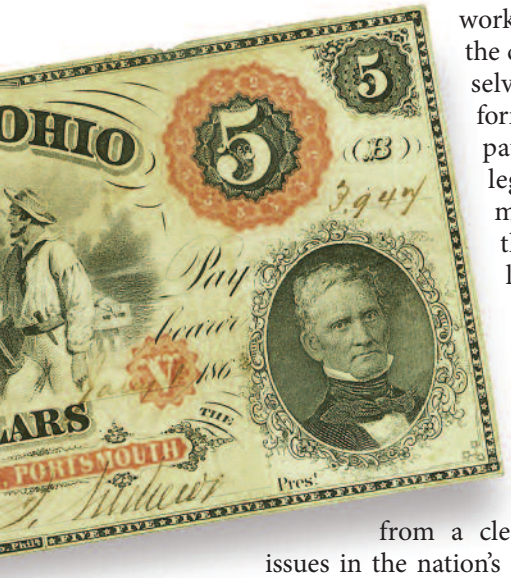
▸ ONLINE
For more about these bank notes:
minneapolisfed.org
The Region, September 2010

An Antebellum Lesson

Bank insurance systems before the Civil War provide a clear message for policy today about the importance of incentives, authority and exposure to loss

Douglas Clement
Editor

The U.S. financial system has evolved radically since its earliest years. Bank functions and organization, financial flows, international networks, government supervision, the currency and buildings themselves—all have been transformed by the nation's historical path, economic growth, and legal and political development.[1] It might therefore be thought that little can be learned about current regulatory matters from analysis of banking systems from centuries ago. Such a conclusion would be seriously wrong.



In reality, some of today's most difficult dilemmas would benefit from a clearer understanding of how issues in the nation's earliest years were addressed and, at times, resolved. This is particularly the case for what is one of the central challenges of modern financial system regulation: Ensuring that government policy does not promote excessive risk-taking by financial institutions. This issue of "moral hazard" (a seemingly archaic term with overtones of puritanical judgment) is paramount in current policy debates. How can government provide the right measure of protection for banks and other financial institutions without encouraging risky behavior?

With explicit provision of deposit insurance and implicit assurance of bailouts, governments have, in the public interest, long supported banks so that they're not subjected to runs by panicked depositors. Such runs can spread rapidly, destroying confidence and freezing liquidity throughout the financial system. And the more interconnected a financial institution is with other parts of the financial system, the greater the threat its collapse represents. This was precisely the rationale for the government's controversial bailout of AIG during the recent financial crisis.

But excessive support or insurance will exacerbate risk-taking, provoking the very financial instability it seeks to curb. A bank that assumes a government rescue may take risks it otherwise wouldn't. The issue of moral hazard has long been a concern to insurance providers of all sorts—health, auto and homeowner insurance companies all worry about clients taking excessive (and potentially expensive) risks, and insurance rates usually reflect the provider's beliefs about the customer's likely risk-taking behavior.

A recent piece of research by Minneapolis Fed Senior Research Officer Warren Weber, "Bank Liability Insurance Schemes Before 1865" (Working Paper 679 online at minneapolisfed.org), draws on the history of bank regulation before the Civil War to paint a vivid picture of how the right financial arrangement can discourage excessive risk-taking.

Warren Weber
Senior Research Officer
Federal Reserve Bank of Minneapolis

Genesis and background

"In truth, the motivation for this paper was Gary's conference," said Weber in an interview, referring to the conference in honor of former Minneapolis Fed President Gary Stern held at the Minneapolis Fed, April 23-24. (Go to "Events" on the Research

page at minneapolisfed.org and to "Too Big to Forget" in the June 2010 *Region*.) Because moral hazard in banking was one of Stern's central concerns, Weber sought out a historical example. "My focus in this paper was moral hazard and the monitoring of risk in the context of 'deposit insurance,' and it turns out that there is a very clear illustration from the mid-1800s."

Many assume that the Federal Deposit Insurance Corp. (FDIC), launched in 1933, was the first significant insurance scheme for banks in the United States, but Weber notes that some states insured deposits well before that. Eight states enacted deposit insurance programs between 1909 and the 1920s, and the National Currency Act of 1863 that established the National Banking System provided for explicit U.S. Treasury guarantee of notes issued by national banks.

But two types of bank insurance schemes were also in effect prior to the Civil War: insurance funds and mutual guarantee systems, and these are the focus of Weber's study. Though they went out of existence a century and a half ago, their history still sheds light on current regulatory quandaries. To initially develop his understanding of the systems, Weber used a classic 1958 study by Carter Golembe and Clark Warburton, *Insurance of Bank Obligations in Six States*, a book-length report created for the FDIC. "Golembe and Warburton really were the pioneers in this, pulling together massive amounts of data and archival information about these early insurance schemes," he observed.

The analysis in terms of incentives, moral hazard and exposure to loss is all Weber's, however, and his examination of the insurance plans provides clear lessons. "I'd argue that their experience demonstrates the critical importance in bank regulation of incentives, the authority to change those incentives, and the question of who bears loss—these points are essential to controlling moral hazard," he noted. Or as he phrases it in the working paper: "[R]egulatory incentives matter. … The schemes that provided the most control of moral hazard were those that had a high degree of mutuality of losses borne by all banks participating in the scheme."

Weber points to recent testimony by Allan Meltzer, a Carnegie Mellon University economist and historian of the Federal Reserve, who testified about bank supervision early this year before the U.S. House Financial Services Committee. (See an interview with Meltzer in the September 2003 *Region* and a review of his book *A History of the Fed*, *Part 1* in the December 2003 *Region*, both issues online at minneapolisfed.org.)"We cannot have deposit insurance without restricting what banks can do," said Meltzer. "The right answer is to use regulation to change incentives—making bankers and their shareholders bear the losses."[2]

The chief lesson of the mid-1800s bank insurance schemes is that when all members of the insurance plan are liable for losses incurred by others, they have an incentive to monitor the behavior of fellow members. And successful schemes provided the power to change it to reduce risk.

The pre-Civil War experience supports Meltzer, writes Weber in his working paper. "The incentives set up by the insurance scheme regulations were important for how well the moral hazard that accompanies any insurance schemes was contained." But Weber contends that the evidence suggests more. "It could be useful to think about expanding the class of agents that could (should?) be made to bear losses from a bank's behavior *beyond the shareholders* of that bank," he writes (emphasis added). "The class could be expanded to include other banks if they were to also have the power or authority to modify the incentives that a bank faces."

The chief lesson of the mid-1800s bank insurance schemes, Weber says, is that when all members of the insurance plan are liable for losses incurred by others, they have an incentive to monitor the behavior of fellow members. And successful schemes provided not only the incentive to *monitor* behavior, but the power to change it to reduce risk. "It's an overused expression," admits Weber, "but having 'skin in the game' makes all the difference to reducing moral hazard. Also, the ability to do something about risk-taking by others is another crucial element."

Historical context

Weber's paper begins with a review of money and banking in the antebellum period when these insurance schemes were active, and his description is a startling reminder of how much has changed.

- There was no central bank of any sort.
- Unlike the *fiat* money of today, the United States had a *commodity* money standard. A dollar was defined in terms of grams of silver or gold, and the federal government issued gold and silver coins.[3]
- But coins were "only a small fraction" of the money supply in the United States, Weber writes. "By far the predominant media of exchange were the notes issued by banks. … Virtually every bank in existence during this period issued its own notes … [that] were redeemable [in gold and silver coins] on demand at that bank."
- Banks were plentiful relative to the U.S. population, growing in number from 356 in 1830 to 705 in 1840, and then doubling to 1,421 by 1860. There are nearly 8,000 FDIC-insured banks today, but the ratio of banks to people in 1860 was nearly twice as high as it is now.
- Bank regulation was exclusively state-based (no federal regulation), and in most states, banks were restricted to a single location.

Two bank insurance systems

In this antebellum period, two types of schemes were established to insure liabilities of member banks: insurance funds and mutual guarantee systems.

Under an *insurance fund* (called a "safety fund" in some states)—established in three states—banks paid a fraction of their capital to the state's bank authority, which would use this insurance fund to reimburse creditors of a bank that failed. Payments to creditors were capped by the funds, though member banks "could potentially be required" to make further contributions.

Under a *mutual guarantee system*—also established in three states—member banks were legally responsible for full repayment of losses incurred by creditors of any of its failed members, "only limited by the market value of assets of all banks."

Weber offers a significant level of detail for the insurance funds established in New York and Vermont and for the mutual guarantee systems established in Indiana and Ohio. The insurance systems established in Michigan and Iowa are ignored because they only existed for a short period.

Insurance funds

The New York and Vermont insurance funds were established in 1829 and 1832, respectively, and lasted until 1863 when all banks became part of the National Banking System.[4] The funds had similar structures. They guaranteed all liabilities, but when a

bank failed, its creditors were paid from the fund only after the failing bank's assets had been completely liquidated, a process that could take some time.

To fund the insurance pool, banks were assessed a percentage of capital, ranging as high as 3 percent in New York and 4.5 percent in Vermont. If the insurance fund was exhausted, additional assessments could be levied until it was replenished, but annual contribution requirements were limited for each bank. A bank could opt out of the fund when its charter expired and regain a portion of its contributions.

Founders of these funds were well aware of the moral hazard such safety nets would create. Weber quotes from an account of the legislative debate over establishing the New York fund:[5]

> [A]nother representative, Mr. Hubbell, pointed out that the very existence of such a fund would relax "public scrutiny and watchfulness which now serve to restrain or detect malconduct."

Founders of these funds were well aware of the moral hazard such safety nets would create. ... To mitigate that problem, states established restrictions on bank activities and supervision of bank conduct.

To mitigate that problem, both states established restrictions on bank activities and supervision of bank conduct. New York stipulated that banks could issue notes of a value no greater than two times capital stock (or shareholder equity), and Vermont set a note issuance limit of three times shareholder equity. New York's law limited loans and discounts to no greater than 2.5 times equity.

Bank commissioners were also established in both states to supervise banks belonging to the insurance fund, and Golembe and Warburton note that such supervisory agencies were an innovation at that time. Weber, though, is skeptical about their effectiveness, pointing out that there were only three bank commissioners in each state to supervise all insured banks. (There were 90 banks in the New York fund when it began; total membership declined over time. Membership in the Vermont fund fluctuated, with a maximum of 16 banks.) He also notes that supervisors weren't authorized to close banks for bad banking practices, only for illegal acts or insolvency. Moreover, "bank commissioners were prohibited from owning stock in any bank," he writes. "As a result, they had no direct stake in the gains or losses from the activities of the banks they supervised." That is to say, supervisors had no financial skin in the game.

Still, Weber emphasizes that supervisors—then and now—are motivated by far more than personal financial gain. In most instances, supervisors are and were highly competent and work to the best of their ability to identify weaknesses in the banking system and have them corrected. And supervision works in part because supervisors know that their careers and reputations depend on solid job performance. A direct financial stake in a bank's health adds another important element to a supervisor's incentive structure.

Mutual guarantee systems

Weber then describes the mutual guarantee systems in Indiana and Ohio. (Again, Iowa's lasted just a short time.) Both state systems were called the "State Bank of …," and all member banks were called "branches" of the State Bank. But the terms were misleading—the "State Bank" did no business of its own, and each "branch" operated as an independent bank, with its own stockholders, notes and profits. Indiana's system, with 13 branches, operated from 1834 to 1857. Ohio's had (effectively) 34 branches, operating from 1845 to 1863.

To mitigate moral hazard, these systems instituted restrictions on note issuance, loans and discounts similar to those implemented by the New York and Vermont insurance funds. But there were significant differences in supervision, according to Weber. "The supervision of the Branches was done by a state board comprising members appointed by the state legislature and *one director from each branch*," he writes (emphasis in original). The state board, which examined each branch two to three times a year, could close a branch, limit its dividend payments, and restrict loans and discounts.

Moreover, "each member of the system was mutually responsible for at least some of the liabilities of the other banks in the system." Indiana's branches were required by law to guarantee "all debts, notes, and engagements of each other." Ohio's law required that "[e]ach solvent branch shall contribute … to the sum necessary for redeeming the notes of the failing branch."

The upshot was that in a mutual guarantee system, each branch shared in the losses but not the profits of its fellow system members, and was able to supervise the others (by virtue of having one of its directors on the state board). "In other words, the 'regulators' had a direct, one-sided financial stake in the outcome of the branches they regulated," writes Weber. And because each branch was accountable for losses of others, it had every reason to monitor the banking practices of other branches. In sum, every branch had the motive, means and opportunity to protect the health of its peers.

Runs, failure and coverage

Did these systems work? Not entirely, according to Weber's analysis. To explain this conclusion, he gives an account of how each state's banks fared—in terms of runs, failures and coverage for creditors—during national bank panics.

There were two significant panics during that historical period; the first began on May 4, 1837, with banks in Natchez, Miss., suspending payment on their notes. Panic spread quickly, and by May 19, "virtually all banks in the country had suspended payments."[6] Banks resumed payment by the middle of 1838, but a second wave of suspension started in 1839, spreading across the nation with the exceptions of banks in New York and New England. These two waves of bank panic were followed by a severe economic contraction that lasted until 1843.

A second major panic began in the late summer of 1857, most likely starting in Ohio and spreading in subsequent months to Philadelphia, New York and Boston, followed by a contraction that continued until 1858.

So, how well did the insurance plans serve their members during these crises?

Bank runs

Unfortunately, finds Weber, "it is evident that these insurance schemes did not prevent bank runs during the panics of 1837 and 1857." In 1837, banks in New York suspended payment on their notes on May 9, just five days after the Natchez suspensions, and banks throughout New England, including Vermont, did so the following day. Nor did Indiana's mutual guarantee system prevent the potential for runs there. Branches of the State Bank suspended payment in May 1837. (Ohio's system didn't begin until 1845.)

Weber argues that the New York and Vermont insurance funds *may* have led to an early resumption of payment in those states. They resumed in May 1838, while banks in most of the

The Suffolk System

A strong pattern among banks in four states doesn't prove a theory, of course. But another banking system in the antebellum period, the Suffolk Banking System of New England, offers further support for the importance of exposure to loss and authority to restrain risky banking activity. The System provides an example in which a motivated party (the Suffolk Bank of Boston) could and did take action to curb risk-taking by (and improve survival of) interconnected banks (those who joined Suffolk's note-clearing system) whose potential losses would negatively affect its interests.*

Suffolk was a regional note-clearing system—not a bank liability insurance scheme—run by the Suffolk Bank of Boston from 1825 to 1858. By the early 1830s, most banks in New England belonged to the Suffolk System because it enabled them to hold smaller levels of coins and other reserves than would otherwise be required to redeem the notes they issued. Banks could borrow from the Suffolk Bank and pay off the Suffolk loans when their own loans and other assets matured. Another benefit: Notes issued by member banks exchanged at par throughout New England, increasing value and convenience for bank customers. In exchange for these benefits, the Suffolk Bank required its members to keep an interest-free deposit at Suffolk (or another Boston member bank) of 2 percent of bank capital.

If a member bank failed, the Suffolk Bank would be stuck with losses on the bank's notes held on its balance sheet, as well as any overdraft advances made to that bank. (The losses would be borne by Suffolk alone, not mutually by all members as in a mutual guarantee system.) And potential losses could be quite substantial. In the 1830s and 1840s, observes Weber, member banks owed Suffolk about $700,000 on average, climbing to about $1 million in the 1850s. Bank notes held by Suffolk were about $450,000 in the 1830s, rising to roughly $700,000 in the 1850s. These numbers loomed large compared with Suffolk's total capital stock of approximately $1 million in the 1840s and 1850s.

"Thus, the Suffolk Bank had an interest in monitoring the actions of banks that were members of the system," writes Weber. "And it did." He quotes as evidence a letter from Suffolk's president to a Vermont member bank commenting that "too large a portion of your loan … cannot be relied upon at maturity to meet your liabilities."

"Further, the Suffolk Bank had the power to affect the behavior of member banks," writes Weber. Whenever it felt compelled to do so, it notified debtor banks to pay off loans due. Otherwise, the bank's notes would be redeemed by Suffolk for gold and silver coin—solid collateral.

The Suffolk System's apparent ability to reduce bank failure is suggested by Weber's failure rate data from System members in four New England states (Maine, Massachusetts, New Hampshire, Vermont) compared with four other eastern states (Maryland, New Jersey, New York, Pennsylvania). In the Suffolk System, only 24 of 354 banks failed—a rate of 6.8 percent, less than half the 14.5 percent rate of bank failure (47 of 325) in the other four states.

—Douglas Clement

*In earlier work with Arthur Rolnick and Bruce Smith, Warren Weber studied the Suffolk Banking System to evaluate the claim that it was an effective and efficient privately run interbank payments system. They conclude that the System's history of extraordinary profitability suggests that note clearing is a natural monopoly and that "there is no consensus in the literature about whether or not the unfettered operation of markets in the presence of natural monopolies will produce an efficient allocation of resources." Rolnick, Arthur J., Bruce D. Smith and Warren E. Weber. 1998. "Lessons from a Laissez-Faire Payments System: The Suffolk Banking System (1825-58)." *Federal Reserve Bank of Minneapolis Quarterly Review* 22, Summer, pp. 11–21.

country didn't begin paying until August. And the second wave of the 1837 panic didn't cause suspension in New York and Vermont as it did elsewhere. "It is not clear how much of this early resumption and lack of second suspension can be attributed to the … insurance funds [in those states], however," Weber acknowledges; many other New England banks had similar suspension/resumption patterns.

As for the 1857 panic, only Ohio's mutual guarantee system was truly in effect at the time, and while none of the branches suspended payment on their notes, Weber suggests that a variety of unrelated actions taken by the state's bank authorities made it more difficult for note holders to run banks by presenting notes for redemption. The guarantee system was not necessarily a crucial factor.

Bank failures

The story with regard to bank failures is mixed. To analyze failure experience, Weber compares different states with a variety of types of banks.[7] It's a complex picture, but the bottom line seems to be that failure rates for banks that operated under the two insurance systems were "roughly the same as or somewhat higher than those of uninsured chartered banks in the same state or in similar states."

> To control the increased risk-taking that government deposit insurance encourages, the activities of insured banks must be restricted by those parties who have an incentive for doing so.

The stunning exception, though, is Indiana. "There were no failures of the branches of the State Bank of Indiana," Weber writes. But he defers his proffered explanation until the conclusion of the paper. A hint: Think "exposure."

Insurance coverage

A third criterion for evaluating the success of these schemes is the degree to which creditors were made financially whole in the event of bank failure. As Weber points out, doing so was a central rationale for the FDIC, established by the Banking Act of 1933. Representative Henry B. Steagall, a key proponent of deposit insurance, said its purpose was to supply the public with "money as safe as though it were invested in a government bond" and to "prevent bank failures, with depositors walking in the streets."

Weber's thorough analysis of the data finds that results on this criterion varied significantly: Mutual guarantee systems fared far better than insurance funds.

Ten banks that were members of New York's insurance fund made claims on the fund after the crisis of 1837. The first three claims were completely covered. But in 1841, four banks failed, placing claims of over $1.7 million, well beyond the $572,000 available. The next year, three more banks failed; they claimed $532,000 from the fund, which had only $497,000. To cover these claims, New York issued nearly $1 million in state bonds and "all creditors of the failed banks were paid off by the end of 1847."

Because the special bonds allowed the New York fund to pay off all losses, it could be argued that the insurance scheme provided complete coverage. But "in another sense," writes Weber, "at least some creditors suffered losses due to the time delay in receiving final payment." Note holders needing quick access to funds would commonly have to accept a discount of between 30 percent and 50 percent of their notes' face value.

The Vermont situation was worse still. Two members of that state's fund failed and made claims on the fund. Creditors of one failed bank were paid in full after it failed in 1839, *but not until 12 years later*. At the second bank failure, in 1857, less than half the amount claimed was paid off.

The mutual guarantee systems in Indiana and Ohio provided much better coverage for creditors. No branch of the Indiana system failed, so no creditors suffered loss. And though four Ohio branches failed, other members of the system were assessed to redeem in full the notes of the failed four.

In a side note, Weber mentions an interesting parallel to today's policy discussions. In 1855, faced with imminent branch bank failures, the president of Ohio's state board advocated making fund

advances to the branches experiencing liquidity problems, "the object being to sustain the Branch during a period of general alarm, when [its] failure … would have, in all probability, carried several others with it." His rationale, observes Weber, bears remarkable likeness to that used by regulators during the recent crisis in justifying large bailouts to avert broader financial collapse.

Lessons (still to be) learned

The experience of these bank insurance systems has clear implications for controlling moral hazard, notes Weber, with close application to today's financial system, different though it may otherwise be.

Meltzer had it right in his House testimony, says Weber. He stressed that to control the increased risk-taking that government deposit insurance encourages, the activities of insured banks must be restricted by those parties who have an incentive for doing so. To repeat, Meltzer said regulators should "change incentives [by making] bankers and their shareholders bear the losses." Increasing the required amount of capital held by banks would provide *share*holders (among others) added incentive to watch their bank's risk levels. Contingent debt plans would convert debt into equity in the event of bank failure, providing *bond*holders with an incentive to monitor bank actions.

But the lessons of history teach that losses can usefully be shared beyond the equity or debt holders of a particular bank. "All of the pre-Civil War bank liability insurance schemes had at last partial mutuality of losses borne by all banks participating in the scheme," he writes. Expanding the parties exposed to loss from bank risk-taking could be effective. (A provocative if implausible proposal: Create a system whereby the "too-big-to-fail" banks analyzed in the 2009 stress tests are mutually liable for losses of the others. That financial exposure would offer a powerful incentive to monitor competitors' risk-taking.)

But supplying incentive to monitor banking behavior would do little without also providing the ability to change behavior that might inflict (mutual) losses. "The difference between the insurance fund … schemes and the mutual guarantee schemes," writes Weber, "is that the latter also gave survivors (banks that did not fail) the power to regulate the activities of member banks." In the insurance fund systems, bank commissioners were prohibited expressly from owning bank equity; it was a prohibition that made them impartial, perhaps, but also left them without a direct financial interest in curbing risky behavior by the banks they supervised. In the mutual guarantee systems, a director of each branch sat on the state regulatory board, with means as well as motive to restrict imprudent actions of fellow system banks.

But even among the mutual guarantee systems, there was a significant difference in results. Indiana's system achieved a far better outcome than Ohio's in a key respect: no bank failures, and therefore no need for some members to cover losses of others. The explanation?

"The reason for the different outcomes, in my opinion," writes Weber, "is the difference in the amount of 'skin in the game' of the branches of the two systems. It was much higher for the branches of the State Bank of Indiana." By calculating the fraction of capital that an average branch would have to pay out to creditors should another average branch fail, Weber computed the level of capital exposure of the Indiana branches between 1835 and 1856 and Ohio branches between 1846 and 1861.

While levels varied widely from year to year, the general capital exposure of an Indiana branch was about 20 percent, whereas the exposure of an Ohio branch was on the order of 5 percent. Thus, each Indiana branch had much more to lose if a fellow branch failed, and therefore far greater incentive to curb risky behavior by others. What accounted for Ohio's lower exposure? Two factors, explains Weber: The Ohio system guaranteed only bank notes, not "all debts, notes and engagements" as in Indiana, and there were more branches over which to spread losses (roughly 33 in Ohio versus 13 in Indiana).

The most effective system, in other words, must ensure that those with the authority to restrict bank activities will bear the potential loss of increased risk-taking. But the pre-Civil War experience illustrates another important point. "The incentives do not have to apply solely to the shareholders," writes Weber. "[T]he evidence seems to suggest that degree of mutuality [of losses borne] affected the outcomes."

The bottom line

The experience of the insurance funds and mutual guarantee systems of the mid-1800s thus provides powerful lessons for controlling moral hazard today, says Weber. Relying on supervision alone isn't sufficient because supervisors don't bear financial losses if the institutions they oversee fail. "Supervision is fine, and necessary, and there's no question that supervisors then and now were very competent and committed to carrying out their responsibilities," he said. "But if this historical episode is any guide, getting incentives right is critical, and creating direct financial incentives seems to work." In implementing deposit insurance or other measures to limit bank runs and systemic failure, policymakers should consider designing systems that include a higher extent of financial loss-sharing among involved parties, and that provide members with the means to change incentives of other members.

"Regulatory incentives matter for controlling moral hazard," he writes in summing up the pre-Civil war experience with bank insurance liability plans. "The schemes that provided the most control of moral hazard were those that had a high degree of mutuality of losses borne by all banks participating." R

Endnotes

[1] See, for example, "The 'Monster' of Chestnut Street" in the September 2008 *Region* and "The Bank that Hamilton Built" in the September 2007 *Region*, both issues online at minneapolisfed.org.

[2] Meltzer, Allan H. 2010. Testimony to the U.S. House Financial Services Committee, March 17. Meltzer touched on many issues in his testimony, but control of moral hazard was central, and using incentives rather than supervision was his key point: "Trust stockholders' incentives not regulators' rules. Incentives are not perfect, but they are better. … Real financial reform requires that bankers, not regulators, monitor the risk on their balance sheet and accept their losses from mistakes. … That will make for more prudence. I repeat my frequent comment: Capitalism without failure is like religion without sin. It doesn't work well."

[3] Textbook definitions of these terms, from N. Gregory Mankiw's *Principles of Economics*, are that commodity money "takes the form of a commodity with intrinsic value," while fiat money is "money without intrinsic value that is used as money because of government decree." But Weber notes that the true source of value for fiat money remains a debated issue.

[4] Weber suggests that the New York fund essentially stopped providing insurance in 1842.

[5] Chaddock, R. E. 1910. *The Safety Fund Banking System in New York State, 1829-1866*. S. Doc. No. 581, 61st Cong., 2nd sess. Washington, D.C.: Government Printing Office.

[6] Suspended banks would not redeem their notes or deposits for gold and silver, but remained open for other business.

[7] For example: banks with state charters but without insurance, and so-called free banks, which were allowed to operate without a state charter but with restrictions on note issuance.

Tax Buyouts

Raising government revenue without distorting work decisions[1]

Marco Del Negro
Federal Reserve Bank of New York

Fabrizio Perri
Federal Reserve Bank of Minneapolis and University of Minnesota

Fabiano Schivardi
Università di Cagliari and EIEF

Introduction

Little is certain about the United States' fiscal future beyond this: Given foreseeable trends in economic growth, future tax revenues will not cover forecasted mandatory and discretionary expenditures; therefore, a large and growing budget deficit is highly probable.[2] While policymakers may be able to enact modest spending cutbacks, they will undoubtedly need to consider options for raising taxes as well.

Unfortunately, when they do so, they will face a further unpleasant economic reality: Taxes often introduce distortions and inefficiencies that depress economic activity. Indeed, taxes generally undercut the incentive to generate the income on which they are levied.

This economic policy paper addresses that quandary by offering an option with a number of appealing features:

- It allows governments to raise revenues without the labor-discouraging distortion common to income taxes.
- Its elimination of economic distortion contributes to economic activity and well-being.
- Because it allows citizens free choice to opt for an alternative tax arrangement, it is politically viable.

To be specific, this paper suggests that a tax buyout program could achieve the goal of raising revenues without distorting work incentives and thereby diminishing economic activity. The buyout is a contract between the government and individual citizens whereby each person has the option in each tax period to pay a fixed price in exchange for a set

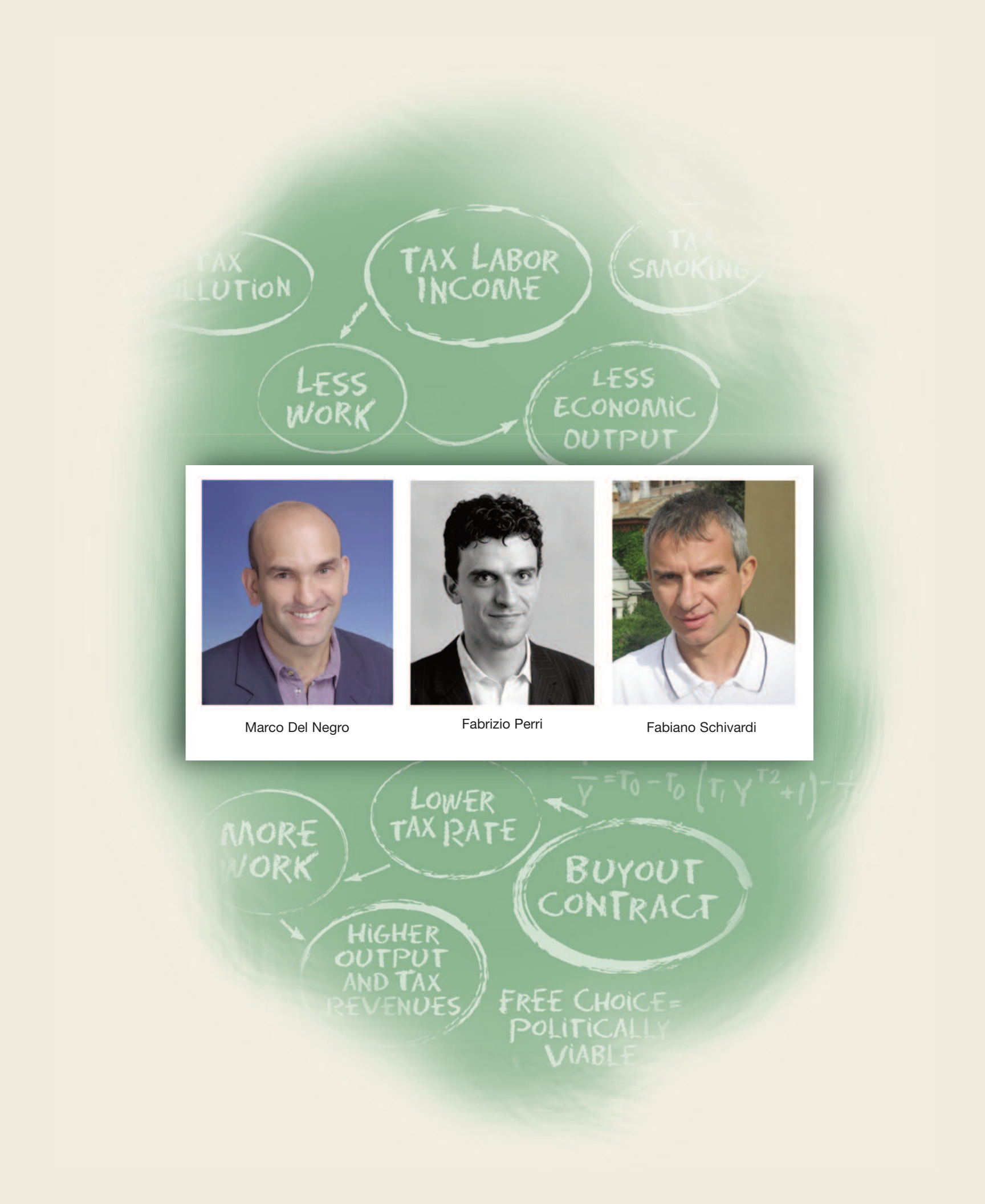
Economic Policy Papers *are based on policy-oriented research by Minneapolis Fed economists, officers and other staff. The papers are an occasional series of publications written for a general audience.*

ABSTRACT

Due to cyclical and structural factors, including the fiscal response to the 2007-2009 recession, and rising Medicare, Medicaid and Social Security expenditures, the U.S. government is facing unprecedented levels of spending, now and for years to come. To maintain fiscal sustainability, revenue levels must increase, but doing so through higher taxes depresses economic activity and is politically difficult.

This paper proposes a fiscal instrument, which we call a "tax buyout," that would allow the government to raise at least part of the needed revenues in a politically viable way, and without stifling economic activity.

The essence of a tax buyout is to offer citizens the choice to decrease their tax rate for a period of time in exchange for a fixed payment. The tax rate reduction avoids the inhibition of economic activity, the fixed payment allows government to raise revenues, and presenting the buyout as a choice ensures political feasibility. Our initial estimates indicate that a well-designed tax buyout program would have significant quantitative importance in a national economy. Prior to public implementation, however, a number of concerns and possibilities for improvement should be considered.



Marco Del Negro

Fabrizio Perri

Fabiano Schivardi

reduction in his or her marginal tax rate for a given period (say, one year).

We call it a "buyout" because it allows individuals who purchase the contract to effectively pay off a percentage of their regular (and distortionary) taxes with a lump-sum payment to the tax collection authority. Participation is voluntary and involves no risk from the individual citizen's point of view: Only those who would gain from entering the contract in a given period (after any uncertainty about their labor income is resolved) will do so.

This paper begins by discussing the distortion problem addressed by this plan, including background from related research. It then describes the model we developed to analyze how a tax buyout program would work in a dynamic macroeconomy, including quantitative estimates of the impact such a program might have on the U.S. economy during a time of high fiscal pressures such as those now present. We conclude with suggestions of further issues that should be addressed to make tax buyouts a concrete policy option—an important goal in a period of substantial and growing fiscal deficits.

The views expressed here are ours, and not necessarily those of others in the Federal Reserve System.

Background and description of tax buyouts

The idea of a tax buyout focuses on an issue that is central to economic analysis: the disincentive effect of taxation. Taxes are sometimes imposed on activities that society wishes to discourage, such as smoking or pollution; in such cases, the disincentive is intentional. But when a government seeks to generate revenue by imposing taxes on a worker's earnings, it spites itself. A tax on labor income discourages work because the worker knows that each hour of labor will generate less take-home pay. The resulting decrease in work effort leads to less economic output, which in turn leads to a lower tax base—undercutting the revenue generation that is the very goal of imposing such taxes. That distortion of economic incentives is a key challenge to tax policy, and to economic research as well.

In an effort to understand how to design a better tax system, British economist and Nobel laureate James Mirrlees analyzed these "labour-discouraging effects" in a classic 1971 paper, and much of modern "optimal fiscal policy" research—including ours—is based on his model.[3] Mirrlees recognized that any labor tax system must cope with "asymmetric" information: A citizen knows more about his or her ability to work than does the government. Given that asymmetry, how do policymakers design a tax system that provides maximum incentive to work and minimal distortion to labor supply, economic growth and revenue generation? Mirrlees' solution was to design a system that is "incentive compatible," meaning that it must give workers a pecuniary incentive to reveal their true work abilities—that is, it must be designed such that workers' self-interest will induce them to provide as much labor as they can.

That, indeed, is the intent of our buyout proposal: To establish a tax scheme that draws forth as much work effort as possible by offering individuals the chance to purchase a "buyout" contract that decreases their marginal tax rate. And our challenge is to determine whether there is, in fact, a contract price low enough to draw people into the program, but high enough to generate sufficient revenue to fund the buyout scheme *and* other government expenditures. Our analysis suggests that such a program is not only feasible, but also quantitatively significant and politically viable.

A tax on labor income discourages work because the worker knows that each hour of labor will generate less take-home pay. The resulting decrease in work effort leads to less economic output, which in turn leads to a lower tax base—undercutting the revenue generation that is the very goal of imposing such taxes.

It should be noted that other economists have suggested related ideas. During a milder economic downturn in the early 1990s, Harvard economists Alberto Alesina and Philippe Weil proposed a two- (or more) tiered tax schedule under which taxpayers could purchase a lower marginal tax rate.[4] "The tax payers who select to 'buy' the reduction in the

marginal rate, i.e., who choose the new tax schedule, will be the most productive workers: under the new tax schedule they will work and consume more. … [The] introduction of the second tax schedule does not reduce total tax revenues. More generally, additional revenue-neutral or revenue-increasing Pareto improvements can be achieved."

Similarly, in a 1994 paper, the University of Michigan's Joel Slemrod and his co-authors investigated a two-bracket income tax structure and found that "a second tax bracket allows the lower marginal tax rate on high-wage people to coax out … greater labor supply from the most productive segment of society, with the increased tax revenue used to lower the tax burden of the least productive segment. Although the calculated optimal tax system features declining marginal tax rates, it still generally features increasing average tax rates, so that it is progressive but not graduated, in the standard sense of these terms."[5]

This research provides important background but does not include several elements that are potentially important to fully evaluate the impact of these schemes. These previous models are static—analyzing economies at just one point in time—and the people acting in these models are essentially identical to one another in every way but work ability. Our research extends this idea into an economy that is dynamic (it evolves over time) and incorporates "heterogeneous agents" (meaning that people in our model vary substantially in numerous characteristics relevant to labor supply, income and taxes). In addition, we look at the broad macroeconomy and also at the idea of tax schedules that are "nonlinear" (tax rates for different income brackets can vary dramatically—tax rate graphs are curves, not straight lines). We believe that this appraisal renders our buyout scheme a pragmatic proposal that, with refinement, could be used to address current challenges in fiscal policy.

Step-by-step analysis

We first conduct an abstract exercise with a mathematical model of a national macroeconomy to see if the tax buyout idea is sensible at a theoretical level. Do the basic relationships among critical variables in our model economy—tax rates, labor supply, consumption levels and the like—result in the buyout plan inducing enough extra labor, and therefore extra output and tax revenue, to more than pay for itself? The answer is yes.

While our model is designed to represent crucial economic incentives and relationships, like all such models, it abstracts from reality in a number of respects. Nonetheless, in our analysis we take a step-by-step approach to incorporating increasing levels of realism into the model and at each step evaluate whether the important result of a tax buyout drawing forth additional work effort can be achieved.

We begin with a very basic model: an economy that examines static relationships between individuals and government with a constant tax rate and perfect information. By "perfect" information, we mean that the work capability of every individual is public knowledge: Tax collectors know how much work everyone is able to perform, so pretending to be disabled to avoid work and collect government insurance benefits isn't an option. In this (unrealistic) case, our model demonstrates that the offer of a contract to reduce an individual's tax rate in exchange for a set contract payment will be accepted by everyone in the population, will increase total labor and well-being, and will leave government revenues unchanged.

To understand the logic of this result, consider a simplified example. There are two workers: Alice and Ben. Alice earns a high labor income, while Ben earns a low one. Both pay taxes, and the more they earn, the more taxes they pay. If the government knows their ability exactly, it can offer them a tax buyout contract that involves Alice and Ben paying a fixed amount (high for Alice and low for Ben) in exchange for a reduction of their tax rate to zero. If the fixed amounts are chosen equal to the pre-buyout tax receipts, the government will not lose money from the contract.

But what incentive do Alice and Ben have to buy into this program?

The key difference between standard taxation and the tax buyout program is that in the tax buyout, the government asks for a *fixed* amount. So if Alice or Ben works an extra unit (another hour, day or week, say), she or he is the sole beneficiary of the extra revenue—they no longer have to pay a portion of it to the government as under standard taxation. This increases their incentive to work and

thus will increase their income and, ultimately, their well-being.

In other words, by lowering an individual's tax rate, the tax buyout removes what Mirrlees called the "labour-discouraging effect" of labor income taxes, thereby eliminating the inefficiency due to distortionary taxation. That releases a surplus that can then be shared by both government and individuals.

We then extend this to the more realistic scenario considered by Mirrlees in which information about work ability is imperfect, or "asymmetric": The government does *not* know how much work every person is capable of. Can a tax buyout program create the incentive compatibility that Mirrlees showed is necessary?

Going back to the Alice and Ben example, now the government does not know which one is the more productive worker and thus cannot offer a tailored contract (a high price for Alice and a low one for Ben), but instead must offer a single contract. In this case, we find that the buyout contracts can nonetheless be priced at a level high enough to generate positive revenues for the government, but low enough to attract enough individuals to buy them.

In the Alice and Ben example, if the government offers the buyout at the price equal to the pretax liabilities of Alice, then Alice will take the buyout (and this will increase her labor effort and well-being), while Ben will typically not take it, and so his welfare will be unaffected. Still, the buyout is socially desirable because part of the population gains, while another part does not lose.

Some might contend that because Ben, a poor person, is unaffected, while Alice, who is rich, is gaining from the buyout, the program could increase the gap between rich and poor—an arguably unfair outcome. In the paper, we argue that it is possible to construct buyout schemes in which all people, including poor workers, can be made better off by the introduction of the buyout, even when they don't participate in it directly. The idea again is that the contract generates a surplus that can be shared. With a properly designed buyout plan, the government can receive and redistribute some of this additional surplus so that the entire population benefits, not just the most productive.

By lowering an individual's tax rate, the tax buyout removes ... the "labour-discouraging effect" of labor income taxes, thereby eliminating the inefficiency due to distortionary taxation. That releases a surplus that can then be shared by both government and individuals.

Real-world relevance?

The step-by-step analytical modeling demonstrates that the tax buyout idea has substantial theoretical merit. But that leaves aside the issue of quantitative importance. That is, given actual levels and distributions of economic and demographic variables (such as household earnings, wealth levels, tax and interest rates, life span and retirement length), would a tax buyout program have any real dollars-and-cents impact on a multitrillion-dollar economy? Or is this merely an interesting academic proposition without practical application?

To answer this question, we write a more detailed artificial model economy with overlapping generations of heterogeneous (in terms of abilities and luck) households who make labor decisions, consume and accumulate wealth over their lifetimes. We then put this model through a process called "calibration"—essentially, setting the model's parameters so that its basic predictions capture aspects of actual U.S. households that we think are crucial for our policy experiment.

In particular, we calibrate our model to ensure that

(1) households in the model have the same wealth and labor earnings distribution as households in actual U.S. data for 2006, and

(2) the shape of the tax function (i.e., the equation that assigns a household's tax liability as a function of its total earnings and family composition) is consistent with actual U.S. tax code.[6]

A key parameter for our model economy is the so-called Frisch elasticity of labor supply, a measure

of how much workers change their labor supply in response to a change in wages (or taxes), keeping everything else (including their wealth) constant.

This parameter is crucial for our question because if workers are not very responsive to wage or tax changes, then taxes are not very distortionary—that is, tax increases or decreases hardly affect overall welfare and labor supply. If that's the case, the tax buyout, which operates through reduction of distortions, will not yield large benefits.

A very large literature in economics has tried to estimate Frisch elasticity, but economists are still uncertain. In our work, we start by considering a value that lies in the middle range of existing estimates, but we also experiment with different values.[7]

Generating answers

After this calibration process, we run the model through many computer simulations to generate numerical answers for the questions we're interested in:

- What percentage of people will purchase a buyout contract at a given price for a specified reduction in their tax rate?
- What effect will that have on the hours of work they supply?
- How will that affect government tax revenue?
- To what degree will this change in labor supply (through a reduction in tax distortion) alter the nation's economic output?

Our strategy is to consider an economy with a set level of government spending and no tax buyout plan (for example, the U.S. economy before the recent financial crisis), which then unexpectedly faces a 20 percent jump in public expenditures, due, say, to a financial sector bailout or sharply higher Medicare costs (the U.S. economy post-crisis). We then consider two scenarios: one without the buyout offer and one with it.

In particular, we consider the following buyout option: Each citizen has the option of reducing his or her labor income taxes by 5 percent for one year by paying the government the fixed price of $4,500. The contract is very simple to understand and to accept or reject.

An example may help to make the option more concrete. Consider again our friends Alice and Ben. We'll assume that Alice, the more productive worker, earns a labor income of $100,000, while Ben's labor income is $30,000. At the time of filing her taxes, Alice would find it advantageous to accept the buyout because her take-home pay will be $500 higher. In contrast, Ben will *not* buy the contract because doing so would actually reduce his take-home pay by $3,000.[8]

Note that accepting or rejecting the buyout would not involve any additional risk for either Ben or Alice (the decision is taken at the time of filing taxes), but the *essential* element is that Ben and Alice know that the buyout is an option at the beginning of the year, when they decide how much to work. Notice that if Alice knows of the buyout option, she will in general work harder, because she can retain more of the additional dollars she earns, and her additional work is the key social and private benefit of the buyout.

In both cases (with and without buyout), we assume that the government will raise taxes to finance the additional expenditures so that the budget is balanced in every period. By comparing those scenarios, we can judge the quantitative impact of a variety of buyout plans. And because we use a dynamic model, we're able to estimate results over a span of 20 years.

Quantitative results

In our first experimental run-through, we find that in the scenario without the buyout, taxes as a fraction of total income need to rise (in order to balance the budget) from roughly 21 percent to 26 percent. With the tax buyout option, however, taxes would rise to just 24.5 percent.

Given that government expenditures are identical in both scenarios, why would buyouts result in lower average taxes? Because, according to the model, over 8 percent of the population will purchase the buyout contracts, thereby generating additional government revenue. This transformation of part of government revenues from a tax that distorts labor decisions to a lump-sum payment that does not is the essence of the tax buyout contract. And it does so in a revenue-neutral fashion without making anyone worse off.

The reduction in work-supply distortion—the decrease in what Mirrlees called the "labour-dis-

couraging effects" of income taxes—is quantitatively important. Labor supply with the tax buyout scheme is 0.33 percent higher than without it because those who buy the contracts choose to work harder (since their marginal tax rate is lower). Moreover, those people tend to be the most industrious workers, so there is an increase in average labor productivity. Therefore, while national economic output (or alternatively, national income) drops because taxes had to increase to fund higher government spending, it drops less with the buyout program, about 1 percent less. Due to higher overall taxes, wealth and consumption decrease, but the decrease is less severe with buyout contracts.

Changing assumptions

We then run the model under a few different scenarios, changing the size of the buyout, making its price age-dependent and altering the estimate of worker responsiveness to wage changes. The table below shows the results, compared with the results in the baseline scenario, reported in the first row.

As the table indicates, increasing the tax buyout size (or, alternatively, the tax rate reduction) from 5 percent to 10 percent (column 1, rows 1 and 2) means nearly a tripling in price (from $4,500 to $12,900) and half as many buyers. As expected, reducing the size (row 3) lowers the contract price and increases program participation. The larger buyout scenario still has a significant impact on GDP; the smaller buyout less so.

Interestingly, if the price of the buyout contract is varied according to the purchaser's age (row 4), similar to life insurance pricing, it will attract more buyers and generate a bit more revenue. This is because older people have higher wages on average, would benefit more from the reduction of distortion provided by the tax buyout and, hence, are willing to pay a higher price.

As discussed previously, a crucial parameter for evaluating the effectiveness of the buyout is the Frisch elasticity of labor supply. In the table's last row (5), we show results when we consider a low elasticity value. In this more conservative case, the benefits of the buyout are smaller than in the baseline case but remain significant, with gains in GDP exceeding half of 1 percent.

Finally, we looked at how things change over time to get a sense of which types of people are most likely to buy the contract, not just now but in the future. This is one of the clear advantages of using a dynamic rather than a static model. One way of looking at a tax buyout is that it's an opportunity to buy, for a fixed price, a subsidy on one's labor income. And because the subsidy is calculated as a percentage of income, the benefits are greatest for those who earn—or *expect* to earn—high labor income. Bottom line: The people most

Tax Buyout Scenarios

		Buyout size (reduction in marginal tax rate)	Buyout price	Contract buyers as a percentage of taxpayers	Percentage of total tax revenue from buyout contracts	Gain in GDP
		(1)	(2)	(3)	(4)	(5)
Baseline scenario	(1)	5%	$4,500	8.2%	4%	0.95%
Larger buyout	(2)	10%	$12,900	4%	5.6%	0.8%
Smaller buyout	(3)	2%	$1,300	14.5%	2%	0.6%
Age-dependent pricing	(4)	5%	Increases with age	10.1%	4.2%	1.1%
Lower labor elasticity	(5)	5%	$5,100	6%	3.3%	0.55%

likely to buy the buyout contract now or in the future are

- high-wage (and therefore older) people
- people who are patient (because they value the possibility of earning a lot in the future) and
- people with little wealth (because lower wealth induces individuals to work harder).

Our computer simulations find considerable differences over time among people. The types of individuals just listed would significantly benefit from introduction of a tax buyout program even when they don't participate in it initially. An obvious example is young people: Even if they are not buying into the contract now, they will probably earn higher labor income when they're older and therefore be more likely to participate. The program's existence, and the possibility of (literally) buying into it in the future, is highly valued. Thus, in a dynamic world evaluated over the long run, the benefits of a tax buyout program spread well beyond the fraction of people who participate in it at any single point in time.

Further work needed

Before a buyout program is designed and implemented, a number of concerns call for further investigation. By the same token, several promising possibilities could lead to significant improvements in buyout strategy.

The first concern is what economists call a "general equilibrium effect." One consequence of introducing a tax buyout program is that prices (in particular, wages) will change, and perhaps in a direction that is disadvantageous for some. Specifically, the tax buyout's reduction in incentive distortion will result in a labor supply increase. That could reduce wage levels in general and hurt in particular the low-wage, low-productivity people who are least likely to buy the contract. This effect deserves quantitative investigation because its impact likely depends on factors not considered in our model, such as the openness of capital and labor markets.

Another concern arises in regard to the distribution of high and low ability within the total population. The issues here are complex, but they come down to two basic questions: Would the program benefit only high earners, rendering it socially less desirable and politically unpalatable? As we discussed earlier, a possible solution to this issue is to accompany the buyout program with a redistribution policy (financed by the buyout itself) to assist low earners.

The benefits of a tax buyout program spread well beyond the fraction of people who participate in it at any single point in time. A number of concerns call for further investigation. By the same token, several promising possibilities could lead to significant improvements.

And secondly, are there so many high-labor-income people in the population, or people of such high labor income, that offering them the chance to lower their tax bill would significantly undercut general tax revenues? Future research should therefore investigate the benefits of limited buyouts, in which a person's gain from tax reduction is limited to a specified multiple of the contract price. For instance, what if the tax benefits for a buyout contract were limited to, say, twice the contract purchase price? What labor supply, tax revenue and GDP impact would such a program have?

On the more encouraging side, there are many directions in which this buyout idea could be extended to reduce labor effort distortions still further. For example, varying the contract pricing schedule for individuals of high and low work ability could have a beneficial impact. Another possibility: In our current setup, we assume completely asymmetric information, meaning that the government knows essentially nothing about individuals' work abilities. In reality, of course, the government knows quite a bit about its citizens—education levels and earning history, for example—and could alter contract prices accordingly.

Third, it seems likely that labor supply elasticity—again, sensitivity to changes in wage levels—differs among individuals: Some people will

respond more than others to a $5 wage hike. In our quantitative experiments, we plug in just one value for the entire population, but in fact, people with high elasticity would be more likely to buy tax-reducing contracts, leading to higher program participation. And lastly, the tax buyout idea could be expanded to capital income—stock dividends, for instance—and further analysis should estimate the combined effects of buyout programs offered for both labor and capital income.

Conclusion

We believe that a tax buyout initiative is a promising means of addressing likely revenue shortfalls in the United States. By offering citizens the opportunity to decrease their marginal tax rate in return for a fixed payment, governments could reduce the negative impact that labor income taxes have on labor supply decisions, thereby increasing total work effort, raising overall economic output and well-being, and generating higher tax revenues.

Our initial analyses suggest that tax buyout programs can have significant quantitative importance in a national economy, especially at a time when high fiscal needs call for high levels of distortionary taxation. Prior to designing such a program for public implementation, a number of concerns should be addressed and several possibilities for improvement considered. Also, the effects and consequences of such a scheme could be evaluated with alternative methods, for example, by running small-scale experiments such as introducing the buyout for local and state taxes in small communities. R

Endnotes

[1] This policy paper is based on: Del Negro, Marco, Fabrizio Perri and Fabiano Schivardi. 2010. "Tax Buyouts." Research Department Staff Report 441, Federal Reserve Bank of Minneapolis. The authors thank Doug Clement for many insightful comments and excellent editorial assistance.

[2] As noted by Federal Reserve Chairman Ben Bernanke in recent congressional testimony, "[I]n the absence of further policy actions, the federal budget appears to be on an unsustainable path. A variety of projections that extrapolate current policies and make plausible assumptions about the future evolution of the economy show a structural budget gap that is both large relative to the size of the economy and increasing over time. … To avoid sharp, disruptive shifts in spending programs and tax policies in the future, and to retain the confidence of the public and the markets, we should be planning now how we will meet these looming budgetary challenges." Statement by Ben S. Bernanke, chairman, Board of Governors of the Federal Reserve System, before the Committee on the Budget, U.S. House of Representatives, June 9, 2010.

[3] Mirrlees, James A. 1971. "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies* 38(2), pp. 175-208.

[4] Alesina, Alberto, and Philippe Weil. 1992. "Menus of Linear Income Tax Schedules." NBER Working Paper 3968.

[5] Slemrod, Joel, Shlomo Yitzhaki, Joram Mayshar and Michael Lundholm. 1994. "The Optimal Two-Bracket Linear Income Tax." *Journal of Public Economics* 53, pp. 269-90.

[6] The distribution of earnings and of wealth for U.S. households is computed using the most recent waves of two widely used economic surveys: the 2007 Current Population Survey and the 2007 Survey of Consumer Finance. For further details, see the original paper.

[7] In particular, we consider a value of the Frisch elasticity of 0.75, which implies that on average a worker who faces, say, a 10 percent reduction in wages while keeping his or her total resources constant would reduce his or her labor supply by 7.5 percent.

[8] The savings from the tax buyout are 5 percent of labor income. For Alice, this is 5 percent of $100,000, a $5,000 savings that exceeds the buyout contract cost of $4,500 by $500. She'll take home an extra $500 from choosing the buyout. For Ben, however, the buyout would yield a saving equal to 5 percent of $30,000, i.e., $1,500 that falls $3,000 short of the $4,500 contract cost.

BOOK REVIEW

BOOK REVIEW

This Time Is Different

Eight Centuries of Financial Folly

By **Carmen M. Reinhart** and **Kenneth S. Rogoff**
Princeton University Press
463 pages

Reviewed by **Kevin L. Kliesen**
Business Economist
Federal Reserve Bank of St. Louis

> *Nobody can hope to understand the economic phenomena of any, including the present, epoch who has not an adequate command of historical facts and an adequate amount of historical sense or what may be described as historical experience.*
> —Joseph Schumpeter[1]

Toward the end of his life, Harvard economist Joseph Schumpeter remarked that of the three building blocks of economics—theory, statistics and history—economic history "is by far the most important."[2] The importance of economic history is on grand display in *This Time Is Different: Eight Centuries of Financial Folly*. In their book, Carmen Reinhart (University of Maryland) and Kenneth Rogoff (Harvard University) convincingly remind us that economic crises are recurring events. (See an interview with Rogoff in the December 2008 *Region,* online at minneapolisfed.org.) This fact naturally leads to two important conclusions: There will be more in the future and, accordingly, financial reform legislation will not prevent future crises. But Reinhart and Rogoff also remind us, in a way that Schumpeter would no doubt appreciate, that economic policymakers repeatedly fail to fully grasp one of the key lessons of history: Economic misfortune falls upon countries that fail to heed the consequences of excessive debt accumulation.

Overview

Building on a historical narrative that uses an extensive data set of their construction, Reinhart and Rogoff (hereafter R&R) show that periods of excessive public debt accumulation generally do not end well. Over time, many countries have defaulted on their debt (including restructuring) for a variety of reasons and by a variety of methods (inflating away the real value of the debt has been very popular). These defaults, they show, can produce detrimental spillover effects. Recent defaults by Russia (1998) and Argentina (2001) come to mind, and the possibility of a future restructuring by Greece looms large for its foreign creditors (for example, European banks)—and for European policymakers.

One drawback of R&R's analysis, which they readily admit, is that it focuses almost entirely on debt issued by governments, or sovereigns, rather than by the private sector. In the financial crisis of 2007-09, which they term the "Second Great Contraction," the accumulation of private debt (chiefly mortgage debt of the dodgy variety) and the collapse in nominal house prices eventually helped trigger a banking and financial crisis of immense proportions and a collapse in economic activity. In response, federal government outlays in the United States and other advanced economies rose enormously, which resulted in huge budget deficits that have significantly boosted debt-to-GDP levels.

Since emerging and developing countries tend to rely heavily on foreign creditors such as large multinational banks, sharply higher debt-to-GDP ratios in the context of weakening economic fundamentals can lead to "sudden stops"—that is, credit is withdrawn abruptly, leading to a cascade of defaults. In advanced economies, which have better credit and inflation histories, and thus sharply lower probabilities of default, rising debt-to-GDP ratios tend to weaken economic growth.[3]

Research Digest

The Region often includes one or two feature articles about economists at the Minneapolis Fed and their current work. Research Digest provides shorter summaries of recent economic research papers.

In this issue, the Digest discusses Erzo Luttmer's efforts to explain employment growth patterns of U.S. companies by merging two competing theories, and a paper by Veronica Guerrieri and her coauthor on the link between asset price volatility and fund managers' career concerns.

Explaining Growth

Economist Erzo Luttmer blends two competing theories to generate a model that helps account for patterns of employment growth in U.S. companies.

PHOTOGRAPH BY DONG HO

Growth patterns—of plant and animal species, but also of cities, companies and nations—have long fascinated natural and social scientists. Economists are no exception. If economists could grasp the essential mechanisms that explain widespread patterns in growth data,[1] they could better understand how and why economies grow as they do. (And in the current era of high unemployment and anemic economic growth, such explanations would be of particular value.) In a recent paper, Minneapolis Fed consultant Erzo G. J. Luttmer delves into this long-term puzzle, with promising if still tentative results.

Luttmer's focus is companies—how many employees they have, how quickly those employment numbers grow and most important: Why? In "On the Mechanics of Firm Growth" (Staff Report 440, online at minneapolisfed.org; also forthcoming in *Review of Economic Studies*), he begins with several observations and questions:

> More than half of all [U.S.] firms … have no more than four employees. But there are also almost a thousand firms with more than ten thousand employees each and these firms employ as much as a quarter of the

Research Digest

Firms grow their organization capital by having employees work with pieces of the existing capital to create still more capital. Luttmer refers to this as "blueprints" that can be used to create new capital. "Starbucks implementing its store formula in many places would be a good example that fits my model well."

U.S. labor force. What accounts for the large amount of heterogeneity in firm size? How does this heterogeneity evolve over time? Some benchmark answers to these questions are needed.

Two prominent theories offer competing explanations for firm employment growth patterns, notes Luttmer, but both fail to match the facts well enough. The first idea is that the skewed distribution of firm size—where a very small number of firms employ a very large percentage of all employees[2]—is the result of big (and randomly distributed) differences among firms in *productivity growth*; some firms become more efficient over time than others in generating output with given levels of inputs.

A second potential explanation is that skewed firm size distribution results from random distribution of *organization capital* (a term coined by economists Edward Prescott and Michael Visscher in 1980), meaning that companies are defined by their accumulated information: ideas or methods developed by a firm. Firms grow their organization capital by having employees work with pieces of the existing capital to create still more capital. Luttmer refers to this as "blueprints" that can be used to create new capital. "Starbucks implementing its store formula in many places would be a good example that fits my model well," he explains.

Theory fusion

While both theories have been studied carefully by economists, neither is entirely satisfactory. So Luttmer blends the two in a hybrid model that goes a long way toward explaining firm growth patterns as seen in U.S. data.

His model starts with organization capital theory as its base. "In the model," Luttmer writes, "a firm produces one or more differentiated commodities using labor and commodity-specific blueprints." In other words, by using employees and organization capital. A new firm is born when an entrepreneur produces a "start-up blueprint." Then this new firm can hire more workers, combine them with blueprints and seek to develop more blueprints for still new commodities. That is, it can attempt to grow. Over time, of course, blueprints can become obsolete, as the information they represent is superseded by blueprints held by other firms. (Say, for instance, a new search engine surpasses Google's.)

This theory—based on blueprints, or organization capital—produces results that match actual U.S. size distribution of firms, but it predicts far too high an average age for big firms: about 750 years rather than the 75 years seen in actual 2008 U.S. data for companies with over 10,000 employees.

How can the theory be modified to account for the relatively young age of large firms seen in the data? By supplementing it with productivity shocks.

"Suppose," writes Luttmer, "that some new firms enter with an initial blueprint of higher quality." That is, they receive a random productivity shock. The higher profits that result from the high-quality blueprint encourage the firm's managers to copy it rapidly, and "if copies stay within the firm, then these new firms will grow fast." The

Research Digest

growth rate will eventually decline (assuming that the quality advantage is transitory), but if the rapid growth period varies by company, and if it isn't expected to last too long, this variation allows for the appearance of large firms that are young—that is, a median age of about 75 years, a match to empirical reality.

This version of organization capital theory, infused with elements of productivity theory, "can match the overall size distribution, the amount of [firm] entry and exit, as well as the relatively young age of large firms," writes Luttmer. It doesn't hold strictly to Gibrat's law—that firm growth rates are independent of firm size—but "the mean growth rates of surviving firms behave [as] in the data: roughly independent of size for most firms and significantly higher for the smallest firms."

It is, on the whole, a strong step toward a better understanding of the mystery of why companies—and economies—grow as they do.

—Douglas Clement

Organization capital theory, infused with elements of productivity theory, "can match the overall size distribution, the amount of [firm] entry and exit, as well as the relatively young age of large firms," writes Luttmer.

[1] Two examples as applied to firms: *Gibrat's law*, a firm's growth rate is independent of its size; and the *rank-size rule*, if you take all the firms in an economy and rank them top-to-bottom by employment, the second-largest firm will have half the employee count of the largest, firm no. 3 will have a third as many as no. 1 and so on. *Zipf's law* states the rank-size relationship in terms of probability—the likelihood that a firm has a size greater than S is proportional to $1/S$.

[2] U.S. firm employment data reliably exhibit this striking, skewed pattern, referred to as a Pareto distribution. Italian economist Vilfredo Pareto noted in 1906 that income, city populations and even peas in pods exhibit similar distributions. He documented, for example, that 20 percent of Italy's population held 80 percent of the nation's wealth. Zipf's law (see footnote 1) is a specific type of Pareto distribution.

Research Digest

PHOTOGRAPH BY BETH ROONEY

Asset Bubbles and Rat Races

A new Minneapolis Fed staff report by Veronica Guerrieri and her coauthor examines the effect of portfolio managers' reputation-seeking on asset prices.

During the financial crisis of 2007-09, macroeconomists came under fire for ignoring the importance of financial markets and particularly the role of financial intermediaries, such as investment banks and portfolio fund managers, in fueling asset price bubbles.

Recent Minneapolis Fed research by Veronica Guerrieri of the University of Chicago and Péter Kondor of the Central European University ("Fund Managers, Career Concerns, and Asset Price Volatility," SR 446 online at minneapolisfed.org) focuses on this relationship and suggests that financial professionals' concern about their reputations and careers plays a direct role in asset price volatility.

The authors start with two observations about financial markets. First, the risk premium—the higher average return on risky securities like junk bonds compared with risk-free assets like government bonds—increases during recessions and decreases in economic upswings. Second, investors often don't handle their portfolios themselves, but hire fund managers to do the job for them. In essence, Guerrieri and Kondor put forward a theory that makes use of the second fact to explain the first.

They start with a model in which investors can park their money in either a risk-free asset that pays a low but guaranteed rate or invest it in a risky bond that might pay a higher return but also might end up worthless if the bond issuer defaults. Investors outsource this decision to fund managers, whose pay is based directly on the portfolio's return.

The model gets interesting when some managers know more than others. The authors assume some managers know for sure whether the risky bond will default, but

Research Digest

It is therefore in every fund manager's interest to maintain a reputation—deserved or not—for being savvy about asset quality. The economists' model demonstrates that such career concerns "distort their investment decisions and magnify asset price volatility." Guerrieri and Kondor call this price distortion a "reputational premium."

others know only the probability that it might. This is akin to knowing a coin will land heads-up versus only knowing that you have a 50-50 shot.

Naturally, investors want to hire better-informed managers, not the less-informed, but they can't tell the difference beforehand. So at the end of every period, investors compare their manager's performance to the best manager's and attribute returns to the skill of their manager. If their manager was too exuberant and put their money in risky bonds that defaulted, or played it too safe and missed out on the relatively higher average returns, investors will fire the old manager and hunt for a new one.

It is therefore in every fund manager's interest to maintain a reputation—deserved or not—for being savvy about asset quality. The economists' model demonstrates that such career concerns "distort their investment decisions and magnify asset price volatility." Guerrieri and Kondor call this price distortion a "reputational premium."

Here's how it works: In financial recessions, default risk is high. To compensate uninformed managers for investing in risky assets—because their reputations will be damaged if the assets default—the reputational premium is positive, so the return on such assets has to be high. (And by definition, assets with high returns are those with low prices.)[1]

During boom times, the opposite occurs: Default risk is low, so the "reputational premium" is low, and even negative. Smaller returns are required to induce fund managers to buy assets, so managers tend to buy higher-priced assets than they would in the absence of career concerns. Thus, the model replicates the countercyclical risk premium, and procyclical price movements, seen in the real world. That is, during a recession, the risk premium rises and asset prices fall, and vice versa during an economic boom: The premium falls and prices rise.

Indeed, the economists illustrate this with some empirical observations from recent financial swings. "Our model suggests," they write of the dot-com bubble, "that hedge funds were willing to buy technological stocks at highly inflated prices because of their fear of losing reputation and hence funds if they missed the high returns generated by the bubble. This is consistent with the additional fact … that the largest hedge fund, Tiger Fund, which refused to invest in technology stocks, experienced severe fund outflows in 1999 compared to its main competitor who did invest in technology stocks, Quantum Fund."

Guerrieri and Kondor expand their model to incorporate the tendency for high risk of default today to imply high risk tomorrow, and likewise for low risks. This "persistent default risk" makes asset prices even more volatile, but it adds a second effect as well.

With persistent risk, a smaller share of uninformed managers keep their jobs in high-risk times, so future prices will reveal more. This makes the cost of getting fired greater, which increases the reputational premium. The price of the risky bond can change even if the actual probability of default doesn't change, indicating that some movements in asset prices

Research Digest

are not driven by fundamentals.

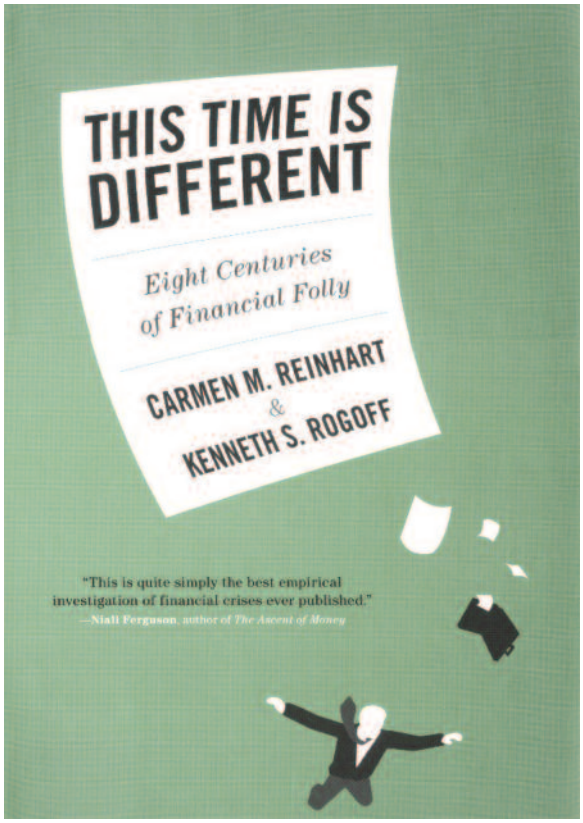
To reiterate, this phenomenon can help account for financial crises. In periods such as the tech stock boom of the 1990s and the housing boom of the last decade, the yield spreads between high-risk and low-risk securities tend to drop very low and then skyrocket with the onset of crisis. The model explains that feature in terms of reputational effects.

While far from the only theory of excess asset price volatility or risk premium swings, this new research is the first to explain these phenomena in terms of fund managers' career concerns. And given the list of intriguing extensions and applications the authors discuss, it won't be the last.

—Joe Mahon

[1] An asset's return is the financial benefit it yields compared with its price: Return=Yield/Price. So, the lower the price, the higher the return, and vice versa.

With persistent risk, a smaller share of uninformed managers keep their jobs in high-risk times, so future prices will reveal more. This makes the cost of getting fired greater, which increases the reputational premium. The price of the risky bond can change even if the actual probability of default doesn't change, indicating that some movements in asset prices are not driven by fundamentals.



As the world's largest economies return to their long-run rates of economic growth (relative to the growth downturn of the recent recession), their governments will increasingly be forced to undertake fiscal readjustments. The European sovereign debt crisis in the first half of 2010 has demonstrated that the retribution for inaction can be swift—plummeting currencies and sovereign credit ratings. The debate at present, though, is whether this adjustment should occur before or after these countries have returned to their normal growth paths. R&R remind us that returning to this growth path after financial crises is often a protracted affair.

Outline of the book

The book comprises six parts; broadly speaking, the first three are historical, the last three topical. And the six parts are further divided into a total of 17 chapters.

Part I provides a historical overview of the various financial crises discussed in the book: banking and currency crises, external debt default and inflation crises. To give readers a sense of the historical significance of the "this time is different syndrome," chapter 1 discusses five well-known episodes over the past 75 years that most readers will be familiar with: (1) the buildup to the emerging market defaults of the 1930s, (2) the debt crisis of the 1980s, (3) the debt crisis of the 1990s in Asia, (4) the debt crisis of the 1990s and early 2000s in Latin America and (5) the financial crises prior to the Second Great Contraction in the United States.

A common thread running through this discussion: hubris. Households, governments, institutions, financial market participants, economists and businesses (have I missed anyone?) consistently underestimate the fragility of highly leveraged economies. Why? Innovations or improvements in the real economy or in public policies lead many to conclude that a boom is based on solid fundamentals. In the mid to late 1990s, for example, an acceleration in productivity growth, financial innovations and new risk management tools were thought by many to explain the rapid rise in prices of financial and tangible assets (the so-called tech boom).[4]

Not surprisingly, R&R show that serial default is a common occurrence among many of the world's less-developed countries—although more-advanced economies also default on their debts. R&R tell us that Greece, the most recent poster child for fiscal malfeasance, was in "continual default" from 1800 to just after World War II. We also learn that, perhaps surprisingly, default tends to occur at debt levels well below the Maastricht Treaty threshold (60 percent of GDP).[5] For middle-income countries from 1970 to 2008, more than half of all defaults occurred at debt-to-GDP levels below 60 percent.

This raises an interesting question: What is a safe debt threshold? R&R argue that a nation's safe debt threshold depends heavily on its historical record of defaults (if any) and its past inflation performance. Typically, only countries with good repayment and inflation histories are able to regularly access global capital markets.

Part I also contains one of the most impressive aspects of this book—its historical data set. The authors focus primarily on 66 countries that accounted for about 90 percent of global GDP in

1990. Although the book's title suggests that the analysis covers the past 800 years, and some of the data extend back to the 13th century, the core analysis is generally based on data from 1800 to the present.

A key to the book's success—besides its accessibility to the noneconomist—is the use of this data set. For instance, in parts II and III, R&R discuss crises associated with default on government debt held by foreign purchasers (external debt) and domestic purchasers (domestic debt). The level of detail is impressive. One of the more interesting chapters (chapter 4) discusses the theoretical aspects of debt crises and why countries choose to default. A vexing question is why lenders repeatedly trust some less-than-reliable borrowers, since lenders presumably know that default over the life of the loan is a good possibility. In this vein, readers of parts II and III will learn, among other things, the following key points:

1. Despite the limited ability of creditors to fully recoup their losses, countries are nonetheless able to borrow from foreign creditors because of concerns about access to international capital markets (for example, borrowing to buy food in case of a natural disaster) or facilitating trade or foreign direct investment, or for reasons related to diplomatic relations. They argue that countries do not repay their debts for the opportunity to borrow even more in the future.
2. Historically, banking crises that originate in global financial centers tend to be contagious because they produce a "sudden stop" in lending to smaller countries—particularly crisis-prone countries, which often borrow in excess when times are good. As these crises unfold, falling commodity prices and rising interest rates in smaller countries help to precipitate sovereign debt crises. In short, smaller countries that borrow too much are exceptionally vulnerable when global growth slows.
3. Defaults on external debt frequently occur in clusters. These types of defaults can occur regionally, such as the wave of European debt defaults after the Napoleonic Wars, or internationally, as during the Great Depression. Cluster defaults have been reduced because of large lending programs by the International Monetary Fund and the World Bank.
4. Governments of emerging markets often view favorable shocks as permanent developments and, as a result, increase government spending and borrowing.
5. Economic conditions before and after a default on domestic debt are considerably worse than for a default on external debt. For instance, from 1800 to 2008, the inflation rate in the three years following a crisis averaged nearly 120 percent for domestic defaults, but only 32 percent for episodes of external default.

The Second Great Contraction

The second half of the book, parts IV through VI, is a topical discussion of banking and inflation crises, the aftermath of financial crises, the recent U.S. subprime crisis and the international dimensions of the subprime crisis. The book concludes with historical composite measures of financial turmoil and the typical "what have we learned" chapter. There are also several appendixes listing data sources.

In the authors' view, banking crises are remarkably similar in how they affect rich and poor countries. In this sense, they are an "equal opportunity menace." At the same time, banking crises can take different forms across the income strata of nations. For instance, financial repression is a type of banking crisis that only poor countries tend to experience: Depositors in poor or developing countries deposit funds in banks (because there are few or no alternatives), and then the bank is directed by the government to purchase debt issued by the government. The situation is sometimes made worse by the government instituting interest rate caps at a low nominal rate and then generating much higher rates of inflation.

A second type of crisis is the traditional bank run. The bank funds its assets, which tend to be long-term loans, with short-term liabilities (demand deposits). During a crisis, depositors withdraw their funds in a sudden panic—a bank run, which forces banks to liquidate assets, often at "fire sale" values—which further magnifies the crisis. In the United States, deposit insurance has effectively eliminated bank runs, at least in the formal banking sector. However, as Gorton (2010) details, "runs" did happen in the shadow banking system in the 2007-09 financial crisis. These runs

occurred because some financial firms refused to renew repurchase agreements or they imposed sizable "haircuts," which forced a significant amount of deleveraging—that is, reducing debt through rapid asset sales—by banks and other financial intermediaries.[6]

One reason banking crises are protracted affairs is the amplification mechanisms that stem from this deleveraging.[7] Using their data set, R&R show that real house prices typically rise sharply prior to a banking crisis and then fall sharply during the crisis and even after the crisis ends. A decline in real house prices, they argue, produces much more virulent banking crises than a decline in stock prices. This may help explain the relative mildness of the 2001 recession, which came on the heels of the collapse in the prices of technology stocks.

Another difference among types of asset price collapses is the marked increase in public indebtedness after a banking crisis triggered by a collapse in real housing prices. R&R find that in peak-to-trough cycles of real housing prices and banking crises, there is little quantitative difference between emerging and advanced economies. More important, unlike debt defaults, they argue that no country has been able to "graduate" from banking crises. Banking crises seem to be an enduring feature of the economic and financial landscape.

Sovereign credit risk

According to R&R, increased public indebtedness is the true legacy of banking crises. Focusing on crisis episodes for 13 emerging and advanced economies in the post-World War II period, they show that real (central) government debt increases by 86 percent in the three years following the crisis. And since real GDP falls, according to their data, by an average of more than 9 percent during an average two-year crisis, the result is a near doubling of the debt-to-GDP ratio in a relatively short time. In the United States, the federal debt-to-GDP ratio in nominal terms is projected to rise from about 36 percent in fiscal year 2007 to about 67 percent in 2012. In the aftermath of the Second Great Contraction, rising public indebtedness has run head-on into subdued economic recovery.

The authors note that recessions in advanced economies tend to spill over onto emerging market economies. This "collateral damage" may linger if advanced economies take longer than usual to return to their normal growth rate. If history is any guide—and of course, that is the premise of their book—then the fallout from the Second Great Contraction will be an "elevated number of defaults, reschedulings, and/or massive IMF bailouts" for emerging market economies. Indeed, one of the key legacies of banking and financial crises is rising public indebtedness and increased sovereign credit risk. But in the aftermath of the recent crisis, it is generally the advanced economies rather than the emerging market economies which, so far, have seen rapidly rising debt-to-GDP ratios.[8]

Admittedly, default on sovereign debt is an extremely low-probability event for most advanced countries. R&R show that Canada and the United States have managed to avoid this outcome over their relatively short histories, while default in other advanced economies in the 20th century, such as France, Germany and the United Kingdom, generally occurred only during periods of exceptional turmoil (the aftermath of wars or hyperinflation).[9] Still, with debt-to-GDP ratios in the advanced countries rising to ignominious levels, the question is not so much whether the advanced economies will default on their debt in the Russian or Argentinean sense, but whether default will occur in a different form.

One old-world favorite, they argue, is debt default through debasement—devaluation of the currency. In the old days, a monarch could reduce the gold or silver content of coins to finance wars or other large expenditures. Debasement is much easier under a modern fiat currency system, since the monetary authority can generate unexpected increases in inflation, so that debt can be repaid in currency with significantly less purchasing power than when first issued. But with many people more worried about deflation risks than inflation risks these days, the possibility of debasement seems remote. Nonetheless, R&R ominously warn that quiet periods of inflation "do not extend indefinitely." Perhaps those who can't fathom an acceleration of inflation in the foreseeable future would be wise to ponder why this time is different.

It is difficult to conceive that today's central bankers would countenance an unexpected surge in inflation as a way to reduce real debt burdens. Yet,

there have been alarming discussions—if only conjectural at this point—that the world's major central banks should contemplate raising their explicit or implicit inflation targets. Why? To better escape the zero nominal bound problem in the future.[10] Although they do not address this issue directly, R&R warn that there are clear inflation risks from rising levels of domestic debt. But at the same time, they warn that a strict inflation targeting regime can only be justified if there are equally strict regulations against excessive leverage. This seems like cold comfort to those who worry about the dangers of high and rising inflation in an era of aging populations and exploding debt-to-GDP ratios.

Four expensive words

The legendary British investor Sir John Templeton might not have been the first to utter the words, but his quip that "this time is different [are the] four most expensive words in the English language" rings loud and clear through R&R's analysis. In the final chapter of their book, they argue that no country—regardless of its size or importance—is immune to the syndrome of believing that times—and financial prospects—have changed, because so few people remember the key lessons from history.

What is needed, the authors contend, is an entirely new international regulatory institution that would collect, analyze and disseminate cross-country data designed to improve macroprudential oversight. Only an international authority, they claim, would "provide some degree of political insulation from legislators who relentlessly lobby domestic regulators to ease up on regulatory rule and enforcement."

But would such a supranational financial regulator with a long institutional memory have prevented the worst of the 2007-09 financial crisis? Perhaps a better question is whether the benefits of an all-powerful regulator would exceed its costs—or whether the world's countries would be willing to cede some of their sovereignty to prevent a once-in-100-years crisis, let alone a vastly smaller crisis.

Recently, Wilkinson, Spong and Christensson (2010) assessed the effectiveness of the information and analysis provided before and during the financial crisis by the Financial Stability Reports (FSRs) published by the central banks of four countries: The United Kingdom, Sweden, the Netherlands and Spain. They concluded that

> these four FSRs were generally successful in identifying risks that played important roles in the crisis—although they underestimated its severity. While it is not clear that FSRs helped to reduce the damages, it would be a mistake to dismiss them as a useful tool. Overall, publishing FSRs appears to be a worthwhile exercise that encourages central banks and international authorities to identify and monitor important trends and emerging risks and to develop a better understanding of the underlying structure of domestic and global financial markets.

This evidence of modest effectiveness suggests that R&R, and maybe even many policymakers themselves, should temper their enthusiasm for how much a new supranational financial regulator might accomplish. It might help identify risk, but likely won't prevent crises.

History teaches important lessons for designing future economic policies. In that regard, it is difficult to believe that R&R could have timed the release of their book any better. But if, as they insist, everyone regularly underestimates the fragility of highly leveraged economies, then what are the global implications of an aging population that, based on current policies, will produce future debt-to-GDP ratios that would make a third-world dictator blush? It can't be pretty.

This Time Is Different: Eight Centuries of Financial Folly belongs on the short list of economic books that key policymakers should carefully read—if for no other reason than to remind them that when they hear economists, analysts and even other policymakers utter the popular refrain "this time is different," what should immediately come to mind is not smooth sailing ahead, but storm clouds building on the horizon. Or, in the immortal words of Charles Kindleberger, financial crises are "hardy perennials."[11] Maybe Schumpeter was onto something after all. R

Endnotes

[1] Quoted in McCraw (2007), p. 250.

[2] Ibid.

[3] In a subsequent paper, R&R examine real GDP growth (median) at various levels of federal government debt for 20 advanced economies, from 1790 to 2009. They find that real GDP growth is 3.9 percent per year when a government's debt-to-GDP ratio is below 30 percent. But when the debt-to-GDP ratio rises to 90 percent or higher, the median level of annual GDP growth falls to 1.9 percent. See Reinhart and Rogoff (2010).

[4] Mian and Sufi (2010) argue that two main competing explanations seek to explain the 2000s housing and credit boom. On the one hand was a shift in the demand for credit chiefly due to real factors (for example, former Fed Chairman Alan Greenspan's productivity-driven New Economy). On the other hand was an increase in the supply of credit driven by financial innovations, such as securitization. Using microeconomic data, Mian and Sufi find data support for the latter explanation.

[5] The Maastricht Treaty is a 1992 agreement that created the European Union and set the rules for membership in the euro-area.

[6] A haircut refers to the difference between the market value of the collateral pledged by the borrower and the amount of the funds lent. For example, a 10 percent haircut means that the lender will loan to the borrower 90 percent of the value pledged as collateral.

[7] These dynamics are discussed in the context of the financial accelerator models of Bernanke and Gertler (1990) or Kiyotaki and Moore (1997).

[8] See Buiter (2010).

[9] In the United States, three states repudiated their debts from 1841 to 1842; in the late 1800s, 10 defaulted.

[10] See "IMF Tells Bankers to Rethink Inflation," which appeared in the Feb. 12, 2010, *Wall Street Journal*. Briefly, the zero nominal bound problem refers to a situation in which the central bank—seeking to boost growth or reduce the probability of deflation—cannot lower its nominal interest rate target below zero.

[11] See Kindleberger and Aliber (2005).

References

Bernanke, Ben S., and Mark Gertler. 1990. "Financial Fragility and Economic Performance" *Quarterly Journal of Economics* 105(February), pp. 87–114.

Buiter, Willem. 2010. "Sovereign Debt Problems in Advanced Industrial Countries." *Global Economics View,* Citigroup Global Markets, April 26.

Gorton, Gary. 2010. *Slapped by the Invisible Hand: The Panic of 2007*. Oxford University Press.

Kindleberger, Charles P., and Robert Aliber. 2005. *Manias, Panics, and Crashes: A History of Financial Crises*. Wiley, 5th ed.

Kiyotaki, Nobuhiro, and John Moore. 1997. "Credit Cycles." *Journal of Political Economy* 105(February), pp. 211–48.

McCraw, Thomas K. 2007. *Prophet of Innovation: Joseph Schumpeter and Creative Destruction*. Harvard University Press.

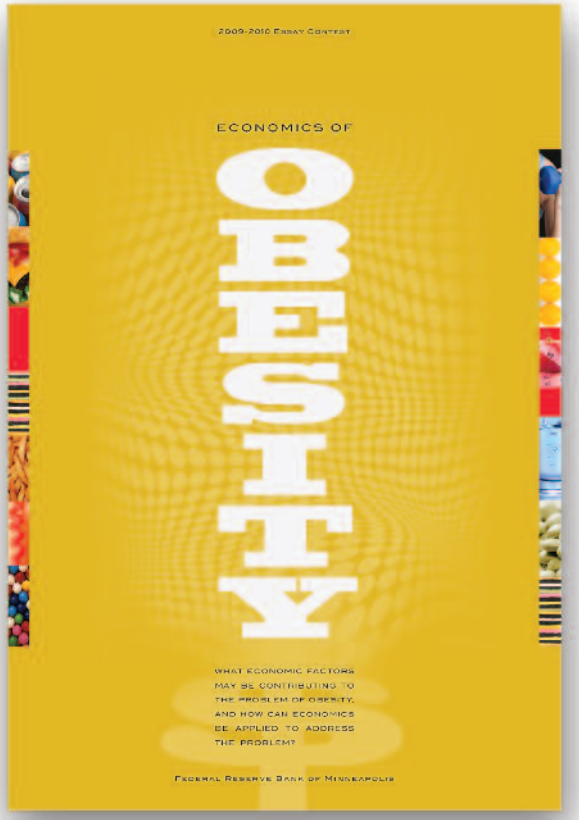
Mian, Atif, and Amir Sufi. 2010. "The Great Recession: Lessons from Microeconomic Data." *American Economic Review: Papers and Proceedings* 100(May), pp. 51–56.

Reinhart, Carmen M., and Kenneth S. Rogoff. 2010. "Growth in a Time of Debt." *American Economic Review: Papers and Proceedings* 100(May), pp. 573–78.

Wilkinson, Jim, Kenneth Spong, and Jon Christensson. 2010. "Financial Stability Reports: How Useful During a Financial Crisis?" *Federal Reserve Bank of Kansas City Economic Review* 95(1st Quarter), pp. 41–70.

2009–2010 Student Essay Contest

The Economics of Obesity



Essay Question

What economic factors may be contributing to the problem of obesity, and how can economics be applied to address the problem?

This spring the Minneapolis Fed held its 22nd Annual Student Essay Contest, which is open to high school juniors and seniors in the Ninth Federal Reserve District. The contest drew 243 essays from schools throughout the district. Submissions were divided into two categories: standard and advanced economics classes. The essay selected as the best over both categories is published here. Other top essays can be found at minneapolisfed.org under the Student Resources section of Community & Education.

Fifteen finalists in each division received a $100 U.S. savings bond. First- and second-place winners from both divisions received additional savings bonds. A paid summer internship at the Minneapolis Fed was awarded to the overall winner, Michael Graham of the Blake School in Minneapolis.

For almost all of the human past, the prospect of starvation was a real threat to most people. While scarcity is still the pervasive fact of economics, modern industrial economies have an abundance of low-cost food. As a result, the United States and other countries have seen an increase in rates of obesity. The health care costs of obesity are high, and some claim that increasing obesity rates inflict costs on the rest of society. For this reason, there might be a case for public action to reduce obesity.

Students were asked to explore why obesity has increased and what sorts of policies (if any) can combat this increase.

Student Essay Contest Winner

Economics of Obesity: Causes and Solutions

Michael Graham
Blake School
Minneapolis, Minnesota

The incidence of obesity in America has exploded over the past quarter century. The percentage of obese Americans—those having a body mass index (BMI) over 30 (about 30 pounds overweight for a 5'4" woman)[1] —has sharply risen from 15 percent in 1980 to just over 34 percent in 2006.[2] Notwithstanding author J. Eric Oliver's whimsical claim that obesity is not intrinsically harmful,[3] it is (as he agrees) at *least* a microcosm of Americans' fundamental mismanagement of their dietary and exercise needs. The impacts are marked: Obese people spend 42 percent more on health care ($1,429 more per year),[4] obesity costs the nation $75 billion in direct costs each year,[5] the total cost of obesity is as high as $139 billion per year (indirect costs include absenteeism, disability and workers' compensation)[6] and obesity is linked to approximately 300,000 deaths each year.[7] Notably, many of these costs are borne by private hospitals, the government and businesses rather than the obese citizens themselves, an important economic concept.[8]

Obesity's red herrings

Unfortunately, many policymakers are misled by red herring culprits for obesity. To be sure, it cannot be a decrease in exercise; Americans' energy expenditure habits have been static over the time period.[9] It cannot be cultural changes; data showing the same trends among fresh immigrants to the United States suggest that there is not a driving cultural force behind obesity.[10] It cannot be fast food restaurants' "super-sized" bundles; there has been no discernible increase in calories per meal.[11] It cannot be poverty; there is a *decreasing* gap between obesity rates of different socioeconomic population segments over the time period,[12] with much of the remaining gap attributable to varying genetic predispositions to obesity associated with race.[13]

The true culprit: Snacks

A litany of studies has shown that Americans have fundamentally increased their caloric intake over the past quarter century, and this increase fully accounts for America's ballooning obesity rate.[14] This increase is due to an increase in meals per day; since 1975, average snacks per day has increased by 60 percent.[15] Moreover, these snacks are often high in calories and low in nutritional value: "[S]ales of high-salt, high-calorie snack foods have skyrocketed, while sales of fruits and vegetables (excluding potatoes) has only increased marginally," particularly in the soft drink sector.[16] In addition to these factors, obesity itself has powerful biological and social positive feedback mechanisms that only add to the problem. First, as Oliver explains, "To consume about three hundred calories, all one needs to do is eat a seventy-cent bag of potato chips, a Snickers bar, or six Oreo cookies. To burn off three hundred calories … the average person needs to walk vigorously for about three miles."[17] Second, studies demonstrate that children's diet and exercise habits mimic those of their parents.[18] So, then, generational progression is not enough to combat obesity. Not only are the obese faced with an uphill battle reversing their state, but obesity begets more obesity as time progresses.

Market failures

Two major market failures have produced the caloric increase: the detrimental externalities of obesity and consumers' inability to efficiently allocate between the present and the future. The detrimental externalities of obesity are manifest. Obese citizens pay for little of the total cost of their obesi-

ty. Because much of the cost is passed on to private hospitals, the government and businesses, citizens actually become more obese than they themselves are willing to pay for. The resulting societal detriments burden everyone in the economy.

People's precarious tendency to buy more obesity than the socially optimal level is compounded by their failure to adequately allocate between the present and the future. Beginning in the late 1970s, numerous technological innovations in food preparation greatly increased the efficiency of food production in terms of both time and monetary investment.[19] This led to the widespread development of processed foods (those foods most often used as snacks) and to major time savings in food preparation. For instance, "the average time mothers spend preparing meals at homes has declined by more than 50 percent in the last two decades."[20]

While these developments may seem beneficial, the vastly lowered costs of eating have combined with widespread hyperbolic discounting to produce the increase in caloric intake.[21] This is because individuals deviate "from the usual standard rational choice models of uniform discount rates" by engaging in hyperbolic discounting—using a short-run discount rate that is larger than the long-run discount rate.[22] Thus, when food becomes readily available to individuals, the high marginal utility of eating is not properly compared to the future costs of increased intake, resulting in overconsumption.[23]

The economic solutions: Deliberate taxation

The total costs of obesity to American society, while intrinsically incalculable, in combination with the widespread market failures provide strong justification for judicious government political-economic action to realign incentives and correct the failures. The application of two indirect taxes would make great strides toward correcting these failures while, provided effective implementation, minimizing unplanned excess burdens.

Given the clear and widespread detrimental externalities of obesity, the government should institute an excise tax (specifically, a value-added tax) on foods with a high-caloric content but low nutritional value. While there have been historical impasses with implementing such taxes, the initial expenditures in defining which foods are included under the tax would easily be recovered by shifting more cost onto consumers, thus reducing total consumption of obesity (for analytical purposes, obesity is considered a good that people consume) to its socially optimal level. In addition, the government should create a payroll tax for obese citizens to supplement higher premiums for the obese to pay, thus further shifting the marginal social cost back onto obese citizens.

Lastly, the government should earmark at least some of the revenue raised from these taxes to fund community education programs. The cost-efficacy of well-selected education programs is especially appealing; one such program doubled the market share for low-fat and fat-free milk in several communities through campaigns that cost as little as 22 cents per person.[24] Indeed, another program was found to have a benefit-cost ratio of 10.64 in terms of expenditures on the program versus dollar benefits of avoiding or delaying health care costs and losses of productivity associated with obesity.[25]

Preliminary analysis of some nutritionally detrimental foods has demonstrated that a marginal tax rate of 20 percent or more would be necessary to instill change in consumer preferences due to a relatively high price inelasticity of demand. However, there is compelling evidence that consumers' demand for soda is elastic enough to support lower marginal tax rates and still result in reduced consumption (with more of the tax incidence placed on the consumer because of the elasticity). Indeed, a 10 percent increase in the price of soda has been shown to *halve* consumption.[26] Moreover, the public campaigns financed by the taxes would serve to shift the cross-elasticity of demand such that an increase in price of nutritionally detrimental foods would result in increased demand for healthier substitutes. For example, a negative shift in demand for high-fructose corn syrup would both send consumers searching for alternatives (e.g., juice) and shift the production possibilities frontier such that more farmers would grow alternatives (e.g., fruit) in response to the shift in consumer preferences.

While the value-added tax on nutritionally detrimental foods and the payroll tax are both regressive, the government should not reject them as solutions on equity concerns. First, those at the bottom of the income distribution could be insulat-

ed from this incidence by increasing the value of food stamps toward healthy food items. Additionally, even if citizens with comparatively lower incomes were taxed more, this would ultimately be beneficial since obesity leads to lower wages in the workforce and increased personal medical costs.[27] Lastly, public campaigns would have funds to address whatever food concerns there are in income-disadvantaged communities in the status quo, thereby reducing the excess burden on those citizens. Thus, the cost-benefit analysis of these taxes would always prove to be beneficial to the income disadvantaged.

Conclusions

When the tantalizing but ultimately misleading potential causes of obesity in America are eliminated, the true guilt of increased caloric intake due to widespread snack consumption becomes clear. And when the astounding detrimental externalities of obesity and modern food processes' tendency to exacerbate citizens' behavioral tendency to discount hyperbolically are considered, the necessity of government intervention becomes equally clear. Indeed, through shifting the marginal private cost of being obese toward the true marginal social cost, the taxes would serve to rein in obesity to its decidedly much lower socially optimal level of equilibrium. In the end, these policies would serve to better maximize utils in the American economy and therefore constitute the optimal economic decision. R

Endnotes

[1] Kinsey, Jean. "The Economics of Overeating." Lecture, College in the Schools Teacher Workshop, University of Minnesota, Minneapolis, Oct. 23, 2009.

[2] Levi, Jeffrey, Serena Vinter, Liz Richardson, Rebecca St. Laurent and Laura M. Segal. "F as in Fat: How Obesity Policies Are Failing in America." Healthy Americans. healthyamericans.org/reports/obesity2009/Obesity2009Report.pdf (accessed Jan. 11, 2010), 10.

[3] Oliver, Eric J. Fat Politics: *The Real Story Behind America's Obesity Epidemic*. (London: Oxford University Press, 2006), 3–5.

[4] Kinsey.

[5] Levi et al., 27.

[6] Levi et al.

[7] Seiders, Kathleen, and Ross D. Petty. "Obesity and the Role of Food Marketing: A Policy Analysis of Issues and Remedies." *Journal of Public Policy & Marketing* 23, no. 2 (2004): 153–69. jstor.org (accessed Jan. 10, 2010), 153.

[8] Kinsey.

[9] Cutler, David M., Edward L. Glaeser and Jesse M. Shapiro. "Why Have Americans Become More Obese?" *Journal of Economic Perspectives* 17, no. 3 (2003): 93–118. jstor.org (accessed Jan. 11, 2010), 103.

Oliver, 150.

[10] Antecol, Heather, and Kelly Bedard. "Unhealthy Assimilation: Why Do Immigrants Converge to American Health Status Levels?" *Demography* 43, no. 2 (2006): 337–60. jstor.org (accessed Jan. 11, 2010).

[11] Cutler et al., 101.

[12] Zhang, Q., and Y. Wang. "Trends in the Association between Obesity and Socioeconomic Status in U.S. Adults: 1971 to 2000." *Obesity Research* 10 (2004): 1622–32. http://www.ncbi.nlm.nih.gov/pubmed/15536226 (accessed Jan. 11, 2010).

Morrill, Allison C., and Christopher D. Chinn. "The Obesity Epidemic in the United States." *Journal of Public Health Policy* 25, no. 3/4 (2004): 353–66. jstor.org (accessed Jan. 11, 2010), 355.

[13] Levi et al., 27.

Finkelstein, Eric A., and Laurie Zuckerman. *The Fattening of America: How The Economy Makes Us Fat, If It Matters, and What To Do About It.* (New York: Wiley, 2008), 52.

[14] Cutler et al., 101.

Loureiro, Maria L. "Obesity: The Economics of a 'Super Size' Problem." *CHOICES Magazine*, Fall 2004. http://www.choicesmagazine.org/2004-3/obesity/2004-3-02.htm (accessed Jan. 11, 2010).

Morill et al., 359.

[15] Cutler et al., 101.

[16] Oliver, 134.

[17] Oliver, 151.

[18] Oliver, 53.

[19] Cutler et al., 106.

[20] Oliver, 136.

[21] Cutler et al., 113.

[22] Epstein, Richard A. "Behavioral Economics: Human Errors and Market Corrections." *University of Chicago Law Review* 73, no. 1 (2006): 111–32. jstor.org (accessed Jan. 11, 2010), 130.

[23] Cutler et al., 112.

Lakdawalla, Darius, Tomas Philipson and Jay Bhattacharya. "Welfare-Enhancing Technological Change and the Growth of Obesity." *American Economic Review* 95, no. 2 (2005): 253–57. jstor.org (accessed Jan. 11, 2010), 253.

Hayne, Cheryl L., Patricia A. Moran and Mary M. Ford. "Regulating Environments to Reduce Obesity." *Journal of Public Health Policy* 25, no. 3/4 (2004): 391–407. jstor.org (accessed Jan. 11, 2010), 392–93.

Finkelstein et al., 19.

[24] Nestle, Marion, and Michael F. Jacobson. "Halting the Obesity Epidemic: A Public Health Policy Approach." Public *Health Reports* 115 (2000): 12–24. http://cspinet.org/reports/obesity.pdf (accessed Jan. 11, 2010), 10.

[25] Kuchler, Fred, Abebayehu Tegene, and J. Michael Harris. "Taxing Snack Foods: Manipulating Diet Quality or Financing Information Programs?" *Review of Agricultural Economics* 27, no. 1 (2005): 4–20. web.ebscohost.com (accessed Jan. 11, 2010), 18.

[26] Kinsey.

[27] Viner, Russel M., and Tim J. Cole. "Adult Socioeconomic, Educational, Social, and Psychological Outcomes of Childhood Obesity: A National Birth Cohort Study." *British Medical Journal* 330, no. 7504 (2005): 1354–57. jstor.org (accessed Jan. 11, 2010).

The Region

Public Affairs
Federal Reserve Bank of Minneapolis
P.O. Box 291
Minneapolis, Minnesota 55480-0291

Change Service Requested

Presorted Standard
U.S. Postage
Paid
Permit No. 2235
Minneapolis, MN

The main thing that concerns me is the threat of *persistent* high unemployment, and here the European experience of the last three decades fills me with dread.

—Thomas Sargent