FEDERAL RESERVE BANK OF ST. LOUIS

REVIEW

SECOND QUARTER 2014
VOLUME 96 | NUMBER 2

Special Centennial Section

100 YEARS™
FEDERAL RESERVE SYSTEM

REVIEW

Volume 96 • Number 2

Special Centennial Section

Review Now Published Quarterly

Review is published four times per year by the Research Division of the Federal Reserve Bank of St. Louis. Complimentary print subscriptions are available to U.S. addresses only. Full online access is available to all, free of charge.

Online Access to Current and Past Issues

The current issue and past issues dating back to 1967 may be accessed through our Research Division website:
http://research.stlouisfed.org/publications/review. All nonproprietary and nonconfidential data and programs for the articles written by Federal Reserve Bank of St. Louis staff and published in *Review* also are available to our readers on this website.

Review articles published before 1967 may be accessed through our digital archive, FRASER: **http://fraser.stlouisfed.org/publication/?pid=820**.

Review is indexed in Fed in Print, the catalog of Federal Reserve publications (**http://www.fedinprint.org/**) and in IDEAS/RePEc, the free online bibliography hosted by the Research Division (**http://ideas.repec.org/**).

Authorship and Disclaimer

The majority of research published in *Review* is authored by economists on staff at the Federal Reserve Bank of St. Louis. Visiting scholars and others affiliated with the St. Louis Fed or the Federal Reserve System occasionally provide content as well. *Review* does not accept unsolicited manuscripts for publication.

The views expressed in *Review* are those of the individual authors and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System, or the Board of Governors.

Subscriptions and Alerts

Single-copy subscriptions (U.S. addresses only) are available free of charge. Subscribe here:
https://research.stlouisfed.org/publications/review/subscribe/.

Our monthly email newsletter keeps you informed when new issues of *Review*, *Economic Synopses*, *Regional Economist*, and other publications become available; it also alerts you to new or enhanced data and information services provided by the St. Louis Fed. Subscribe to the newsletter here: **http://research.stlouisfed.org/newsletter-subscribe.html**.

Copyright and Permissions

Economic Data

General economic data can be obtained through FRED (Federal Reserve Economic Data), our free database with more than 200,000 national, international, and regional data series, including data for our own Eighth Federal Reserve District. You may access FRED through our website: **http://research.stlouisfed.org/fred2**.

ISSN 0014-9187

Monetary Policy in the United States: A Brave New World?

Stephen D. Williamson

This article is a reflection on monetary policy in the United States during Ben Bernanke's two terms as Chairman of the Federal Open Market Committee, from 2006 to 2014. Inflation targeting, policy during the financial crisis, and post-crisis monetary policy (forward guidance and quantitative easing) are discussed and evaluated. (JEL E52, N12)

Federal Reserve Bank of St. Louis *Review*, Second Quarter 2014, *96*(2), pp. 111-21.

Ben Bernanke chaired his last Federal Open Market Committee (FOMC) meeting in January 2014 and departed from the Board of Governors on February 3 after eight years as the head of the Federal Reserve System. So, the time is right to look back on the Bernanke era and ask how central banking has and has not changed since 2006.

There is plenty in the macroeconomic record from 2006 to 2014 to keep economists and policy analysts busy for many years, so in this short piece we can only scratch the surface of what is interesting about the Bernanke era. I will focus on three issues: (i) inflation targeting, (ii) Fed lending and other interventions during the financial crisis, and (iii) post-crisis Fed policy, in particular experiments with forward guidance and quantitative easing (QE).

INFLATION TARGETING

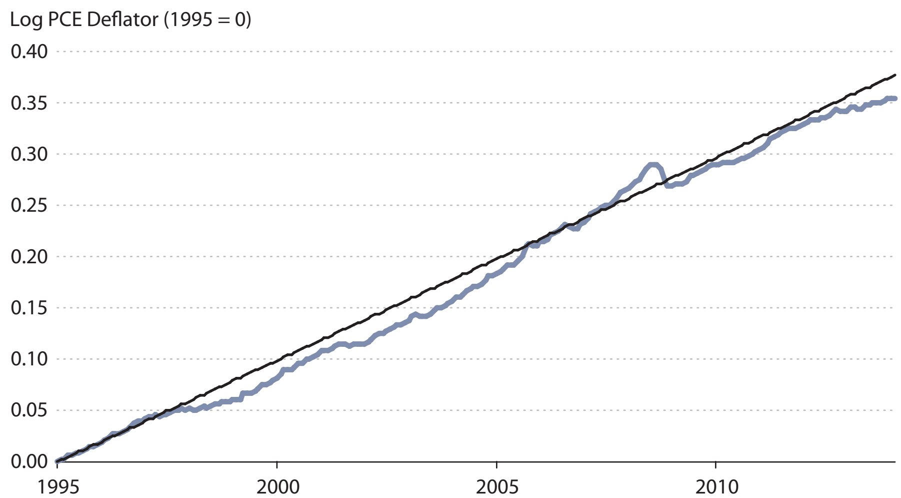
When Bernanke began his first term in 2006, I think the big change people expected was an inflation-targeting regime for U.S. monetary policy, similar to what exists in New Zealand, Canada, and the United Kingdom, for example. While that may have been in the cards, by the time Bernanke had settled into the job, events had overtaken him and he clearly ended up with much more than he bargained for. In terms of how Fed officials think about their jobs and how the public thinks about the role of the central bank, the Fed's objectives and its toolbox are far different from what existed, or was envisioned, in 2006.

Ben Bernanke was on record prior to his time as Fed Chair as a supporter of inflation targeting (see, for example Bernanke, 2004), which, as noted, had been adopted in other central

Stephen D. Williamson is the Robert S. Brookings Distinguished Professor in Arts and Sciences at Washington University in St. Louis, a research fellow at the Federal Reserve Bank of St. Louis, and a visiting scholar at the Federal Reserve Bank of Richmond.

Figure 1

PCE Deflator and 2 Percent Trend



banks, including those in New Zealand, Canada, and the United Kingdom. Bernanke (2004) argued that, in spite of the Fed's successes in controlling inflation, beginning in the Volcker era, it would be an improvement if the Fed stated a specific target for the inflation rate.

While Fed communications during Bernanke's term changed in ways that reflected more explicitly the dual mandate that comes from the U.S. Congress, the Fed ultimately stated explicitly that its target was a 2 percent per year increase in the raw personal consumption expenditures (PCE) deflator. In early 2012, a set of long-run goals for monetary policy were laid out by the Fed, which included this specific inflation target (see Board of Governors, 2012). This version of inflation targeting differs from that in some other countries (e.g., New Zealand or Canada) where there is an explicit agreement between the national government and the central bank that there will be an inflation target that is periodically renegotiated. In the United States, the central bank took it upon itself to come up with the inflation-targeting approach.

Figure 1 shows the path for the PCE deflator from January 1995 to February 2014, along with a 2 percent growth trend. This is quite remarkable, as inflation has actually not strayed far from the 2 percent trend path over this 19-year period. Thus, the Fed's announced inflation-targeting approach was simply ratifying a policy that had implicitly been in place for a long time. Thus, while announcing the 2 percent inflation target may have been important in reinforcing the Fed's commitment to low inflation, it appears that the Fed could have announced such a target in 1995 and that this would not have constrained its behavior. Perhaps surprisingly, once the Fed had announced the 2 percent inflation target, it began missing it on the low side, currently by about as much as at any time between January 1995 and the present.

THE FINANCIAL CRISIS: CENTRAL BANK LENDING, BAGEHOT'S RULE, AND MORAL HAZARD

It will take us many years to sort out what happened during the financial crisis and why. We do not even have all the facts yet, and I am sure there will be important revelations when the principals (hopefully including Bernanke) write about it. For starters, it is useful to read the transcript from a conference at the Brookings Institution in January 2014 (see Brookings Institution, 2014). This part of the discussion, between Chairman Bernanke and Liaquat Ahamed at the conference, relates to the principles of central banking in a financial crisis:

> **Mr. Ahamed:** You've said somewhere that the playbook that you relied on was essentially given by a British economist in the 1860s, Walter Bagehot. And his dictum was that in a financial crisis, the Central Bank should lend unlimited amounts to solvent institutions against good collateral at a penalty rate. How useful in practice was that rule in guiding you?
>
> **Mr. Bernanke:** It was excellent advice. This was the advice that's been used by Central Banks going back to at least the 1700s. When you have a market or a financial system that is short of liquidity and there's a lack of confidence, a panic; then the Central Bank is the lender of last resort. It's the institution that can provide the cash liquidity to calm the panic and to make sure that depositors and other short-term lenders are able to get their money.
>
> In the context of the crisis of 2008, the main difference was that the financial system that we have today obviously looked very different in its details, if not in its conceptual structure, from what Walter Bagehot saw in the 19th century. And so the challenge for us at the Fed was to adapt Bagehot's advice to the context of a modern financial system. So for example, instead of having retail depositors lining out, you know, standing in line out the doors as was the case in the 1907 panic, for example, in the United States; we had instead runs by wholesale short-term lenders like repo lenders or commercial-paper lenders, and we had to find ways to essentially provide liquidity to stop those runs.
>
> So it was a different institutional context, but very much an approach that was entirely consistent I think with Bagehot's recommendations.

Bagehot's name is often mentioned in discussions of the financial crisis, but the idea that what the Fed did during this recent episode is entirely consistent with Bagehot's recommendations may be entirely wrong. As Bernanke points out, Bagehot lived in an entirely different economic environment. In 1873, when Bagehot published *Lombard Street* (Bagehot, 1873), the Bank of England operated under the Bank Charter Act of 1844 (also know as Peel's Bank Act), which gave the Bank of England a monopoly on the issue of currency in the United Kingdom (except for some grandfathering). The United Kingdom operated under a gold standard, and part of the stock of Bank of England notes had to be backed 100 percent by gold. There was also a fiduciary component to the note issue, not backed by gold, and in practice this was a substantial fraction of the outstanding notes—sometimes a majority.

Banking panics during Bagehot's time were effectively flights from the liabilities of private financial intermediaries to gold and Bank of England notes. Bagehot's recommendation was to stem such panics through unlimited lending to private banks by the Bank of England against good collateral, *at a penalty rate*. Why the penalty rate? Bagehot appeared to be concerned with protecting the gold reserves of the Bank of England and seemed to think the high penalty rate on lending would be an efficient way to limit lending.

But was central bank lending actually an important part of the response to panics that occurred during the 19th century in the United Kingdom? Possibly not. The key problem during these panics was a shortage of media of exchange, so the best response of the Bank of England would have been to put more Bank of England notes into circulation, the stock of gold being essentially fixed (except for international gold flows). During the panics of 1847, 1857, and 1866, which occurred after Peel's Bank Act and before Bagehot published *Lombard Street*, Parliament acted to suspend the Bank Act, which permitted an expansion in the Bank of England's fiduciary issue of notes. Whether the notes got into circulation through lending or asset purchases by the Bank of England perhaps is irrelevant.

Thus, whether Bagehot's prescriptions were actually an important part of a crisis response during his own time is subject to some doubt. More to the point, it is very hard to say that any important intervention by the Fed during the financial crisis looked much like what Bagehot recommended. For example, a very large lending program was carried out using the Term Auction Facility, under which lending peaked at about $480 billion in March 2009. It seems the rationale for this program was the reluctance of financial institutions to borrow due to the stigma associated with standard discount window lending. Under the Term Auction Facility, the Fed chose a quantity of funds to lend, then auctioned this off. That is hardly unlimited lending at a penalty rate.

So, I think it would be useful if we declared a moratorium on the use of Bagehot's name in instances where the name is being used only to lend credence to some contemporary idea. We could probably make a case to do the same for Wicksell, Minsky, and Keynes, for that matter, but that would be taking us too far afield for this article.

It is common, I think, to rate the Fed highly for its performance during the financial crisis. Of course, it is also important to analyze carefully the Fed's behavior in 2007-09 and to be critical, so that we can learn and not repeat mistakes, if there were any. For example, consider the following. An interesting idea is that moral hazard associated with "too big to fail" is not only a long-term problem, but a problem that can present itself in the context of a crisis. For example, the Fed played a very important role in the Bear Stearns collapse. The Fed lent to Bear Stearns just prior to its unwinding in March 2008 and helped engineer its sale to JP Morgan Chase by taking some of Bear Stearns's assets onto the Fed's balance sheet. This Fed intervention, we could argue, then created the expectation among large financial institutions that the Fed was ready to intervene, and this may have led Lehman Brothers to forgo actions that may have circumvented its bankruptcy in September 2008. The Lehman bankruptcy seems to have been critical in precipitating a systemic crisis. So, perhaps the Fed made key errors in the instance of Bear Stearns, which had important consequences later in the year.

Another problem may have been excessive concern by the Fed regarding the solvency of money market mutual funds (MMMFs). These are institutions originally designed to—as much as possible—mimic the functions of commercial banks, without being subject to the same regulations. To the extent that MMMFs borrow short, lend long, and guarantee their creditors a nominal return of zero, they can be subject to the same sorts of instabilities as commercial banks. But of course there is no deposit insurance for MMMFs, only the implicit insurance that came into play during the financial crisis, involving both the Fed and the Treasury. But why should the Fed care? MMMFs choose to operate outside of standard banking regulation, and MMMF shareholders should be able to understand the consequences of holding these uninsured shares. And the systemic consequences of a MMMF failure should be slight. Clearly the shareholders lose, but the assets are highly liquid, and it seems easy for another financial institution to step in and pick up the pieces. If we think there is some fire sale phenomenon, that may give the central bank cause to intervene through temporary asset purchases, but why step in to save the MMMFs?

POST-CRISIS POLICY: FORWARD GUIDANCE AND QUANTITATIVE EASING

For good or ill, there has been a dramatic shift in how central bankers—not only in the Fed System but also in the world—think about policy. The key post-crisis innovation in the United States was (is) an extended period with the short-term policy interest rate at close to zero. This has been combined with the payment of interest on reserves, an increase in the size of the Fed's balance sheet, a dramatic change in the composition of assets on the balance sheet, an increase in the use of forward guidance, and a change in the interpretation of the dual mandate.

The pre-2008 period now seems very strange and simple. As a point of reference, this is the FOMC statement that followed the June 26-27, 2001, FOMC meeting, when Alan Greenspan was Chairman of the Federal Open Market Committee (see Board of Governors, 2001):

> The Federal Open Market Committee at its meeting today decided to lower its target for the federal funds rate by 25 basis points to 3-3/4 percent. In a related action, the Board of Governors approved a 25 basis point reduction in the discount rate to 3-1/4 percent. Today's action by the FOMC brings the decline in the target federal funds rate since the beginning of the year to 275 basis points.
>
> The patterns evident in recent months—declining profitability and business capital spending, weak expansion of consumption, and slowing growth abroad—continue to weigh on the economy. The associated easing of pressures on labor and product markets is expected to keep inflation contained.
>
> Although continuing favorable trends bolster long-term prospects for productivity growth and the economy, the Committee continues to believe that against the background of its long-run goals of price stability and sustainable economic growth and of the information currently available, the risks are weighted mainly toward conditions that may generate economic weakness in the foreseeable future.
>
> In taking the discount rate action, the Federal Reserve Board approved requests submitted by the Boards of Directors of the Federal Reserve Banks of Boston, New York, Philadelphia, Atlanta, Chicago, Dallas and San Francisco.

If you have been reading recent FOMC statements, you'll first notice that the above is remarkably short by comparison (202 words vs. 790 in the January 2014 FOMC statement;

see Board of Governors, 2014). The policy action relates to only one policy variable—the target federal funds rate; there is only a brief description of the state of the economy; there is no attempt to explain why the Fed is taking the action it did—we're left to infer that the state of the economy had something to do with why the federal funds rate target was lowered; there is some vague attempt to address the dual mandate (price stability and sustainable economic growth), but no mention of what economic variables (e.g., PCE inflation rate or unemployment rate) the Fed might be looking at; there is little forecast information, only mention of risks; and there is no forward guidance (i.e., no information about contingent future actions).

In light of the way the Fed now conducts policy, we might think of the above FOMC statement as much less forthcoming than it should have been. But remember that, at the time, people were generally inclined to think highly of the Fed's performance. We may remember quibbles, but there was certainly no consensus that the American public was getting a bad deal from the Fed. But clearly, changes were thought to be necessary during the Bernanke era.

The FOMC has apparently become much more activist. There seems to be a consensus view among FOMC participants—among whom PhD economists with high-level research records are increasingly prominent—that the Fed can and should make the world a better place. Alan Greenspan had much more practical background in the private sector compared with Ben Bernanke who, previous to his Fed appointment, had been a high-profile academic researcher at Princeton University. Greenspan's approach was simple: He wanted to control inflation (implicitly, as shown in Figure 1, at about 2 percent per year) and was willing to reduce the federal funds rate in response to low aggregate economic activity. Just how the behavior of the Fed under Bernanke differed from the behavior of the Greenspan Fed will certainly be a subject for study among economic researchers. We might, for example, like to know whether researchers could make inferences about how the Greenspan Fed would have behaved during the financial crisis.

A recent paper (Williams, 2014) does a nice job of showing us the role that formal economics played in post-financial crisis monetary policy in the United States. New Keynesian models were an important influence, of course, and some of the work on such models had already been done pre-financial crisis—for example, by Eggertsson and Woodford (2003), who studied monetary policy at the zero lower bound. A paper by Werning (2012), which came out post-financial crisis, played an important role in post-financial crisis policy discussions, as did Michael Woodford's Jackson Hole Conference paper (Woodford, 2012).

To think about monetary policy at the zero lower bound (on the short-term nominal interest rate), one first has to provide some justification for why it could be optimal for a central bank to choose zero as the overnight interest rate target. In the simplest New Keynesian models (Werning's, for example), that is done by supposing that the representative agent's discount factor is high for some period of time. Then, the real interest rate should optimally be low, but at the zero lower bound the real interest rate is too high, given sticky prices. In this model, then, the central bank would like to ease more at the zero lower bound, but the only game in town is to make promises about future central bank actions. Further, those promises may not be time consistent, so commitment by the central bank is key to making the policy work.

Bernanke's FOMC got into the business of making promises about the future in a big way. We are now very familiar with the words *forward guidance*. Initially, forward guidance took the form of "extended period" language, to the effect that the federal funds rate target was to remain low for an extended period of time. As the recovery from the past recession failed to proceed as well as expected, more explicit language was added to the FOMC statement about what "extended period" might mean, first in terms of calendar dates and then as a threshold for the unemployment rate.

In a New Keynesian model, it is clear how forward guidance works. The model will tell us when the central bank should choose a target short-term nominal interest rate of zero, when liftoff (departure from the zero lower bound) should occur, and what the nominal interest rate target should be away from the zero lower bound. All of those things are determined by the path followed by the exogenous shocks that hit the economy. In practice, we do not know how good the model is, we cannot observe the shocks, and even if we believe the model it will not allow us to identify the shocks. So, forward guidance must be specified in some loose way, in terms of something we actually can observe. Ultimately, the FOMC chose the unemployment rate.

What went wrong with forward guidance? If you read Woodford's Jackson Hole paper in 2012, you might understand why we might be in for trouble. Woodford takes close to 100 pages to tell us how forward guidance works, which may be a signal of a potential pitfall of forward guidance policy. People have enough trouble understanding conventional central bank actions, let alone complicated and evolving statements about what may or may not happen in the future. First the Fed committed to the zero lower bound for a couple of years, then for another year. Then it committed to a threshold for the unemployment rate. What does that mean? Is the Fed going to do something when the unemployment rate reaches 6.5 percent? No, not necessarily. It seems there could be an extended period at the zero lower bound. So, ultimately, the FOMC came full circle on this. Current Fed statements about forward guidance specify an extended period, now described as a considerable time after the asset purchase program (QE) ends.

This points out the dangers of commitment, with respect to things that the Fed cannot control. Central banks can indeed control inflation, and the benefits of committing to stable inflation are well understood. One might have thought—pre-financial crisis—that it was also well understood that the central bank's control over real variables—the unemployment rate, for example—is transient. Also, there are so many factors other than monetary policy affecting real economic variables that making promises about real economic activity is extremely dangerous for a central bank. If the FOMC has discovered those ideas again, this is not yet reflected in its policy statements.

Finally, probably the most important novel element in post-financial crisis U.S. monetary policy is QE and the associated increase in the size of the Fed's balance sheet. To understand how the Fed thinks about QE, useful references are Williams (2014) and Chung et al. (2012). Fed officials want to convince us that QE works, and they have tended to downplay the experimental nature of the QE programs. It has typically been argued by Fed officials that QE works just like conventional monetary policy (see, for example, Bernanke, 2012). Other than the fact that QE has its effects by reducing long bond yields rather than reducing short rates, the argument is that the transmission mechanism for monetary policy is essentially the same. But what exactly is the science behind QE? This quote from Bernanke (Brookings Institution, 2014) is representative:

> **Mr. Ahamed:** —but we were really operating blind. So in devising QE and all these other unconventional monetary policies, were you pretty confident that the theory would work or that whatever you—going into it?
>
> **Mr. Bernanke:** Well, the problem with QE is it works in practice, but it doesn't work in theory.

In fact, a model in which QE does not work is the basic Woodford New Keynesian model (see Woodford, 2003). In a Woodford model, the only friction results from sticky prices (and maybe sticky wages). Otherwise, financial markets are complete, and there is nothing that would make particular kinds of asset swaps by the central bank matter. Indeed, baseline New Keynesian models leave the liabilities of the government and the central bank out of the analysis completely. So, in the context of post-financial crisis monetary policy, one model—the New Keynesian model—is being used to justify forward guidance, but it is going to have to be another model—if there is one—that justifies QE. That could be a problem, as we might like to know how those policy tools work together or not.

Absent a model, the Fed went ahead with QE anyway, supported by some ideas about how financial market segmentation might produce the desired result (Bernanke, 2012). Basically, if the market in Treasury securities is segmented by maturity, or if the market in mortgage-backed securities (MBS) is segmented from the market in Treasuries, then if the Fed purchases long Treasuries or MBS, the prices of these securities will rise and their yields will fall. Then, the argument goes, declines in long bond yields will increase spending in exactly the way that reductions in the federal funds rate target would stimulate spending under conventional conditions.

So, clearly, the Fed's view is that, though the theory might be murky, it is obvious that QE works in practice. But perhaps not. For a summary of the evidence, see Williams (2014). The empirical evidence takes two forms—event studies and time-series studies. Basically, what researchers find is that announcements of QE and actual QE purchases are associated with subsequent declines in nominal bond yields and changes in other asset prices. There are two problems with the interpretation of the evidence. The first is that, to confidently measure the effects of a particular policy, we need an explicit theory of how the policy works. Clearly, the relationship between the theory—such as it is—and the measurement is tenuous in this case. Second, it is possible that QE could move asset prices—in exactly the way posited by Fed officials—even if QE is actually irrelevant. To see this, suppose a world where QE is irrelevant and everyone knows it. Then, of course, QE will not make any difference for asset prices—or anything else. But suppose, in the same world, that the central bank thinks that QE works and financial market participants believe the central bank. Then QE will in fact cause asset prices to move as the central bank says they will, at least until everyone learns what is going on. Models like that would be very difficult to work out, but this phenomenon could well be part of what we are dealing with.

Given that Bernanke and the other members of the FOMC had little idea what QE might actually do in practice, how did they go about determining the magnitude of the interventions? For example, QE2 was planned as $75 billion per month in purchases of long-maturity Treasuries over 8 months, and the current program was an open-ended purchase of $85 billion per month in long Treasuries and MBS. Obviously, some judgment was made that $75 to $85 billion per month in purchases was about right. Why? Chung et al. (2012) perhaps give us some clues. The important part of their paper in this respect is section 3, page 66. There is an exercise in which the quantitative effects of QE are measured. The basic idea is to take the Fed's FRB/US model as a starting point. In that model, we can conduct policy experiments such as changing the future path for the federal funds rate to determine how that matters. But, it seems, one cannot conduct experiments in that model to determine how changes in the composition of the Fed's balance sheet matter. Certainly that model will not tell us the difference between long-maturity Treasury securities and MBS. So, what is done is to translate a given quantity of asset purchases into a given change in the federal funds rate by making use of the available empirical evidence and then conducting the experiment as a change in the path for the federal funds rate.

There are two problems with that. First, we may not want to take seriously what the FRB/US model tells us about conventional monetary policy, let alone unconventional policy. What is in the FRB/US model is documented in Tulip (2014). The model is basically a dressed-up version of the types of macroeconometric models that were first built in the 1960s. Indeed, the FRB/US model is descended from the FRB/MIT/Penn model (see Nelson, 1972)—a late-1960s-vintage macroeconometric model comprising several hundred equations—basically an expanded IS/LM model. Robert Lucas (1976) and Christopher Sims (1980) were quite convincing when they told us we should not believe the policy experiments that were conducted using such models. My best guess is that there is nothing in the current version of the FRB/US model that would change our minds about the policy relevance of large macroeconometric models.

The second problem with the implementation of QE is that there may be aspects of these programs that work in quite different ways from conventional monetary policy. There is no good reason to think that swapping outside money for T-bills (effectively what happens under conventional conditions) is somehow the same as changing the maturity structure of the outstanding consolidated government debt. Indeed, in Williamson (2014), QE has some very different effects from conventional monetary policy.

Ultimately, though, QE in itself is not a big concern. Either it works or it does not; and, if it does not work, there is no direct harm. There is no good reason to think that QE is inherently inflationary. For example, in Williamson (2014), the long-run effect of QE is to lower inflation. Why? In that model, QE substitutes good collateral for less-good collateral, which relaxes financial constraints and lowers the liquidity premium on safe collateral. Real bond yields rise and, given a fixed short-term nominal interest rate (zero, say), inflation must fall.

As well, the size of the Fed's balance sheet is not a concern. For example, if the Fed expanded its balance sheet by issuing interest-bearing reserves to purchase T-bills, we should not be worried. If the Fed's balance sheet is expanded by purchasing long-maturity Treasury bonds, we should also not be worried, provided there are no unknown effects of changing the maturity structure of the Fed's assets. Finally, even though the Fed is now borrowing short and lending long, in a massive way, we should not care if, in the future, the Fed starts earning negative profits. Economically, that does not matter. In essence, what matters is the consolidated balance sheet of the Fed and the U.S. Treasury, and it is irrelevant whether the Fed pays the private sector interest on reserves or the U.S. Treasury pays the private sector interest on government debt.

CONCLUSION

So what should we make of the state of monetary policy in the United States as Ben Bernanke leaves the Fed? Primarily, I think, the Fed is unsettled about what it should be doing, and the public is uncertain about what it should be expecting from the Fed. It is fine to experiment, and everyone understands that the human race would get nowhere without experimentation. But when the Fed experiments, it needs to cast a critical eye on those experiments. I think it is unclear whether forward guidance has been of any value, and it is possible that careful a priori thinking could have headed off that experiment.

With QE, I think much theoretical and empirical work needs to be done to clarify the role of such policies. Whether QE has quantitatively significant effects or not, a key question is whether it is appropriate for the central bank to be engaging in debt management, which has traditionally been the province of the fiscal authority. Indeed, while QE was taking place, with the Fed acting to reduce the duration of the outstanding consolidated government debt, the Treasury was sometimes taking actions that would have the effect of increasing the average duration of government debt in the hands of the public. Given the view on the FOMC that QE works as advertised, it seems likely that QE will be put away, in the box of good monetary policy tools, for later use. If the maturity structure of the outstanding consolidated government debt is so important, then there should be a public discussion of who is to determine it—the Fed or the Treasury.

Finally, perhaps the most important lasting change in policy that comes out of the Bernanke era is a greater tendency of the Fed to focus on short-run goals rather than long-run goals. Indeed, the Fed's concern with the state of the labor market, as reflected in recent public statements by Fed officials, may have evolved into a belief that the Fed can have long-run effects on labor force participation and the employment/population ratio. That belief appears to be unsupported by theory or empirical evidence. Further, it is possible that two or three more years with the Fed's policy interest rate close to the zero lower bound will not help to fix whatever ails the labor market, nor will it increase the inflation rate, as low nominal interest rates (for example, as in Japan since the early 1990s) typically lead to low inflation in the long term (see Bullard, 2010). If the Fed falls well short of its 2 percent inflation goal for some period of time, it is not obvious that would be so harmful, but at the minimum it harms the Fed's credibility. ■

REFERENCES

Bagehot, Walter. *Lombard Street: A Description of the Money Market*. London: H.S. King, 1873.

Bernanke, Ben S. "Inflation Targeting." Federal Reserve Bank of St. Louis *Review*, July/August 2004, *84*(4), pp. 165-68; http://research.stlouisfed.org/publications/review/04/07/PanelDisc.pdf.

Bernanke, Ben S. "Monetary Policy Since the Onset of the Crisis." Presented at the Federal Reserve Bank of Kansas City Economic Symposium, "The Changing Policy Landscape," Jackson Hole, Wyoming, August 31, 2012; http://www.federalreserve.gov/newsevents/speech/bernanke20120831a.pdf.

Board of Governors of the Federal Reserve System press release, June 27, 2001; http://www.federalreserve.gov/boarddocs/press/general/2001/20010627/.

Board of Governors of the Federal Reserve System press release, January 25, 2012; http://www.federalreserve.gov/newsevents/press/monetary/20120125c.htm.

Board of Governors of the Federal Reserve System press release, January 29, 2014; http://www.federalreserve.gov/newsevents/press/monetary/20140129a.htm.

Brookings Institution. "Central Banking After the Great Recession: Lessons Learned and Challenges Ahead: A Discussion with Federal Reserve Chairman Ben Bernanke on the Fed's 100th Anniversary." January 16, 2014; http://www.brookings.edu/events/2014/01/16-central-banking-after-the-great-recession-bernanke.

Bullard, James. "Seven Faces of 'the Peril'." Federal Reserve Bank of St. Louis *Review*, September/October 2010, *92*(5), pp. 339-52; http://research.stlouisfed.org/publications/review/10/09/Bullard.pdf.

Chung, Hess; Laforte, Jean-Philippe; Reifschneider, David and Williams, John C. "Have We Underestimated the Likelihood and Severity of Zero Lower Bound Events?" *Journal of Money, Credit, and Banking*, February 2012, *44*(Suppl. s1), pp. 47-82.

Eggertsson, Gauti B. and Woodford, Michael. "The Zero Bound on Interest Rates and Optimal Monetary Policy." *Brookings Papers on Economic Activity*, 2003, *34*(1), pp. 139-211.

Lucas, Robert E. Jr. "Econometric Policy Evaluation: A Critique." *Carnegie-Rochester Conference Series on Public Policy*, January 1976, *1*(1), pp. 19-46.

Nelson, Charles R. "The Prediction Performance of the FRB-MIT-PENN Model of the U.S. Economy." *American Economic Review*, December 1972, *62*(5), pp. 902-17.

Sims, Christopher A. "Macroeconomics and Reality." *Econometrica*, January 1980, *48*(1), pp. 1-48.

Tulip, Peter. "FRB/US Equation Documentation." May 2014; http://www.petertulip.com/FRBUS_equation_documentation.pdf.

Werning, Iván. "Managing a Liquidity Trap: Monetary and Fiscal Policy." Working paper, MIT, April 2012; http://dl.dropboxusercontent.com/u/125966/zero_bound_2011.pdf.

Williams, John C. "Monetary Policy at the Zero Lower Bound: Putting Theory Into Practice." Hutchins Center on Fiscal and Monetary Policy, Brookings Institution, January 16, 2014; http://www.brookings.edu/~/media/research/files/papers/2014/01/16%20monetary%20policy%20zero%20lower%20bound/16%20monetary%20policy%20zero%20lower%20bound%20williams.

Williamson, Stephen D. "Scarce Collateral, the Term Premium, and Quantitative Easing." Working Paper No. 2014-008A, Federal Reserve Bank of St. Louis, March 2014; http://research.stlouisfed.org/wp/2014/2014-008.pdf.

Woodford, Michael. *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton, NJ: Princeton University Press, 2003.

Woodford, Michael. "Methods of Policy Accommodation at the Interest-Rate Lower Bound." Working paper, Columbia University, September 16, 2012; http://www.columbia.edu/~mw2230/JHole2012final.pdf.

The 2009 Recovery Act: Directly Created and Saved Jobs Were Primarily in Government

Bill Dupor

Over one-half of the fiscal spending component of the American Recovery and Reinvestment Act (ARRA; i.e., the Recovery Act) was allocated via grants, loans, and contracts. Businesses, nonprofits, and nonfederal government agencies that received this type of stimulus funding were required to report the number of jobs directly created and saved as a result of their funding. Created and saved jobs represent, precisely, the full-time equivalent of jobs funded by first- and second-tier recipients of and contractors on ARRA grants, loans, and contracts. In this article, the author categorizes these jobs into either the private sector (businesses and nonprofits) or the government sector. It is estimated that at the one-year mark following the start of the stimulus, 166,000 of the 682,000 jobs directly created/saved were in the private sector. Examples of private sector stimulus jobs include social workers hired by nonprofit groups to assist families, mechanics to repair buses for public transportation, and construction workers to repave highways. Examples of government stimulus jobs include public school teachers, civil servants employed at state agencies, and police officers. While fewer than one of four stimulus jobs were in the private sector, more than seven of nine jobs in the U.S. economy overall reside in the private sector. Thus, stimulus-funded jobs were heavily tilted toward government. (JEL E6, H7)

Federal Reserve Bank of St. Louis *Review*, Second Quarter 2014, *96*(2), pp. 123-45.

In February 2009, the U.S. federal government began its largest anti-recession fiscal stimulus in over 70 years when it passed the American Recovery and Reinvestment Act (ARRA; i.e., the Recovery Act).[1] The Congressional Budget Office (CBO)'s most recent assessment is that the cost of the Act will eventually total $821 billion.[2] This article studies the employment effects of the stimulus using a new dataset.[3]

The data consist of legally mandated reports provided by the universe of awardees of stimulus funds. In particular, each recipient of a contract, grant, or loan was required to file a report every three months that included a self-constructed estimate of the number of jobs

Bill Dupor is an assistant vice president and economist at the Federal Reserve Bank of St. Louis. The author thanks John Childs for providing data collected by the Ohio Department of Education and Tim Conley, Duane Dupor, Bill Gavin, Matt Lewis, Mike McCracken, Stephanie Moulton, Jay Shambaugh, and David Wheelock for useful comments and conversations. Alex Bruner and Peter McCrory provided excellent research assistance. A repository containing government documents, data sources, a subject bibliography, and other relevant information pertaining to the Recovery Act is available at billdupor.weebly.com.

directly created and/or saved as a result of its stimulus funding, as well as a general description of these jobs. Created and saved jobs represent, precisely, the full-time equivalent (FTE) of jobs funded by first- and second-tier awardees of and contractors on ARRA grants, loans, and contracts.[4]

Using these reports, I estimate that at the one-year mark of the program, 166,000 of the 682,000 jobs directly created or saved by the Act were in the private sector. Thus, fewer than one of four stimulus jobs were in the private sector. In contrast, more than seven of nine jobs in the U.S. economy overall reside in the private sector.[5]

To my knowledge, this article is the first that uses direct survey evidence to assess how many private sector and government jobs were funded by the Recovery Act. Recent years have seen an ongoing public debate on the differences (or lack thereof) of the stimulative effects of the two types of employment.

While limited in quantity, the direct creation/saving of private sectors jobs was not trivial. For example, the U.S. Department of Transportation administered ARRA projects that directly created and saved thousands of private sector jobs, most of which were in the depressed construction industry; however, transportation jobs made up less than 5 percent of all jobs reported in this period.

This article provides an empirical contribution to the debate on the ability of the government to stimulate the private economy, with particular focus on jobs. One view expressed in this debate is that there is little difference between private and government employment. "'Spending is spending,' said Lawrence J. White, an economist at New York University's Stern School of Business. There is no difference in the multiplier effect from a private sector job or a public sector job" (see Jacobson, 2013). Similarly, in a discussion of the composition of jobs created by the Recovery Act, Blinder (2013, p. 226) writes, "Aren't government jobs *jobs*?"

An alternative view holds that, with respect to boosting the economy, jobs created in the private sector are likely to be more effective stimulus and that the government is ineffective at private sector job creation. In a letter to President Obama sent in February 2011, 150 economists from universities and research institutes signed this statement: "Efforts to spark private sector job creation through government 'stimulus' spending have been unsuccessful." The letter goes on: "To support real economic growth and support the creation of private sector jobs, immediate action is needed to rein in federal spending."[6] Along these same lines, Cohen, Coval, and Malloy (2011) find that state-level fiscal spending shocks, driven by changes in the seniority of various U.S. Congress members, caused a decrease in the corresponding states' corporate employment and investment.

The outcome that I document has a different flavor than the one predicted by advisers to President-elect Obama. On January 9, 2009, Jared Bernstein and Christina Romer wrote "More than 90 percent of the jobs created are likely to be in the private sector."[7,8]

My estimate is also different from at least one media report assessing the results of the stimulus. A *Time* magazine article from February 2010 entitled "After One Year, a Stimulus Report Card," states "about half of the jobs that the government counts as created by the stimulus were state- or local-government-funded positions" (Gandel, 2010).

Analyzing the recipient-reported job creation data provides a new and distinct way to evaluate the stimulus. Other existing methods include aggregate time series analysis and cross-sectional studies. Cross-sectional studies on Recovery Act spending have found a positive, statistically significant jobs effect; however, this job creation was very expensive. Wilson (2012) finds that increasing employment by one worker at the one-year mark of the Act cost $125,000. Conley and Dupor (2013) find that, over the first two years following the Act's passage, it cost $202,000 to create a job lasting one year.[9] Given that the typical employment compensation (wages plus benefits) to a worker in a U.S. job is roughly $40,000 per year, the benchmark point estimates from these studies suggest that the Recovery Act created/saved a job at a cost of roughly three to five times that of the typical compensation.

The recipient-reported data have an advantage over these other methods in that they do not require any identification or statistical modeling assumptions. Furthermore, this dataset is an analog to existing surveys that ask respondents how they would (or did) put a particular tax cut or rebate to use (e.g., spend or save the tax savings), such as Shapiro and Slemrod (2003, 2009). As with those approaches, the responses alone should not be interpreted as the program's entire aggregate effects. One reason is that they do not take into account general equilibrium effects of fiscal policy.

The recipient-reported data do suffer from a drawback that may be less significant in other approaches. The recipient-reported jobs are only those directly created/saved due to this federal spending. These data will tend to *overstate* true job creation if the direct government jobs crowd out private sector employment that would have occurred. On the other hand, these data will tend to *understate* true job creation to the extent that there are jobs "indirectly" created by the spending. Moreover, these data will not include the jobs created indirectly as a result of the tax cut and transfer component of the stimulus.

BACKGROUND ON RECIPIENT-REPORTED DATA

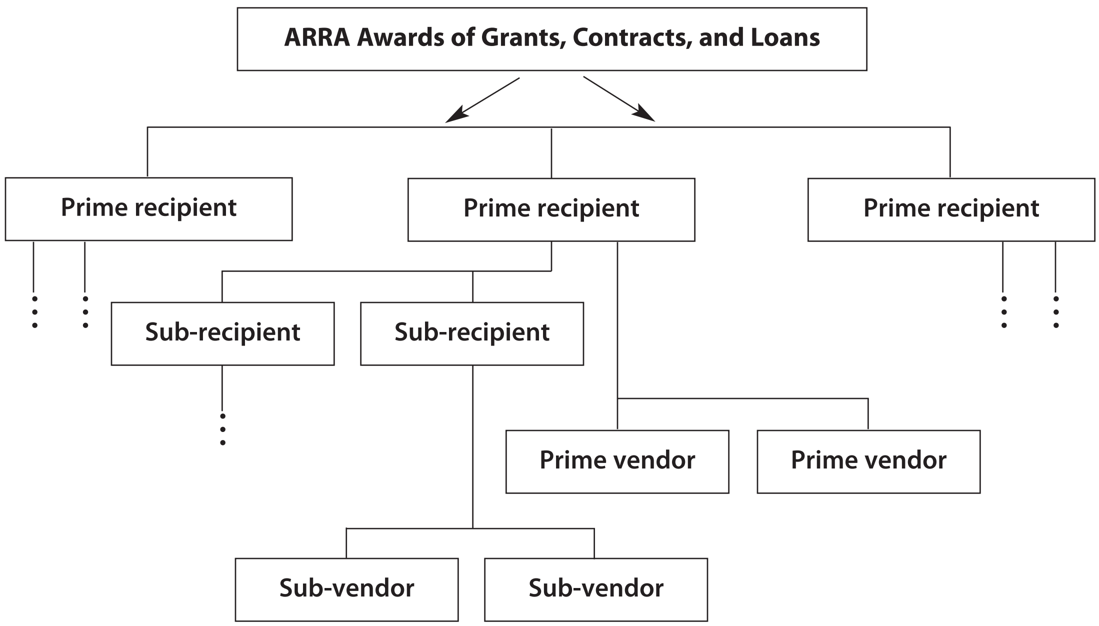
Award Hierarchy and Definition of a Job Saved/Created

The Recovery Accountability and Transparency Board (RATB) was created by the Act and one of the Board's responsibilities was to collect a wide variety of data from primary recipients, known as recipient reports.[10] Primary recipients are one of the four "recipient roles" established by the Board. The other three are sub-recipients, primary vendors, and sub-vendors. Figure 1 illustrates the relationship between the roles. Primary recipients receive award funds from grants, loans, or contracts. Sub-recipients receive stimulus funds though the primary recipients. Vendors and sub-vendors sell goods and services to primary recipients and sub-recipients, respectively.

The responsibility for filing quarterly reports rested with the primary recipients. One required survey field that each recipient had to complete was titled "Number of Jobs." The RATB (2010, p. 19) contains a description of that data field: "Jobs created and retained. An estimate of the combined number of jobs created and jobs retained funded by the Recovery Act during the current reporting quarter in the United States and outlying areas."[11] The field

Figure 1

Award Hierarchy of ARRA Grants, Loans, and Contracts Administered by the Federal Government



description goes on to say "For grants and loans, the number shall include the number of jobs created and retained by sub recipients and vendors." The Office of Management and Budget (OMB, 2009, p. 11) explains that, for grant and loan recipients, "the estimate of the number of jobs created or retained by the Recovery Act should be expressed as 'full-time equivalents' (FTE). In calculating an FTE, the number of actual hours worked in funded jobs are divided by the number of hours representing a full work schedule for the kind of job being estimated. These FTEs are then adjusted to count only the portion corresponding to the share of the job funded by Recovery Act funds."

A few examples are useful at this point. First, the Federal Highway Administration awarded a $342 million Highway Infrastructure Program grant to the Wisconsin Department of Transportation. There were no sub-recipients of this award. There were three primary vendors: one construction firm and two engineering firms. In the fourth quarter of 2009, the Wisconsin Department of Transportation reported that the award had directly created/saved 51.1 jobs. It also reported that the project was less than 50 percent completed.

As a second example, the U.S. Office of Elementary and Secondary Education awarded a $480 millon Education Fund grant to the state of Wisconsin, the primary recipient. The state of Wisconsin, in turn, distributed most of this money to over 400 local school districts, each of which was a sub-recipient of the original grant. In the fourth quarter of 2009, there were

10 sub-vendors on the grant. These were due to expenses created by some sub-recipients to businesses, such as Apple Inc. for computers. At that time, there were no primary vendors, indicating that the primary recipient did not directly buy from vendors. In that quarter, the state of Wisconsin reported that the award had directly created/saved 3,951.56 jobs. With respect to data quality, the recipient-reported jobs data have been scrutinized by state and federal auditors, congressional committees, media organizations, and private citizens.

A Government Accountability Office (GAO, 2009) report based on the third-quarter recipient reports of that year did find some questionable data entries. For example, roughly 4,000 of the more than 100,000 recipient reports showed “no dollar amount received or expended but included more than 50,000 jobs created or retained.” This 50,000 is not trivial, but it is relatively small compared with the 682,000 total jobs reported. In addition, there were “9,247 reports that showed no jobs but included expended amounts approaching $1 billion.” In total, this $1 billion represented less than 2.3 percent of the aid covered by these reporting requirements through the third quarter of 2009. The GAO report did not explore how many of these awards may have generated expenditures without creating jobs. Finally, the GAO found other reporting anomalies but stated that they were “relatively small in number.”

The GAO report (2009) concluded with four recommendations to the OMB to improve the consistency of data collection and reporting. On its website, the GAO posted that two of its recommendations were adopted by the OMB though a December 2009 OMB memo to ARRA fund recipients. One major point of the OMB memo was that “the recipients will no longer be required to make a subjective judgment on whether jobs were created or retained as a result of the Recovery Act. Instead, recipients will more easily and objectively report on jobs funded with Recovery Act dollars.” One of the two recommendations not adopted was moving to an hours worked, wages paid model instead of one of jobs created and saved. The other recommendation that was not adopted was that the “OMB continue working with federal agencies to provide or improve program-specific guidance to assist recipients, especially as it applies to the full-time equivalent calculation for individual programs.”

Since the data I consider are from the first-quarter 2010 report, the RATB and OMB changes resulting from the GAO report may have improved the quality of reports. Unfortunately, the GAO has not issued a comprehensive follow-up (to its November 2009 analysis) of the recipient-reported jobs data. On this point, a GAO report (2011) specifically regarding U.S. Department of Energy ARRA funding did find that “the quality of FTE data reported by recipients to FederalReporting.gov has improved over time.”

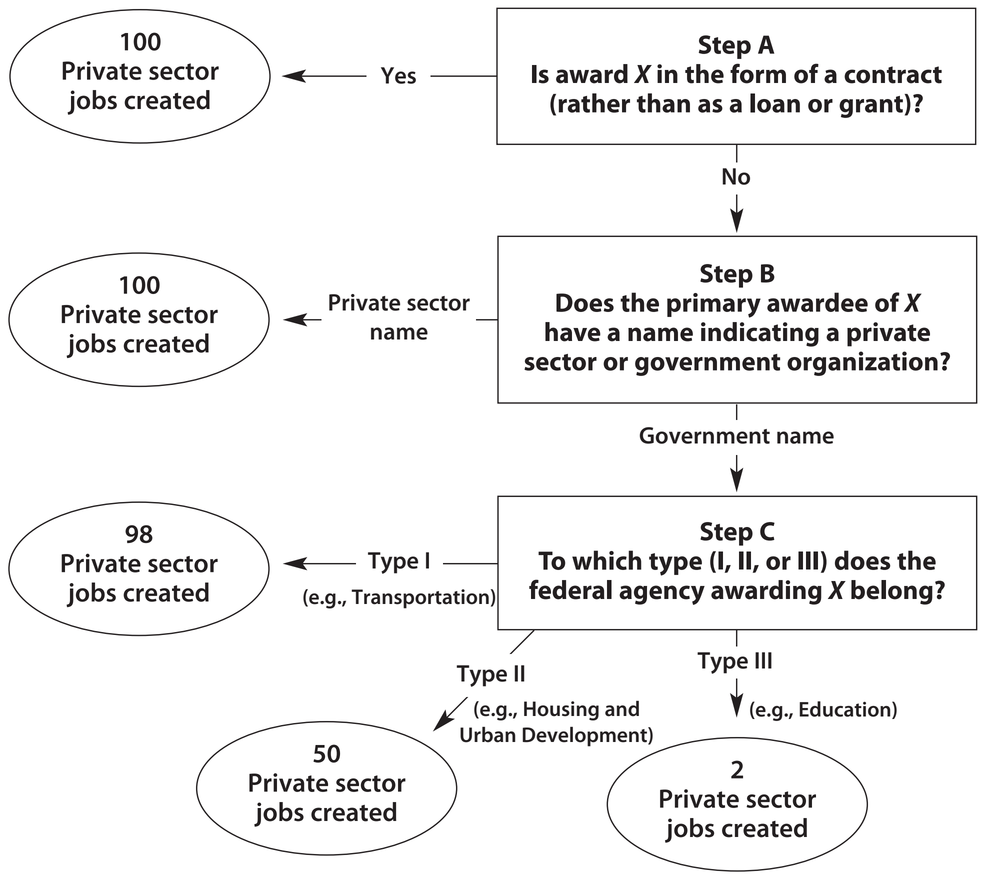
There have been relatively few cases of fraud in the recipient reports. Grabell (2012, p. 285) writes that the RATB received more than 7,500 complaints, which led to over 1,500 investigations. “Only about two hundred cases had resulted in criminal convictions, as of the fall of 2011.”

An Algorithm for Categorizing Job Types

Neither the recipients nor the ARRA oversight board categorizes jobs into the private sector or the government sector; therefore, I perform this task using a three-step procedure as diagrammed in Figure 2.

Figure 2

Algorithm for Determining Allocation of 100 Created/Saved Jobs Between Private Sector and Government



NOTE: See Tables A1 and A2 for common flags to indicate private and government primary awardees (applied in step B).

Step A. In the first stage (designated "Step A" in Figure 2), I assign all created/saved jobs that resulted from stimulus contracts (as opposed to grants and loans) as private sector jobs. My reason is as follows. Analysis of the contract data reveals that the stimulus contracts overwhelmingly were between an agency of the federal government and a private business.

Step B. In the second stage ("Step B"), I begin by sorting each remaining primary recipient into either a government organization or a private sector organization by using the recipient's name. Many recipients were assigned by searching within each name for "flags" (i.e., words and abbreviations) that indicate a government or private sector organization. For example, if "CITY OF" appears in a recipient's name, it is categorized as a government organization. If "CORPORATION" appears in a recipient's name, it is categorized as part of the private sector. Tables A1 and A2 in the appendix contain sample lists of flags used. If the primary recipient's name indicates a private sector organization, I assign all of this recipient's created/saved jobs to the private sector.

One choice I must make in this step is to define a government organization. An organization is treated as part of a government if operational control is by a person or persons serving in the role of a government official or by a person appointed by a government official or agency. The above definitions imply straightforwardly that teachers at public schools and employees at state agencies are part of the government sector. Public state universities, such as the University of Wisconsin, are treated as part of government as well because the operational control of a public state university may be in the hands of a private organization whose members are designated by a state government official. For example, a public university may be controlled by a board of regents (a private organization), but the regents are selected mainly by state government official(s), such as the state's governor.

This definition also clearly implies that private businesses are not part of the government sector. The ownership of a private company, and therefore its direct control, lies outside the hands of a government or government-appointed official. This is not to say that a nongovernment organization may not receive funds from the government. For example, although a private construction company may enter into a legal contract to create a highway for a state's department of transportation, the operational control of the organization is beyond the hands of government. Similarly, a charity, such as the United Way, may receive funding from a government, but my definition implies that it is not part of government. A private organization is any organization that is not part of government, which includes both businesses and nonprofits.

Approximately 1,750 recipients remained uncategorized after applying the above procedure. A research assistant and I split the task of manually assessing the status of each remaining recipient and assigning each to the private sector or government. This process almost always involved locating a website for the organization and reading its description and organization structure. After this manual assignment, 2.6 percent of created/saved jobs had not been categorized.

Step C. The third step ("Step C" in Figure 2) is an attempt to distinguish how many private sector jobs were created/saved through a grant or loan received by a primary recipient *if the recipient is a government organization*. Step C is necessary because a primary recipient can use its award to purchase goods or services from vendors and sub-vendors and also make subawards to sub-recipients. As explained above, only the primary recipient reports the jobs created/saved and this single number is the sum of jobs created/saved from the entire award. Vendors and sub-vendors are in the private sector and sub-recipients may also be in the private sector.[12] If it is assumed that all jobs created by awards for which the primary recipient was part of government are government jobs, it might cause an understatement of the number of private sector jobs created/saved.

To address this issue, I examine in turn each federal agency responsible for awarding Recovery Act funds. Based on the awarding federal agency, I allocate the created/saved jobs between the government and private sector, which amounts to assigning each agency a "private sector percentage." I use three different private sector percentages: 98 percent (mostly private sector, or type I); 50 percent (one-half private sector, or type II); and 2 percent (limited

Table 1

Percentages of Jobs Assigned to Private Sector*

Type I Mostly private sector (98%)	Type II Half private sector (50%)	Type III Limited private sector (2%)
Department of Transportation	Department of Labor	Department of Education
Environmental Protection Agency	Department of Housing and Urban Development	Department of Health and Human Services
Forest Service	Centers for Disease Control and Prevention	Department of Justice
U.S. Army Corps of Engineers		Administration on Aging
Department of Energy		

NOTE: This list contains a sample of federal funding agencies. See Appendix A for a complete list of all federal agencies that funded ARRA-created/saved jobs. *For primary recipients of loans and grants that are government organizations, according to funding federal agency.

private sector, or type III). Table 1 separates some of the federal awarding agencies into the three categories.[13]

For example, I classify projects funded by the U.S. Department of Transportation as being mostly in the private sector (i.e., type I). This classification was made by examining grant descriptions and created/saved jobs descriptions from the recipient reports. The largest number of these jobs came in the form of grants from the Federal Highway Administration to state and local government agencies, which in turn contracted construction companies. Only a small number of jobs were saved/created from employment at the state and local government agencies that oversaw the projects. As a result, 2 percent of the jobs were assigned to government and 98 percent were assigned to the private sector.[14]

As a second example, I classify projects funded by the U.S. Department of Justice as being mostly in the government sector (i.e., type III). The U.S. Department of Justice administered Justice Assistance Grants (JAG) to states and some localities. In reading many descriptions of projects and the jobs those projects created from the recipient reports, I observed that nearly all of the jobs were in law enforcement, the courts, and jail coverage. In my reading of grant descriptions and created/saved jobs descriptions, there was a relatively small number of jobs resulting from hiring vendors and also from sub-recipient grants to nonprofits. As a result, the Department of Justice jobs were classified as type III.

My explanation for these categorizations for each of the main federal awarding agencies appears in Appendix A. Table A3 shows the entire list of awarding agencies and their associated types.

Summarizing the procedure, suppose a particular primary recipient records creating 100 jobs. Then, follow the steps below:

Step A. If the primary recipient's award is a contract (as opposed to a grant or loan), then its 100 jobs are assigned to the private sector; otherwise, proceed to step B.

Table 2

Directly Created and Saved Jobs: Private Sector and Government

Federal department/ agency	Private sector jobs	Government jobs	Percentage of private sector jobs
All	162,567	504,784	24
Education	17,859	456,062	4
Transportation	32,569	626	98
Health and Human Services	20,397	11,390	64
Energy	21,501	460	98
Housing and Urban Development	10,993	9,175	55
Labor	8,738	7,967	12
Justice	17,29	13,065	12
Environmental Protection Agency	9,509	180	98
Agriculture	5,318	164	97

NOTE: These numbers are based on first-quarter 2010 recipient reports by ARRA recipients of contracts, grants, and loans. I was unable to assign 2.4 percent of all created/saved jobs to either the private sector or government sector.

Step B. If the primary recipient has a name indicating that it is in the private sector, then its 100 jobs are assigned to the private sector; otherwise, proceed to step C.

Step C. If the primary recipient's award has a government name, then X of its jobs are assigned to the private sector, where X depends on the federal agency that funds the award. The remaining 100 – X jobs are assigned to the government sector.

DIRECTLY FUNDED JOBS PRIMARILY IN STATE AND LOCAL GOVERNMENT

The top row labeled "All" in Table 2 shows the total number of jobs directly created/saved in the first quarter of 2010, which are then broken down into specific categories. These numbers and those later in the table do not include the 15,000 jobs (roughly 2 percent) that my categorization procedure could not assign as either private or public.

Table 2 shows that about 163,000 of the 667,000 total assigned jobs were in the private sector, which is 24.4 percent of all assigned jobs. This article's headline finding is that saved/created jobs were primarily in government.

As a sensitivity analysis, I consider an increase in the share of private sector workers in the three categories. For type I agencies, I increase the private sector share from 98 percent to 100 percent. For type II agencies, I increase the private sector share from 50 percent to 80 percent. For type III agencies, I increase the share from 2 percent to 12 percent. This modification implies that the percentage of private sector jobs created/saved was 31.2 percent, only slightly greater than the benchmark result. Thus, this adjustment to the model calibration does not overturn my main finding.

For completeness, I report the numbers of jobs by sector in a way that does not drop the small number of jobs that were not assigned. To do so, I construct these numbers by assuming that the fraction of private sector jobs among the unassigned jobs equals the fraction of private sector jobs among the assigned jobs. Under this assumption, the number of private sector jobs is 166,000 among the 682,000 total jobs reported as created/saved. Regardless of whether unassigned jobs are included or excluded, the numbers are nearly identical. I report this adjusted number in this article's abstract and introduction.

Table 2 breaks down the number and type of jobs for the nine largest job-creating/-saving federal agencies. First, the Department of Education created the most total jobs of any agency. These jobs were heavily weighted toward government.

The Department of Education administered several of the Act's largest grant programs, including the Education Stabilization Fund and the Government Services Fund. These grants were for elementary and secondary education. Elementary and secondary education jobs are mainly in government. Furthermore, since local and state government revenues fell dramatically during the recession, the governments used the grants to cover budget shortfalls and pay their workers.

Also, the Department of Education administered the ARRA Government Services grants. These helped state governments meet payrolls of non-education workers.

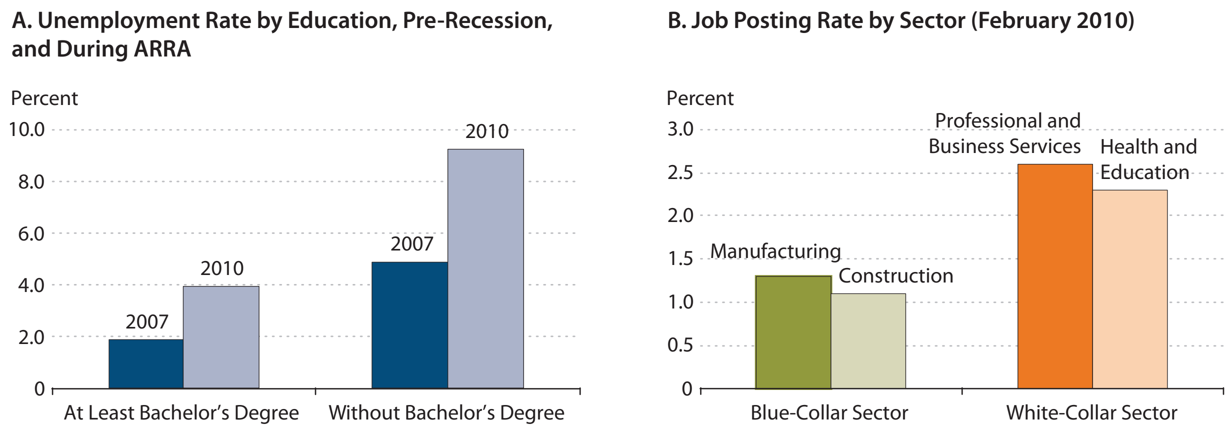
The second-largest job-creating agency was the Department of Transportation. It created private sector jobs almost exclusively; however, the number of government jobs funded by the Department of Education dwarfed the number of private jobs funded by the Department of Transportation by a 14-to-1 ratio. Agencies within the Department of Transportation that specifically oversaw projects included the Federal Highway Administration, the Federal Transit Administration, and the Federal Railroad Administration. The largest source of Department of Transportation jobs resulted from the hiring of highway construction companies by states using funds from the Federal Highway Administration.

Recall that step C in the algorithm required categorizing federal agencies by their intensity of private versus government job creation. While Appendix A gives detailed explanations for the categorization for the larger agencies, I do not provide explanations for the categorizations I made for the relatively smaller agencies. Changing these categorizations has little quantitative effect on the results. For example, switching the type III agencies (for which assignment explanations are not detailed in the appendix) to type I agencies would increase the percentage of private sector jobs by only 2 percentage points (i.e., increase the percentage from 24 percent to 26 percent).

My finding that job support through Recovery Act funds occurred primarily in the government sector may help explain why the U.S. unemployment rate did not fall as predicted.[15] By focusing on the government sector, the positive jobs effects of the Act were likely slanted toward the well educated. This is because roughly 49 percent of state and 47 percent of local government workers have at least a bachelor's degree, whereas for private sector workers this proportion is only 25 percent (see Greenfield, 2007). On the other hand, the labor market during the recession was much weaker for the less educated. In February 2010, the unemployment

Figure 3

Labor Market Conditions Before Recession and During Stimulus Period



NOTE: Unemployment rates are for February and are restricted to those with at least a high school diploma. The job posting rate is the number of job openings divided by employment in the corresponding sector. Data are deseasonalized from the Bureau of Labor Statistics.

rate was 4 percent among persons with a bachelor's degree and over 9 percent among persons without (Figure 3A).

Moreover, on the supply side, job availability (that is, the job posting rate) for white-collar work was more than double that for blue-collar work during the downturn (see Figure 3B). Thus, the stimulus spending studied in this article may have largely missed that part of the labor market in most need of assistance.[16]

A common perception is that the Recovery Act had a strong infrastructure component, suggesting that large numbers of private sector construction (and related industry) jobs should have been created. Job creation through infrastructure spending was complicated by several factors. First, school districts had the option of using stimulus education funds to make capital improvements, which would have implied more infrastructure job creation. In the 2010:Q1 recipient reports, there was almost no U.S. Department of Education spending on infrastructure; schools instead used their stimulus funds to maintain and add to their payrolls and provide pay raises. Second, many programs with funds committed only to infrastructure spent those funds very slowly. For example, only 56 percent of the Act's $48.1 billion transportation allocation had been spent at the two-year anniversary of the Act's passage.

CONCLUSION

A comprehensive understanding of the impact of the American Recovery and Reinvestment Act is far from complete; however, a growing body of research, taken together, appears to be forming a coherent picture of some of the effects of the stimulus. That is, while the stimulus was unsuccessful in creating/saving private sector jobs, it helped maintain and sometimes increase (i) state and local government services, and in turn public sector jobs, as well as (ii) transfer payments to the poor and unemployed. Each of the papers I describe next shows an element or elements that help shape the above picture of the stimulus.

Conley and Dupor (2013) use state-level data to conduct a cross-sectional study of the effect of stimulus jobs. They find that the Act resulted in a statistically significant increase in state and local government employment but not in private employment. Jones and Rothschild (2011), using their own surveys of grant, loan, and contract recipients, find that approximately one-half of the individuals filling positions directly created by Recovery Act funding were leaving other jobs.

Wilson (2012) uses state-level variation through an instrumental variables method to study the job effects of Recovery Act spending. His results concerning the private sector effect of the stimulus are mixed. He finds that overall employment increased as a result of the stimulus. Table 5 in his article breaks down the employment effect by sector. In one specification, he finds a statistically insignificant response of private sector employment to spending. In two others, the effect is positive and statistically significant.

Michael Grabell's (2012) book, *Money Well Spent? The Truth Behind the Trillion-Dollar Stimulus, the Biggest Economic Recovery Plan in History*, intended for a general audience, gives an excellent account of many aspects of the Recovery Act. It contains a detailed narrative account of why the design and execution of the stimulus led to slow and limited direct job creating/saving in the private sector, such as transportation, but rapid job creating/saving in the government sector. He also stresses the importance of the stimulus in funding social safety net programs.[17]

Several times per year since the beginning of the stimulus plan, the CBO has published a low-high interval estimate of the employment effects (combining direct and indirect jobs created/saved) of the program. In its first nine reports, the CBO projected that between 1.3 and 3.3 million persons would be employed in 2010 as a result of the Act; however, in its 10th report (November 2011), the CBO revised its estimate and reported that the Act may have created/saved as few as 650,000 jobs in 2010. Note that the CBO's jobs effect is estimated based on all components of the stimulus.

Examining the National Income and Product Accounts, Aizenman and Pasricha (2011, p. 5) find that "the federal fiscal expenditure stimulus in the U.S. during the great recession mostly compensated for the negative state and local stimuli associated with the collapsing tax revenue and the limited borrowing capacity of the states." Cogan and Taylor (2012) go even further, showing that net lending by state and local governments increased during the period they were receiving stimulus funds.

Ramey (2013), using several different specifications of structural vector autoregressions, shows that in response to an increase in government purchases, government employment rises while private sector employment falls or is unchanged.[18] ■

APPENDIXES

Appendix A: Classification of Agencies by Intensity of Private Sector Job Creation

As explained in the text, federal agencies awarding grants and loans to federal government organizations differ in the extent to which their funding generates creation and retention of private versus government jobs. If step C is reached, then I must make this assessment.[19] I designate three categories of federal agencies: type I—private-sector-intensive job creators; type II—equal division of private and government jobs; and type III—government-intensive job creators. An explanation for the assignment of the largest federal agencies to one of the three types follows.

Type I Agencies: Private-Sector-Intensive Job Creators

Department of Transportation (including Federal Transit Administration, Federal Highway Administration, Federal Railroad Administration)

My examination of job descriptions indicates that projects funded by the U.S. Department of Transportation resulted in mostly private sector jobs. The largest component of this funding came in the form of grants from the Federal Highway Administration to state and local government agencies, which in turn contracted construction companies. For example, in a Pavement Reconstruction and Added Capacity project by the California Department of Transportation, 289 persons were employed in widening lanes and shoulders on State Route 91 in Orange County. The corresponding recipient report's project description states "Jobs are created or retained in the construction and construction management industry such as laborers, equipment operators, electricians, project managers, support staff, inspectors, engineers, etc."

Environmental Protection Agency

My reading of the descriptions of jobs created/saved by the EPA is that most were in the construction industry. I did observe some descriptions associated with the retention of employees at state agencies charged with environmental policy.

Type II Agencies: Equal Division of Job Creators

Department of Housing and Urban Development

I classify half of the jobs created/saved by HUD as government jobs. Most of the larger awards went to state and local governments, which in turn hired or retained government employees to fill program positions. For example, the City of New York was the primary recipient of a $74 million award from HUD as part of the Homelessness and Rapid Re-Housing program. The city reported creating and saving 380 jobs; the description of these jobs included program directors, housing specialists, community liaisons, and outreach workers. The description suggests that these were either government jobs or jobs from nonprofit organizations that help the homeless.

Note that HUD-funded awards did create many private sector jobs. The ARRA Capital Fund supported many projects that modernized existing housing and increased the stock of public housing. For example, these projects funded construction managers, construction workers, and engineers.

The amounts paid by local and state government agencies to these private companies would appear as payments to vendors and sub-vendors. Given that the HUD funding led to both private sector and government jobs, it is classified as 50 percent private and 50 percent public.

National Institutes of Health

Many primary recipients of NIH awards (mainly grants) were nongovernment organizations. As such, these recipients' jobs were counted as private sector in step B of the procedure.[20] Among the awards to a governmental primary recipient, one of the largest in terms of job creation went to the University of Florida, which reported 32.16 jobs. Its job description field states:

> Recipient (UF) funded 11.4 FTE's in various research positions required to advance the work of this project. Our 6 sub-recipients have employed 20.76 FTE's in various positions that support the work being funded under the grant.

Because a substantial number of jobs were created at the sub-recipient level, I then determined whether the sub-recipients were in government. The sub-recipients were Scripps Research Institute, The Washington University, Cornell University Inc., Ponce Medical School Foundation Inc., and the Trustees of Indiana University. Only one of these entities is a government organization. As such, I conclude that the jobs created/saved were a mix of government and private sector, with the majority in the private sector.

Another large grant went to the Research Foundation for the State University of New York, which created 26.26 jobs. In general, I treat research foundations associated with public universities as part of government for reasons detailed in the text. The job created/saved field describes these positions:

> Postdoctoral Associate, Reimbursement to SUNY for faculty and staff time on research projects, Research Aide, Research Project Assistant, Research Scientist, Research Support Specialist, Research Technician I, Senior Postdoctoral Associate, Senior Research Support Specialist.

This description suggests that all or nearly all of the jobs were at the public university.

Next, the University of Miami received a grant resulting in 20.51 jobs, also one of the largest single job-creating NIH awards. The job created/saved field describes these as follows:

> Prime recipient funded a quality coordinator, 3 research associates, a sr. research associate, a clinical research ooordinator, a research assistant, a sr. research analyst, a plebotomist, a sr. manager of research support, a professor, an an associate professor. Subrecipients have funded a study nurse, a research scientist, a medical investigator, a deputy health officer, 9 nurse practitioners, a sr research associate, an assistant professor, 3 associate professors, a professor, a program coordinator, 2 physicians, and 5 research assistants, a site principal investigator, a site coordinator/research assistant, 2 reesarch assistant/counselors, a counselor, STI P.A. and a budet analyst. [*Spelling errors present in original report.*]

Upon examining the sub-recipients of the award, I found that there were both private sector and government organizations.

Overall, I conclude that NIH funding to government prime recipients directly created/saved a similar proportion of private sector and government jobs, and as such it is a type II funding agency.

Department of Labor

I classify the U.S. Department of Labor as a type II agency. The California Department of Employment Development was the primary recipient of a $489 million grant, the largest single award issued by this federal agency. The grant provided "employment services and work readiness and occupational skills training for unemployed adults and youth." The department reported creating 1,797 jobs with the description: "case managers and program support" and "hiring of youth program counselors and mentors." The award was dispersed among mostly city and county employment agencies. Thus, I infer that many jobs created were in government. A substantial fraction of the jobs created/saved also appear to be in the private sector, as both for-profit trade schools were used as sub-vendors on the grant and nonprofit organizations were sub-recipients of a fraction of the award.

The second-largest job-creating award went to the Ohio Department of Job and Family Services. The $138 million award created 1,275 jobs with job descriptions similar to those for the California award above. The Ohio award was largely divided among sub-awardees that were mainly employment departments of various counties. Because of this split of funds, the Department of Labor is classified as a type II agency.[21]

Finally, only 8.3 percent of expenditures went to vendors and sub-vendors, and only 2.4 percent of expenditures were made through federal contracts.

Type III Agencies: Government-Intensive Job Creators

Department of Education: Office of Elementary and Secondary Education

I classify the Department of Education as creating mostly government jobs. The largest part of this department's funding came as State Fiscal Stabilization Fund education grants. These grants were intended to mitigate and avoid state and local cutbacks in education as well as improve instruction.

In terms of parsing the jobs created by this funding, California's jobs descriptions from education awards are particularly useful. California's main Education Fund grant was a $4.39 billion award, of which it had spent $3.95 billion by the quarter I consider. It was California's largest job-creating grant, with 35,393 jobs created and saved; most importantly, its job description provided a useful quantitative breakdown of the jobs. In particular, its description listed 287 jobs as vendor jobs. Thus, less than 1 percent of the jobs created were vendor jobs. Of the remaining jobs, 16,208 were precollege teaching positions and 3,547 were nonteaching positions, including food service, bus drivers, teaching assistants, custodians, office staff, librarians, and instructional aides. The remaining jobs were at public postsecondary schools, the University of California system, the California State University system, and the California Community Colleges System.

I note that the Recovery Act allowed the education state grants to be used for infrastructure improvements to a state's schools. Significant private sector job creation could have come from large infrastructure investment. I do not see this in the data, at least through the quarter studied. In California, only $8.1 million of the $3.95 billion from this grant was spent on infrastructure.[22]

Ohio's main Education Fund grant was a $1.46 billion award, of which it had spent $523 million by the quarter I consider and created 8,465 jobs. The Ohio Department of Transportation provided me with a very detailed breakdown of the jobs created for Ohio's ARRA funds. It includes job descriptions from every school, self-reported by each school's staff.[23] Inspection of the breakdown indicates that nearly all of the jobs went to public school employees: teachers, aides, principals, librarians, and so on.

Department of Education: Office of Special Education and Rehabilitation Services

The U.S. Department of Education administered Special Education grants to states through its Office of Special Education and Rehabilitative Services. This office funded a total of 62,891 jobs.[24] My inspection of the job descriptions for a group of grants to these states indicates that nearly all were for professional and support staff in public education.

The State of California reported creating/saving 5,722 jobs from its Special Education grant. Its job description (listed as FTEs) follows:

> Jobs created or retained include 3164.86 classified jobs, 2360.90 certificated jobs, 193.81 vendor jobs, and 0.00 IHE [Institutions of Higher Education] jobs. Classified jobs include non-teaching positions such as food service, bus drivers, teacher assistants, custodians, office staff, librarians, and instructional aides for special education. Certificated jobs include teaching positions. Vendor jobs represent a variety of different types of jobs.

Nearly every sub-recipient on the grant was a city government, county government, or school district.

The Georgia Department of Education was awarded a Special Education grant of $314 million. It reported creating/saving 2,522 jobs. Its job description (listed as FTEs) states:

> Teachers (693.30); Aides & Paraprofessionals (1528.57); Clerical Staff (27.95); Interpreter (2.63); Technology Specialist (4.00); School Nurse (2.69); Physical Therapist (5.50); Teacher Support Specialist (55.47); Secondary Counselor (3.00); School Psychologist (22.33); School Social Worker (3.91); Family Services/Parent Coordinator (5.00); Bus Drivers (57.30); Other Management (21.07); Other Administration (89.79); Other Salaries & Compensation (11.38); Speech Language Therapist (2.95); Other (15.26)

Department of Justice

I classify jobs funded by the Department of Justice as mostly in the government sector (i.e., type III). The Department of Justice administered Justice Assistance Grants (JAG) to states and some localities. In reading many descriptions of projects and the jobs created by those projects from the recipient reports, I observe that most of the jobs are in law enforcement, the courts, and jail coverage.

For example, the state of Virginia's recipient report stated the funds "afforded the state legislature the opportunity to offset the budget cuts necessitated by declining state revenues.

The Recovery Act JAG funds will be used to provide funds to the Compensation Board for distribution to Sheriff's offices during state fiscal year 2010." Virginia distributed most of its award among a total of 121 city governments, county governments, and (a few) jails. It created/saved 1,789 jobs and listed the following description:

> $23.2 million in Federal funds were used this quarter to offset the State budget cuts. Deputies in 144 Sheriff's offices across the Commonwealth were able to retain employment. While the number of jobs retained may seem a little unusual, the entire $23.2 million dollars was expended in one quarter and not spread over the year as projected in our grant application.

In terms of jobs created, the second-largest grant from the Department of Justice was awarded to the Ohio Office of Criminal Justice Services. It reported creating/saving 398.84 positions (it did not report these as FTEs). The corresponding job description suggests that most of these jobs were in government:

> Description of ARRA JAG jobs created: Courts/prosecution/defense/civil attorney 10; Law enforcement 44; Info technology 5; Community/social/victim service 101; Training/technical assistance 6; Detention/probation/parole/comm corrections 38; Administrative/human resources 10; Construction 1; Policy/Research/Intelligence 11.
>
> Description of ARRA JAG jobs retained: Courts/prosecution/defense/civil attorney 24; Law enforcement 121; Info technology 1; Community/social/victim service 183; Training/technical assistance 1; Detention/probation/parole/comm corrections 51; Administrative/human resources 28; Construction 6; Policy/Research/Intelligence 12.

Hennepin County in Minnesota received a JAG award of $5.7 million and reported 88.9 jobs saved. The description of these jobs stated that they were "all law enforcement" and "all retained." As further evidence, only 3.6 percent of all expenditures was paid to vendors and subvendors.

Table A1

Examples of Strings Indicating ARRA Recipient as a Private Sector Organization or Business

JV	JOINT VENTURE
ARCHITEC	CONSTRUCT
FOUNDATION	CORPORATION
INCORPOR	CORP.
CONSTRUCTION	ENGINEERING
CO.	CHRISTIAN
CATHOLIC	COMPANY
PHARM	(INC)
LLC	(INC.)
L.L.C.	Inc
LTD	COMP.
UNITED WAY	GOODWILL
HABITAT FOR HUMANITY	BIG BROTHERS
BIG SISTERS	METHODIST
ASSOCIATES	BUILDERS
BETH ISRAEL	SEMINARY
PUBLIC/PRIVATE VENTURES	ONESTAR NATIONAL SERVICE COMMISSION
UNIVERSITY OF CHICAGO, THE	UNIVERSITY OF LA VERNE
UNIVERSITY OF NOTRE DAME DU LAC	UNIVERSITY OF ROCHESTER
UNIVERSITY OF SAN DIEGO	UNIVERSITY OF SOUTHERN CALIFORNIA

NOTE: These are 40 of the 542 strings used to identify private sector organizations and businesses used in step C in this article's algorithm.

Table A2

Examples of Strings Indicating ARRA Recipient as a Government Organization

CITY OF	PARISH
COUNTY	DEPART
DEPT	DEPARTMENT
AUTHORITY	AGENCY
TOWNSHIP	DISTRICT
HOUSING COMMISSION	RAPID TRANSIT
AIRPORT	OFFICE OF
VILLAGE	SECRETARY OF
BOARD OF	STATE
DISTRICT	TOWN OF
BOROUGH OF	DIVISION OF
GOVERNOR	ADMINISTRATION
TOWNSHIP	CITY HALL
MUNICIPALITY	HIGHWAY PATROL
PUBLIC SAFETY	COMMUNITY COLLEGE
PUBLIC INSTRUCTION	COMMONWEALTH
NEW JERSEY TRANSIT	TRIBE
TRIBAL COUNCIL	CHEROKEE NATION
SIOUX NATION	HOUSING COMMISSION
ATTORNEY GENERAL	NATIONAL GUARD

NOTE: These are 40 of the 430 strings used to identify government organizations and businesses used in step C in this article's algorithm.

Table A3

Federal Agencies Funding ARRA Grants and Loans: Numbers of Jobs Created/Saved in First Quarter of 2010 and Intensity of Private Sector Job Creation

Federal agency/subagency	No. of jobs created/saved	Agency type
Office of Elementary and Secondary Education	394,740	III
Office of Special Education and Rehabilitative Services	62,891	III
Department of Housing and Urban Development	19,061	II
Administration for Children and Families	17,828	II
Federal Highway Administration	16,769	I
Department of Labor	16,347	II
National Institutes of Health	16,199	II
Department of Energy	15,408	I
Department of Justice	15,152	III
Federal Transit Administration	12,922	I
Department of Education	11,667	III
Environmental Protection Agency	9,292	I
Corporation for National and Community Service	7,327	III
Health Resources and Services Administration	7,043	I
National Science Foundation	3,565	III
Federal Aviation Administration	1,992	I
Forest Service	1,546	I
Rural Utilities Service	1,366	I
National Endowment for the Arts	1,364	I
Assistant Secretary for Public and Indian Housing	1,219	I
Department of Health and Human Services	1,035	III
Federal Railroad Administration	836	I
Department of the Army	601	I
Bureau of Reclamation	562	II
Rural Housing Service	513	I
Maritime Administration	499	II
Centers for Disease Control and Prevention	495	II
U.S. Army Corps of Engineers-Civil Program Financing Only	487	I
Administration on Aging	487	I
Office of Science	392	III
Indian Health Service	345	II
Indian Affairs (Assistant Secretary)	334	II
National Oceanic and Atmospheric Administration	319	II
Community Development Financial Institutions	200	II
National Telecommunication and Information Administration	197	II
Employment and Training Administration	179	II
Food and Nutrition Service	159	II
Economic Development Administration	144	II
Department of the Interior	139	I
National Aeronautics and Space Administration	135	II
Rural Business-Cooperative Service	123	II
U.S. Fish and Wildlife Service	110	II
Federal Emergency Management Agency	107	II

NOTE: Type I, private-sector-intensive; type II, equal division between private and government sectors; type III, government-sector-intensive. Excludes federal agencies creating fewer than 50 jobs.

SOURCE: Recipient reports available at http://recovery.gov.

Appendix B: Recovery Board Definition of a Job Created or Retained

The RATB's definition of a saved (i.e., retained) job or a created job is dependent on the reporting instructions provided to recipients. These instructions differed slightly between recipients of contracts and recipients of grants or loans. I discuss the grants and loans rules first.

FederalReporting.gov, along with the OMB, provided instructions to recipients of grants and loans for calculating the number of jobs created or saved. While recipients have several pages of guidance for calculating a jobs created/retained number, the following paragraph from OMB (2009) summarizes the key aspects:

> The estimate of the number of jobs created or retained by the Recovery Act should be expressed as "full-time equivalents" (FTE). In calculating an FTE, the number of actual hours worked in funded jobs are divided by the number of hours representing a full work schedule for the kind of job being estimated. These FTEs are then adjusted to count only the portion corresponding to the share of the job funded by Recovery Act funds. Alternatively, in cases where accounting systems track the billing of workers' hours to Recovery Act and non-Recovery Act accounts, recipients may simply count the number of hours funded by the Recovery Act and divide by the number of hours in a full-time schedule.

The OMB (2010) gives very similar instructions for reporting jobs created and retained to recipients of ARRA contracts. One distinction made for contractors that does not hold for grant and loan recipients is as follows: "The definition applies to prime contractor positions and first-tier subcontractor positions where the subcontract is $25,000 or more." Moreover, all primary recipients were instructed to not report the employment impact on materials suppliers and central service providers (referred to as "indirect" jobs) or on the local community (referred to as "induced" jobs).

OMB (2010) goes on to say:

> "Job Created" means those new positions created and filled, or previously existing unfilled positions that are filled, that are funded by the American Recovery and Reinvestment Act of 2009 (Recovery Act). This definition covers only positions established in the United States and outlying areas (see definition in FAR 2.101). The term does not include indirect jobs or induced jobs. "Job Retained" means those existing filled positions that are funded by the American Recovery and Reinvestment Act of 2009 (Recovery Act).

NOTES

[1] President Barack Obama signed the Recovery Act during his first month in office. For an early outline of the plan, see Summers (2008). For an early projected impact of the plan, see Bernstein and Romer (2009).

[2] See CBO (2011).

[3] At the time this article was written, the data were available for download at the website Recovery.gov. The site also contained a data user guide; see Recovery Accountability and Transparency Board (2009).

[4] Unless otherwise stated, jobs reported in the text as saved/created are FTEs and are shown as rounded numbers. As a result, they may differ from the exact values in the tables.

[5] At the one-year mark, the Act's funding directed toward grants, contracts, and loans totaled $83.6 billion. Note that there are no recipient-reported jobs data for other ARRA spending because such spending does not create/save jobs directly.

[6] U.S. House of Representatives (2011).

[7] See page 2 of Bernstein and Romer (2009).

[8] Note that Bernstein and Romer's number is not directly comparable with my results because they did not include what they call both direct and indirect employment creation in this quote. I have not found precise definitions of "direct" versus "indirect" job creation in the existing literature.

[9] The above numbers refer to point estimates from the respective studies.

[10] Section 1512 of the ARRA.

[11] The data description was slightly different in the first quarterly reporting period. At that time, recipients were asked to construct a jobs number based on whether a given job would have existed were it not for the Recovery Act. The reason for the change was the subjective nature of the original question. I did not use the first-quarter responses in my study.

[12] The most likely reason that a sub-recipient was in the private sector was because it is a nonprofit organization.

[13] In designing the entire procedure, I tried to minimize (to the greatest amount possible) the discretion (i.e., "judgment calls") needed to categorize jobs as being in the private or government sectors. Step C required the most discretion on my part since the time that would be required to partition individual jobs into the private or government sectors on an award-by-award basis was prohibitive.

[14] See Appendix A for a description of type II projects.

[15] See Figure 1 in Bernstein and Romer (2009).

[16] If the created and saved jobs were in relatively healthy parts of the job market, then one would expect to see a substantial number of Recovery Act job takers coming from other jobs. To this point, Jones and Rothschild (2011) survey findings have shown that approximately one-half of the individuals filling positions directly created by the ARRA were leaving other jobs.

[17] Note that Grabell's overall assessment is that the stimulus may have created/saved millions of jobs (direct plus indirect), which is consistent with some estimates by others, such as the president's Council of Economic Advisers (2010).

[18] The horizon of Ramey's study does not include the ARRA period.

[19] Recall that step C is reached if the award is not a contract and the primary recipient's name indicates it is a non-federal government organization.

[20] Recall that in step B, if the primary recipient has a name indicating that it is in the private sector, then its jobs are assigned to the private sector.

[21] The distinction between government and nongovernment can become muddled with respect to employment services. For example, the largest sub-recipient of the Ohio award is the Central Ohio Workforce Investment Corporation. It has partners in the private sector, such as Goodwill of Ohio. It satisfies this article's definition of a type II agency because, as its website states, its "Board [is appointed] by the Mayor of the City of Columbus and

the Franklin County Board of Commissioners, in conjunction with recommendations made by the Greater Columbus Chamber of Commerce."

[22] The recipient survey includes several questions about infrastructure spending in particular.

[23] These data are a matter of public record and are available from the author on request.

[24] This total includes Special Education and other grants that it administered.

REFERENCES

American Recovery and Reinvestment Act. Public Law 111-5, 111th Congress, February 17, 2009; http://www.gpo.gov/fdsys/pkg/PLAW-111publ5/pdf/PLAW-111publ5.pdf.

Aizenman, Joshua and Pasricha, Gurnain K. "The Net Fiscal Expenditure Stimulus in the U.S., 2008-9: Less than What You Might Think, and Less than the Fiscal Stimuli of Most OECD Countries." *The Economists' Voice,* June 2011, *8*(2), pp. 1-6.

Bernstein, Jared and Romer, Christina. "The Job Impact of the American Recovery and Reinvestment Plan." Council of Economic Advisers and Office of the Vice President Elect, January 9, 2009; http://otrans.3cdn.net/45593e8ecbd339d074_l3m6bt1te.pdf.

Blinder, Alan S. *After the Music Stopped: The Financial Crisis, the Response, and the Work Ahead*. New York: Penguin Books, 2013.

Cogan, John F. and Taylor, John B. "What the Government Purchases Multiplier Actually Multiplied in the 2009 Stimulus Package," in Lee E. Ohanian, John B. Taylor, and Ian Wright, eds., *Government Policies and the Delayed Economic Recovery*. Chap. 5. Stanford, CA: Hoover Institution Press, 2012, pp. 85-114.

Cohen, Lauren; Coval, Joshua and Malloy, Christopher. "Do Powerful Politicians Cause Corporate Downsizing?" *Journal of Political Economy*, December 2011, *119*(6), pp. 1015-60.

Congressional Budget Office. "The Budget and Economic Outlook: Fiscal Years 2011 to 2021." January 2011; http://www.cbo.gov/sites/default/files/cbofiles/ftpdocs/120xx/doc12039/01-26_fy2011outlook.pdf.

Conley, Tim and Dupor, Bill. "The American Recovery and Reinvestment Act: Solely a Government Jobs Program?" *Journal of Monetary Economics*, July 2013, *60*(5) pp. 535-49.

Council of Economic Advisers. "The Economic Impact of the American Recovery and Reinvestment Act of 2009: Third Quarterly Report." Executive Office of the President, April 14, 2010; http://www.whitehouse.gov/sites/default/files/microsites/CEA-3rd-arra-report.pdf.

Gandel, Stephen. "After One Year, a Stimulus Report Card." *Time*, February 17, 2010; http://content.time.com/time/specials/packages/article/0,28804,1964765_1964764_1964758,00.html.

Government Accountability Office. "Recovery Act: Recipient Reported Jobs Data Provide Some Insight into Use of Recovery Act Funding, but Data Quality and Reporting Issues Need Attention." GAO-10-223, November 19, 2009; http://www.gao.gov/assets/300/298632.pdf.

Government Accountability Office. "Recovery Act: Progress and Challenges in Spending Weatherization Funds." GAO-12-195, December 16, 2011; http://www.gao.gov/assets/590/587064.pdf.

Grabell, Michael. *Money Well Spent? The Truth Behind the Trillion-Dollar Stimulus, the Biggest Economic Recovery Plan in History*. New York: PublicAffairs, 2012.

Greenfield, Stuart. "Public Sector Employment: The Current Situation." Working paper, Center for State and Local Government Excellence, 2007.

Jacobson, Louis. "Government Jobs vs. Private Jobs: Which Help the Economy More?" *Tampa Bay Times*, February 23, 2013; http://www.tampabay.com/news/business/economicdevelopment/government-jobs-vs-private-jobs-which-help-the-economy-more/1276248.

Jones, Garrett and Rothschild, Daniel M. "Did Stimulus Dollars Hire the Unemployed? Answers to Questions about the American Recovery and Reinvestment Act." Working Paper No. 11-34, Mercatus Center, George Mason

University, September 2011; http://mercatus.org/sites/default/files/publication/Did_Stimulus_Dollars_Hire_The_Unemployed_Jones_Rothschild_WP34.pdf.

Office of Management and Budget. "Memorandum for the Heads of Executive Departments and Agencies. Updated Guidance on the American Recovery and Reinvestment Act—Data Quality, Non-Reporting Recipients, and Reporting of Job Estimates." Executive Office of the President, December 18, 2009; http://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_2010/m10-08.pdf.

Office of Management and Budget. "Recovery FAQs for Federal Contractors on Reporting." July 2, 2010; http://www.whitehouse.gov/omb/recovery_faqs_contractors.

Ramey, Valerie A. "Government Spending and Private Activity," in Alberto Alesina and Francesco Giavazzi, eds., *Fiscal Policy after the Financial Crisis*. Chap. 1. Chicago: University of Chicago Press, 2013, pp. 19-55.

Recovery Accountability and Transparency Board. "Recovery.gov: Download Center User Guide." Recovery.Gov, 2009; http://www.recovery.gov/arra/FAQ/OtherDLFiles/Download%20Center%20User%20Guide.pdf.

Recovery Accountability and Transparency Board. "Recipient Reporting Data Model v4.0: Final Production Release, for Quarter Ending September 30, 2010." Recovery.Gov, 2010; http://billdupor.weebly.com/uploads/2/2/8/0/22808472/fedrptgdatamodel_v4.0.pdf.

Shapiro, Matthew D. and Slemrod, Joel. "Did The 2001 Tax Rebate Stimulate Spending? Evidence From Taxpayer Surveys," in James M. Poterba, ed., *Tax Policy and the Economy*. Volume 17. Cambridge, MA: MIT Press, 2003, pp. 83-110.

Shapiro, Matthew D. and Slemrod, Joel. "Did the 2008 Tax Rebates Stimulate Spending?" *American Economic Review*, May 2009, *99*(2), pp. 374-79.

Summers, Larry. "The Economic Recovery PlanPolicy Work," in a memo to President-Elect Obama, December 15, 2008; http://s3.documentcloud.org/documents/285065/summers-12-15-08-memo.pdf.

U.S. House of Representatives. "Economists' Statement to President Obama." [Letter from John A. Boehner to President Barack Obama.] February 13, 2011; http://www.speaker.gov/sites/speaker.house.gov/files/UploadedFiles/Boehner-Obama-Letter-2-13-11-1.pdf.

Wilson, Daniel J."Fiscal Spending Jobs Multipliers: Evidence from the 2009 American Recovery and Reinvestment Act." *American Economic Journal: Economic Policy*, August 2012, *4*(3), pp. 251-82.

Representative Neighborhoods of the United States

Alejandro Badel

Many metropolitan areas in the United States display substantial racial segregation and substantial variation in incomes and house prices across neighborhoods. To what extent can this variation be summarized by a small number of representative (or synthetic) neighborhoods? To answer this question, U.S. neighborhoods are classified according to their characteristics in the year 2000 using a clustering algorithm. The author finds that such classification can account for 37 percent of the variation with two representative neighborhoods and for up to 52 percent with three representative neighborhoods. Furthermore, neighborhoods classified as similar to the same representative neighborhood tend to be geographically close to each other, forming large areas of fairly homogeneous characteristics. Representative neighborhoods seem a promising empirical benchmark for quantitative theories involving neighborhood formation. (JEL R2, D31, D58, J24)

Federal Reserve Bank of St. Louis *Review*, Second Quarter 2014, *96*(2), pp. 147-72.

Racial segregation is a striking trait of U.S. cities. Iceland, Weinberg, and Steinmetz (2002) report that 64 percent of the black population would have needed to change residence for all U.S. neighborhoods to become fully integrated in the year 2000. Income differences across neighborhoods have also been well documented. Wheeler and La Jeunesse (2007) report that between-neighborhood inequality in 2000 represented around 20 percent of overall annual household income inequality in Census data. The variation in housing prices across neighborhoods has also been the focus of a large literature.[1]

This article attempts to summarize the landscape of U.S. cities using a small number of representative neighborhoods. The motivation for this effort is twofold. On the one hand, a clear and concise characterization of the American urban landscape may be useful in the construction of theories involving neighborhood formation. On the other hand, a simple representation can be used to impose empirical discipline on quantitative models with a small number of locations. These types of models are important since they can address complex dynamic issues such as the interaction between neighborhood formation and human capital accumulation without becoming computationally infeasible (see, for example, Fernandez and

Alejandro Badel is an economist at the Federal Reserve Bank of St. Louis. The author thanks Christopher Martinek, Brian Greaney, and Brian Bergfeld for research assistance and Juan Sánchez for useful comments.

Rogerson, 1998). Here it is important to highlight that another part of the urban landscape consists of household heterogeneity *within* neighborhoods (see Ioannides, 2004). This article focuses exclusively on variation *across* neighborhoods.

The empirical strategy consists of applying a clustering algorithm to Census 2000 data describing U.S. neighborhoods. The *K*-means clustering algorithm is used here. This algorithm attempts to classify neighborhoods in such a way that neighborhoods within a cluster are similar to one another and dissimilar with respect to neighborhoods in other clusters. The aggregate of all neighborhoods in each cluster is interpreted as a representative neighborhood.

The rest of the article proceeds as follows. The next section describes the data, and the following section explains the clustering algorithm. Subsequent sections describe the clustering results and the representative-neighborhood characterization. These descriptions are followed by a section with robustness exercises. The final section provides conclusions and closing remarks.

DATA

Data for this study are from the 2000 Census of Population and Housing Summary File 3 (SF3) (U.S. Census Bureau, 2000). The SF3 contains geographically coded summary statistics at various levels of spatial aggregation.

This study focuses on the Census-tract level of geographic aggregation. Census tracts are small geographic subdivisions of the United States. According to the Census Bureau, tract boundaries are defined with the goal of obtaining areas containing demographically and economically homogeneous populations of about 4,000 people. These tract features are obviously desirable for classifying neighborhoods into distinct types.

The set of variables used as Census counterparts for income, racial composition, and house prices is described next. Table 1 defines these variables in terms of SF3 variable names.

Variables

Income (*Y*). Two measures of a tract's income are used. First, a tract's *labor* income (earnings hereafter) is measured as the log of average household earnings in the tract. Second, a tract's total income is measured as the log of average household income in the tract.

Racial Composition (*R*). The measure of racial composition used here is the fraction of white households in the tract. This fraction is obtained as the number of non-Hispanic white households divided by the total number of households in a Census tract.

Price of Housing Services (*P*). Three variables in the dataset can be used to construct measures of the price of housing services: median gross rent, median house value, and median owner costs (see the appendix for details). These variables are measures of housing expenditures. Since expenditures are the product of quantity and price, log expenditures equal the sum of a log price component and the log number of units consumed. The price component is isolated here by regressing the log of the median expenditure measure against a set of house

Table 1

Variable Definitions

Variable	Definition (Census code)
Fraction of black HHs	p151b001/(p151a001+…p151g001)
Fraction of non-Hispanic white HHs	p151i001/(p151a001+…p151g001)
Average tract HH earnings	p067001/p058001
Average tract HH income	p054001/p052001
Average white HH income	p153i001/p151i001
Average black HH income	p153b001/p151b001
Average both races HH income	(P153a001+...P153g001-p153i001-p153b001)
	/(P151a001+...P151g001-p151i001-p151b001)
Median gross rent	H063001
Median value (owner-occupied)	H085001
Median selected monthly owner costs	H091001 (owner-occupied with mortgage)
Median number of rooms in unit	H027002 (owner), H027003 (renter)
Distribution of number units in structure	H032003-012/H032002 (owner),
	H032014-023/H032013 (renter)
Median year structure built	H037002 (owner), H037003 (renter)
Distribution of number of bedrooms	H042003-008/H042002 (owner),
	H042010-015/H042009 (renter)
Fraction with telephone service	H043003/H043002 (owner),
	H043020/H043019 (renter)
Fraction with plumbing facilities	H048003/H048002 (owner),
	H048006/H048005 (renter)
Fraction with kitchen facilities	H051003/H051002 (owner),
	H051006/H051005 (renter)
Distribution of heating fuel	HCT010003-011 (owner),
	HCT0010013-021 (renter)
Distribution of time to work	P031003-014/P031002
Fraction of population in group quarters	P009025/P0009001

NOTE: HH, household.

SOURCE: Census Bureau. 2000 Census of Population and Housing—Summary File 3, Technical Documentation, released September 2002.

Table 2

Definition of Variable Configurations

Name	Income	Racial composition	Price of housing services
ben	Log household earnings	% Non-Hispanic whites	Clean IRV (owners)
inc	Log household income	% Non-Hispanic whites	Clean IRV (owners)
prent	Log household earnings	% Non-Hispanic whites	Clean rent (renters)
pcost	Log household earnings	% Non-Hispanic whites	Clean owner's cost (owners)

NOTE: Implicit rental value (IRV) is defined as a percentage of a home's market value.

characteristics and using the residual from this regression as the measure of the price of housing services.[2]

The benchmark measure of (Y, R, P) is composed of the log of mean earnings, the fraction of white households, and the "clean" log value of housing for owners. Results for alternate configurations after replacing one of the variables with an alternative measure are also presented. This changing-one-variable-at-a-time strategy results in three additional sets of variables. Each configuration is denoted by the name of the variable that changes with respect to the benchmark configuration (see Table 2 for the variables in each variable configuration).

Sample Selection

The baseline sample aims to provide a comprehensive picture of the distribution of income, racial composition, and house prices in U.S. metro areas. Metropolitan statistical areas (MSAs) with populations of at least 1 million are considered. Since the focus is on the black and non-Hispanic white populations, the sample is further restricted to MSAs where at least 10 percent of the population is black.

Within each selected MSA, the sample is restricted to Census tracts where less than 50 percent of the population reports being neither black nor non-Hispanic white. To guarantee the exclusion of rural areas, only the Census tracts with at least 100 people per square kilometer are retained.[3] Attention is also restricted to tracts with at least 200 households and no more than 25 percent of the population living in group quarters.[4]

Application of these sample selection criteria results in a set of 28 MSAs in 25 states including 80.7 million people and 17,815 Census tracts. The largest MSA in the sample is New York-Northern New Jersey-Long Island, with 3,850 tracts; the smallest is Raleigh-Durham-Chapel Hill, with 157 tracts. Table 3 presents the number of observations deleted by each criterion. Table 4 lists some summary statistics of the final sample. The section on robustness compares the results obtained under the baseline sample with those obtained under four variations of the sample selection criteria.

Table 3

Sample Selection Criteria

Criteria	Observations dropped	Total observations
Initial without missing values		50,167
MSA population less than 1 million	14,397	35,770
MSA with less than 10% black HHs	14,244	21,526
Population density less than 100/sq km	1,785	19,741
Other race more than 50%	1,421	18,320
Tract with less than 200 HHs	226	18,094
Institutionalized population more than 25%	279	17,815

NOTE: Each observation corresponds to a Census tract. HH, household.

Table 4

Descriptive Statistics: Main Variables

Variable	Mean	SD	5th Percentile	95th Percentile
Fraction black HHs	0.23	0.32	0	0.95
Fraction white HHs	0.66	0.32	0.01	0.97
Fraction other race HHs	0.10	0.11	0.01	0.35
Average tract HH income ($)	63,921	33,981	28,178	125,278
Average tract HH income ($, blacks)	55,927	43,712	20,508	117,500
Average tract HH income ($, whites)	66,413	36,393	26,702	130,605
Average tract HH income ($, other races)	61,015	37,956	21,550	124,834
Median IRV* ($)	13,372	5,593	6,748	22,571
Median gross rent* ($)	8,806	2,413	5,782	12,743
Median selected owner costs* ($)	15,415	3,795	10,323	21,772
Tract population	4,427	2,265	1,536	8,403
Number of HHs in tract	1,701	890	573	3,300
Population density (population/sq km)	3,391	6,150	192	13,605
Fraction of population in group quarters	0.01	0.03	0	0.08

NOTE: HH, household; IRV, implicit rental value. *Statistics reported controlling for certain factors via linear regression; see Data section for details.

CLUSTERING ALGORITHM

Cluster analysis attempts to classify a large set of objects into a small number of groups (clusters). A perfect classification is obtained if the large set is composed of a small number of groups of identical objects. For example, a dataset composed of only zeros and ones can be perfectly classified with two clusters.

A common clustering method consists of minimizing a square error (SE) criterion. The method used is known as the K-means algorithm and creates a partition of a set containing I objects into K mutually exclusive subsets (where $I \geq K$). How? Suppose each element $i \in I$ is described by the vector x_i. The algorithm searches for a partition of I into subsets $\{C_1, C_2, C_3, \ldots, C_k, \ldots, C_K\}$ that minimizes the within-cluster variation of x_i (or the SE) around each group's centroid c_k. The centroid c_k of cluster C_k is usually taken to be the vector of averages of x_i taken over all elements i belonging to the cluster C_k:

$$SE = \sum_{k} \sum_{i \in C_k} \omega_i \left(\mathrm{x}_i - c_k\right) \cdot \left(\mathrm{x}_i - c_k\right),$$

where ω_i is a weighting factor equal to the number of households in each tract.

Conceptually, the optimal partition could be found by computing the SE for every possible partition of I and then choosing the one that produces the smallest SE. In practice, the search needs to be conducted with a heuristic algorithm known as "iterative relocation."[5] A cluster resulting from iterative relocation has two desirable properties. First, each cluster has a centroid, which is the mean of the objects in that cluster. Second, each object belongs to the cluster with the nearest centroid. On the downside, this type of algorithm does not guarantee finding the optimal partition, and its outcome depends on the initial partition. For all clustering exercises reported here, the clustering procedure is applied 10 times using random starting values, and the cluster that minimizes the SE is reported.

Normalization of Data

A cluster's outcome is sensitive to the relative scaling of variables that describe each tract. One solution to this problem is to normalize each component of x_i to have a sample mean of 0 and a sample variance of 1. This method is referred to as z score normalization in what follows.[6] An alternative normalization method, based on the Mahalanobis transformation, accounts for the correlations across components of x_i. This method normalizes the data by the inverse of the sample covariance matrix of x_i, $\hat{\Omega}^{-1}$. In this case, SE becomes the standard error of the mean (SEM):

$$SEM = \sum_{k} \sum_{i \in C_k} \omega_i \left(\mathrm{x}_i - \mathrm{c}_k\right)' \hat{\Omega}^{-1} \left(\mathrm{x}_i - \mathrm{c}_k\right).$$

A comparison of selected results using z score and Mahalanobis normalizations is provided.

Table 5

Cluster Compactness (percent)

Variable configuration	$K = 2$	$K = 3$	$K = 4$	$K = 5$	$K = 6$
z Score normalization					
ben	37	50	57	62	66
inc	37	52	58	63	68
prent	34	52	58	63	67
pcost	36	50	57	62	66
Mahalanobis normalization					
ben	26	43	53	58	62
inc	26	44	54	58	63
prent	26	43	54	59	62
pcost	26	43	54	58	62
Average	31	47	56	60	65

NOTE: This statistic corresponds to the percentage of (Y, R, P) sum of variances explained by between-cluster variation.

RESULTS

This section describes the results of the clustering exercise. The main results are obtained by applying the clustering algorithm once to the full sample of neighborhoods from all MSAs. Alternate results obtained by applying the algorithm separately to each MSA are reported in the "Regional Stability" subsection.

Cluster Validity

How much of the variance can be captured by a clustering representation? One way to address this question is to assess the "compactness" of a cluster.[7] I use an intuitive indicator of compactness to address the validity of the clusters obtained and complement it with a visual summary of the distribution of (Y, R, P) within and between clusters.

The compactness indicator compares the SE from the clustering algorithm with the overall variability of x_i with respect to the vector of sample means, c.[8] In what follows, this measure is referred to as R^2 because of its mechanical similarity to the familiar concept from standard econometric analysis:

$$R^2 = 1 - \frac{SE}{\sum_k \sum_{i \in C_k} \omega_i (x_i - c) \cdot (x_i - c)}.$$

A value of $R^2 = 1$ means that the data consist of K types of identical elements. Table 5 presents the R^2 values obtained for $K = 2, 3, 4, 5, 6$ and each of the selected variable configurations using z score and Mahalanobis normalizations.

Not surprisingly, compactness increases with K. For $K = 2$, R^2 averages 31 percent across all variable configurations and normalizations and its maximum is 37 percent. The average increases to 47 percent when $K = 3$ and increases further to 65 percent as K increases from 3 to 6. Thus, most of the gains in explanatory power occur at $K = 2$ and $K = 3$. These clusterings provide a reasonable degree of compactness while maintaining an acceptable level of complexity. With $K \geq 4$, the complexity becomes substantially greater without a significant increase in explanatory power.

Figures 1 and 2 show a variety of statistics regarding the distribution of (Y, R, P) within and between clusters for $K = 2$ and $K = 3$, respectively. These plots reflect a large degree of similarity across different variable configurations measuring (Y, R, P) and large differences in the distributions of each variable across clusters.

For instance, consider the first column of Figure 1. The blue boxes depict the interquartile range of the distribution of racial configuration (i.e., the fraction of white households in the neighborhood) in each cluster. For all rows, the interquartile ranges of each cluster do not intersect. Average incomes show a similar result (see the second column). In contrast, all interquartile ranges for the average price of housing services of the two clusters intersect, although the central tendency is the same as for income.

The brackets in each plot in Figure 1 represent the range between the 5th and 95th percentiles of each distribution. For the fraction of white residents, the 95th percentile of neighborhoods of type *I* is below the median of neighborhoods of type *II* (depicted as the center of the corresponding blue box) and is also below the mean (depicted by a vertical solid line). For all variable configurations, the 5th percentile of neighborhood *II* is above the mean and the median of neighborhood *I*.

Figure 2 shows the $K = 3$ case. As shown in the first column, the separation of the distributions of racial configuration across clusters becomes larger between neighborhoods of type *A* and neighborhoods of type *B* and *C* than it is between neighborhoods of type *I* and *II* for $K = 2$. In turn, the distributions in neighborhoods of type *B* and *C* overlap substantially. The second column shows a different picture for income: Income distribution in neighborhood *C* is separated from those in neighborhoods *A* and *B*, while those in neighborhoods *A* and *B* exhibit significant overlap. The third column shows that patterns for distributions of house prices behave more like the patterns for income than those for race.

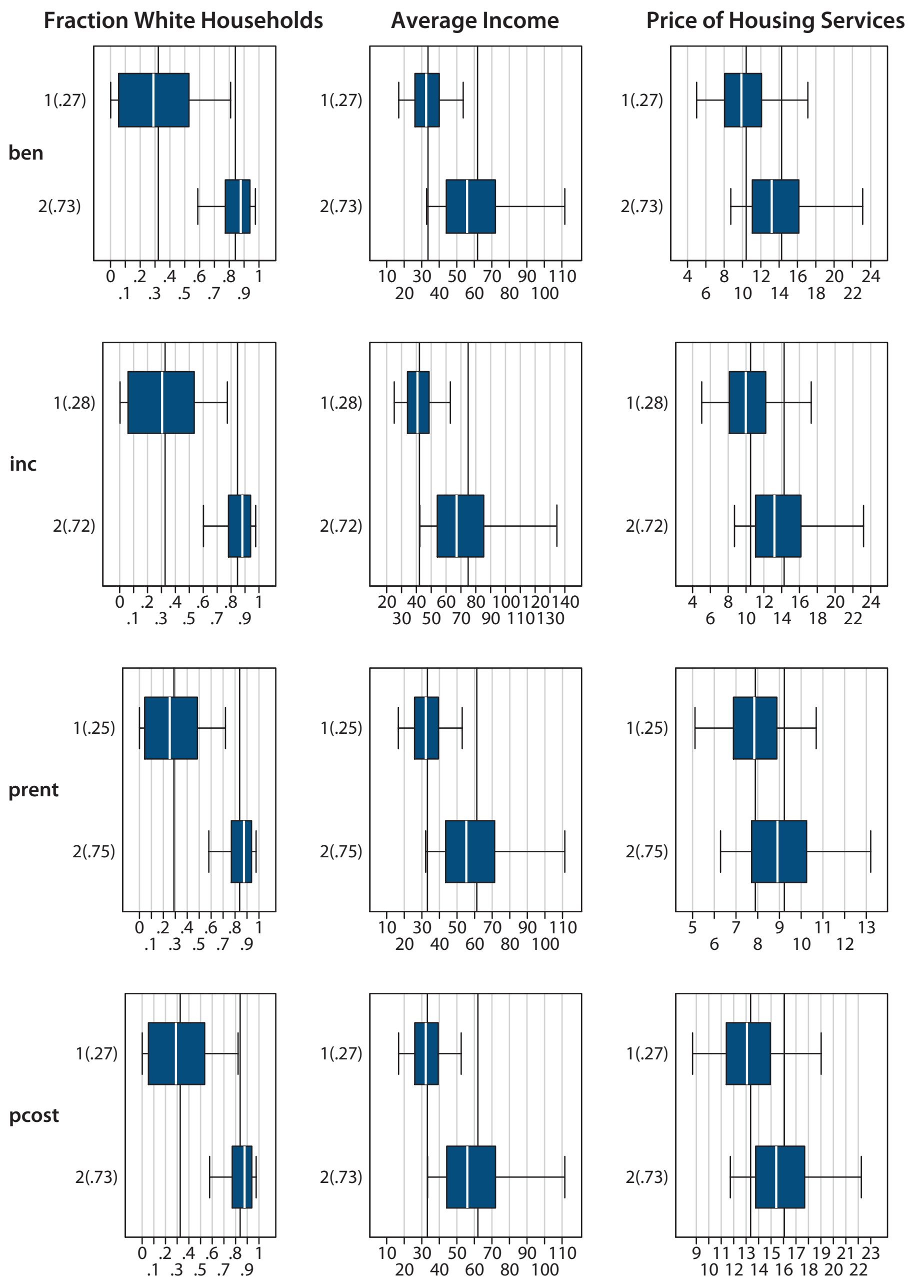
In summary, an off-the-shelf clustering procedure can be used to capture (i) up to 37 percent of the variation in income, racial configuration, and housing service prices across U.S. neighborhoods using only two representative neighborhoods and (ii) up to 52 percent of the variation using three representative neighborhoods.

Spatial Contiguity

Spatial theories of human capital formation emphasize spillovers across geographic locations. A common view states that the strength of these interactions declines with geographic distance. Therefore, the degree to which the tracts in each cluster are spatially contiguous suggests that the classification is potentially consistent with theories featuring spatial spillovers. In contrast, a low degree of contiguity would imply that each cluster is composed of scattered

Figure 1

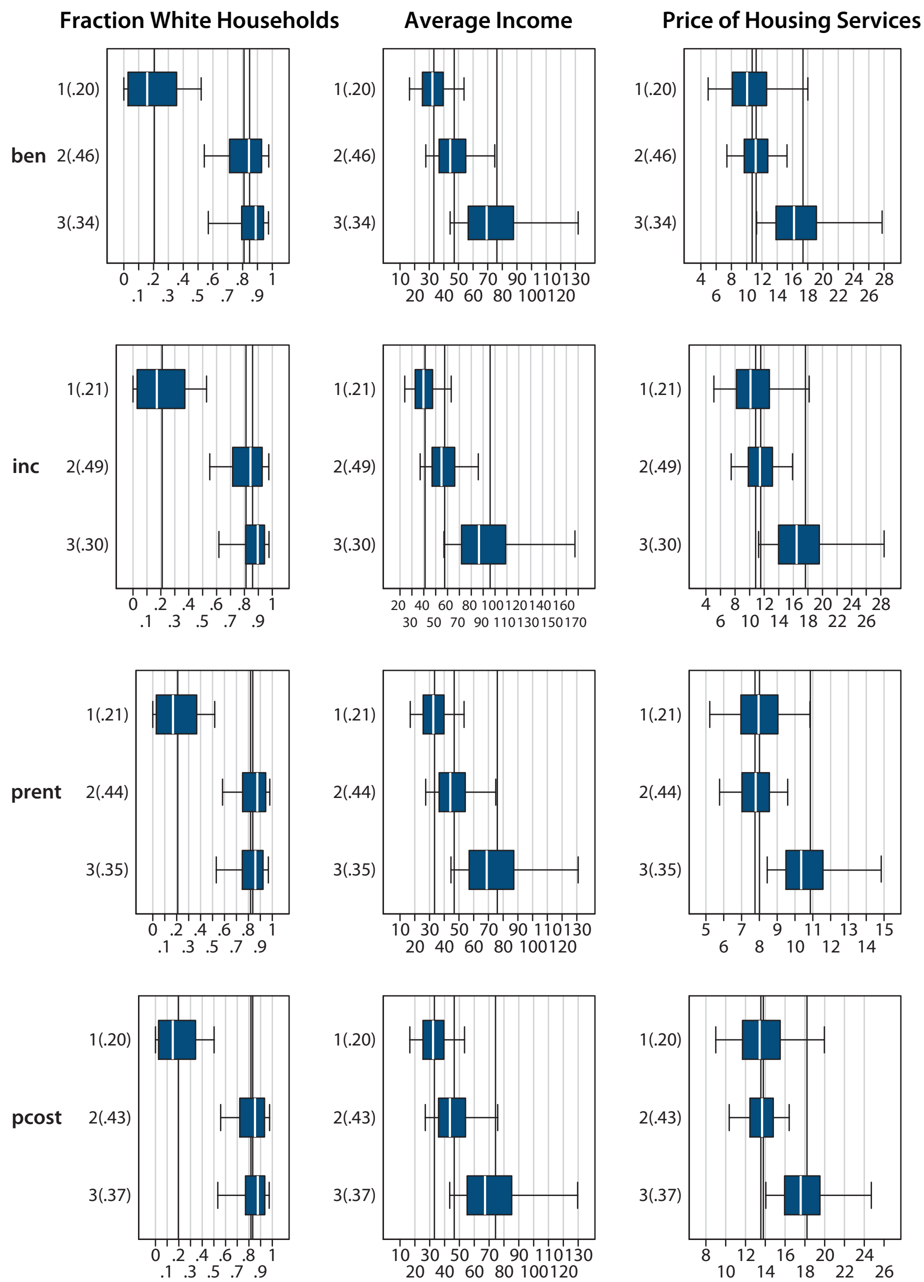
Within-Cluster Distribution of Neighborhood Characteristics: *K* = 2



NOTE: Columns show the plots for (i) racial configuration, (ii) earnings, and (iii) price of housing services measures. Rows correspond to each variable configuration. Within each plot, neighborhood classes (1 and 2 stand for types *I* and *II*, respectively) are listed in the vertical axis (the fraction of households is listed in parentheses). Vertical lines indicate neighborhood means (or centroid c_k). Boxes indicate the range between the 25th and 75th percentiles. Lines within boxes indicate medians. Brackets indicate the range between the 5th and 95th percentiles. All statistics are weighted by the number of households in each tract.

Figure 2

Within-Cluster Distribution of Neighborhood Characteristics: $K = 3$



NOTE: Columns display the plots for (i) racial configuration, (ii) earnings, and (iii) price of housing services measures. Rows correspond to each variable configuration. Within each plot, neighborhood classes (1, 2, and 3 stand for types *A*, *B*, and *C*, respectively) are listed in the vertical axis (the fraction of households is listed in parentheses). Vertical lines indicate neighborhood means (or centroid c_k). Boxes indicate the range between the 25th and 75th percentiles. Lines within boxes indicate medians. Brackets indicate the range between the 5th and 95th percentiles. All statistics are weighted by the number of households in each tract.

Table 6

Cluster Contiguity: $\kappa = 2.5$

	$K = 2$		$K = 3$		
MSA	*I*	*II*	*A*	*B*	*C*
z Score normalization					
Contiguity (%)	40	64	43	32	50
Adjacency	5.6	7.2	5.6	5.7	5.9
Mahalanobis normalization					
Contiguity (%)	41	68	41	28	50
Adjacency	5.6	7.4	5.6	5.7	6.1

NOTE: For a randomly chosen tract *i* of cluster C_k, contiguity equals the expected fraction of tracts of cluster C_k that are connected to *i*. Adjacency equals the expected number of cluster C_k tracts that are directly adjacent to *i*.

geographic areas, so that potential spatial spillovers would not have a large scope of action.

In this article, spatial contiguity is not imposed in any way.[9] However, spatial contiguity serves here as an additional measure of cluster adequacy.

Two strategies are used to assess spatial contiguity. The first computes a simple indicator that measures the fraction of neighborhoods of a class C_k to which the average neighborhood in C_k is "connected." The second strategy presents a few maps indicating the location of each class of neighborhoods in selected MSAs.

To measure contiguity, I begin with a pair of neighborhoods *A* and *B*. The Census Bureau provides the geographic coordinates at one internal point of each neighborhood, denoted as p_A and p_B. A neighborhood's radius can be defined as the radius (r_A, r_B) of a circle with the same geographic area as the corresponding neighborhood. Then, say that neighborhoods *A* and *B* are adjacent if $distance(p_A, p_B) \geq \kappa \max(r_A, r_B)$, where $\kappa \geq 1$ is an arbitrary constant. A *connected* set of neighborhoods is defined as a set of neighborhoods that cannot be separated into two subsets without separating at least one pair of adjacent neighborhoods.

The adjacency parameter of the contiguity indicator is set to $\kappa = 2.5$. This means that two tracts in the same cluster are considered adjacent if the distance between their Census-assigned internal points is less than 2.5 times the larger of their neighborhood radiuses.

Table 6 shows that, using the z score normalization and $K = 2$, type *I* neighborhoods are connected to 40 percent of their own type within an MSA and type *II* neighborhoods are connected to 64 percent of neighborhoods of their own type within an MSA. Thus, representative neighborhoods defined by the clustering procedure describe large geographic areas with homogeneous characteristics. Also, the expected number of same-type tracts adjacent to a randomly selected neighborhood lies between 5.6 and 7.2. Similar results hold using the Mahalanobis normalization.

For $K = 2$, type *I* neighborhoods tend to be substantially less connected than type *II* neighborhoods. This obeys the fact that type *I* neighborhoods form "islands" in a "sea" of type *II* neighborhoods (see the "MSA Maps" subsection below). Connectedness is lower because several MSAs contain more than one "island." For $K = 3$, type *B* neighborhoods tend to be less connected than the other two types.

MSA Maps

Figures 3 to 8 are maps corresponding to selected areas of three MSAs in the sample. For each MSA, the $K = 2$ and $K = 3$ characterizations are depicted in different shades of blue.

The two-neighborhood characterization exhibits a striking degree of contiguity. In the selected MSA, type *I* (low-income) neighborhoods form a small number of large areas, which are surrounded by type *II* (high-income) neighborhoods. This is remarkable given that (i) no geographic location information was used in the clustering procedure and (ii) the number of tracts within each "island" is large. For example, the Washington-Baltimore-Arlington MSA contains 1,453 tracts, of which 378 are type *I*. Almost all of these tracts are grouped into three islands (see Figure 7).

Finally, the three-neighborhood characterization is consistent with the two-neighborhood characterization. The type *I* (low-income) cluster of the two-neighborhood characterization is basically the same as the type *A* cluster in the three-neighborhood characterization. Therefore, the degree of contiguity for type *A* areas is also remarkable in the three-neighborhood characterization (Table 6). Roughly, the type *II* (high-income) neighborhoods of the two-neighborhood characterization are split into two new types (labeled *B* and *C*) when proceeding from $K = 2$ to $K = 3$. In the three-neighborhood characterization, type *B* neighborhoods exhibit the lowest degree of contiguity. These types of neighborhoods appear to the eye as transition areas between the clearly defined "islands" of type *A* and the "sea" of type *C* neighborhoods.

Regional Stability

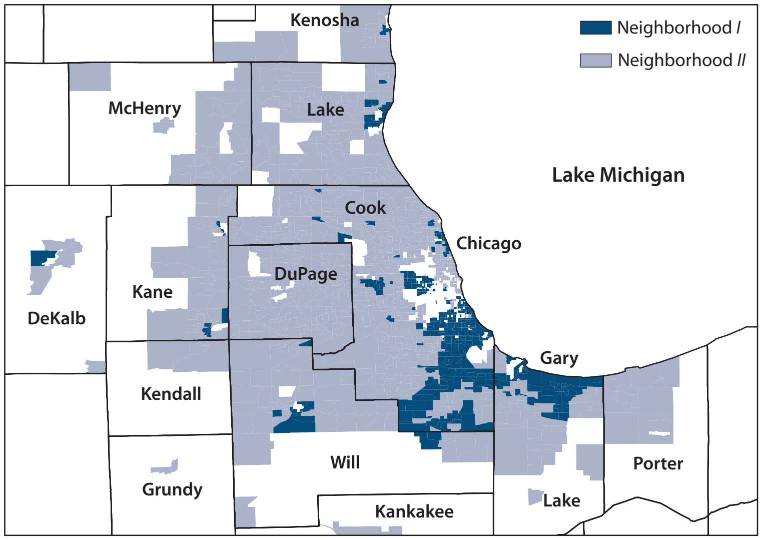
Recall that so far all results correspond to applying the clustering algorithm once to all neighborhoods in the sample. This subsection addresses whether the characterization of neighborhood is meaningful at the MSA level for $K = 2, 3$. The question is approached at two levels. First, do all MSAs contain a roughly similar fraction of each type of neighborhood, or are neighborhoods of each type concentrated in particular MSAs? In other words, are the fractions of each type stable across MSAs? Second, would the classification of neighborhoods differ substantially if centroids were allowed to vary across MSAs? The answers are yes and no, respectively.

First, Table 7 presents the percentage of each neighborhood class by MSA for $K = 2$ and $K = 3$. Each class of neighborhood exists in each MSA in roughly the same percentages, with standard deviations close to 8 percent for $K = 2$ and between 7.6 and 13.7 percent for $K = 3$.

Second, to allow for different centroids across MSAs, I cluster neighborhoods independently for each MSA using the benchmark variable configuration and *z* score normalization for $K = 2, 3, 4$. Then I use the cluster similarity measure to compare these clustering results with those for all MSAs pooled. Table 8 shows the percentage of neighborhoods classified in

Figure 3

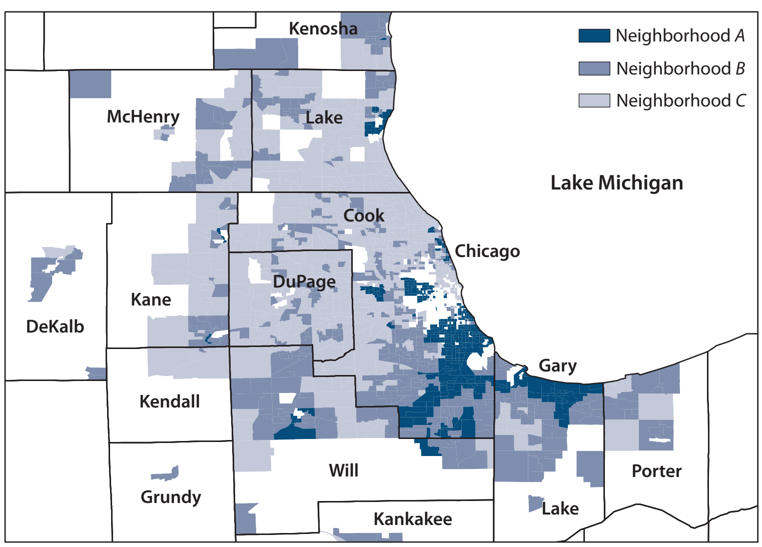
Chicago-Gary-Kenosha MSA: *K* = 2



NOTE: The figure shows selected neighborhoods of the corresponding MSA. Names of counties within the MSA are also listed.

Figure 4

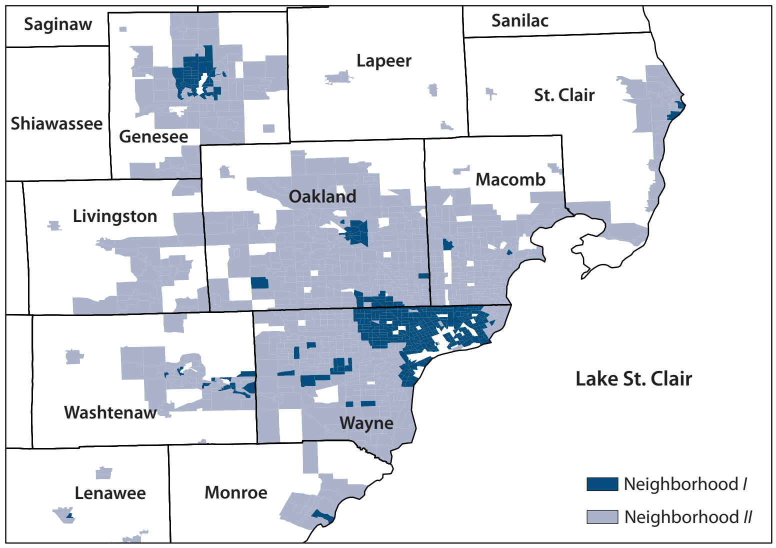
Chicago-Gary-Kenosha MSA: *K* = 3



NOTE: The figure shows selected neighborhoods of the corresponding MSA. Names of counties within the MSA are also listed.

Figure 5

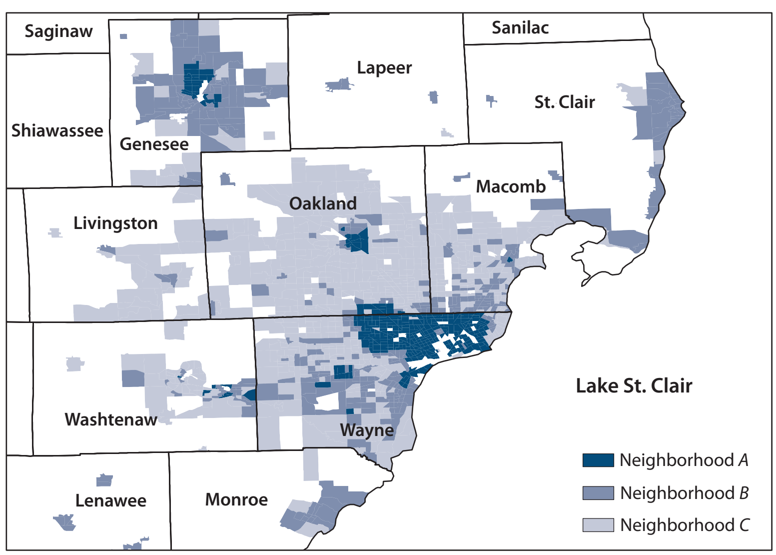
Detroit-Ann Arbor MSA: *K* = 2



NOTE: The figure shows selected neighborhoods of the corresponding MSA. Names of counties within the MSA are also listed.

Figure 6

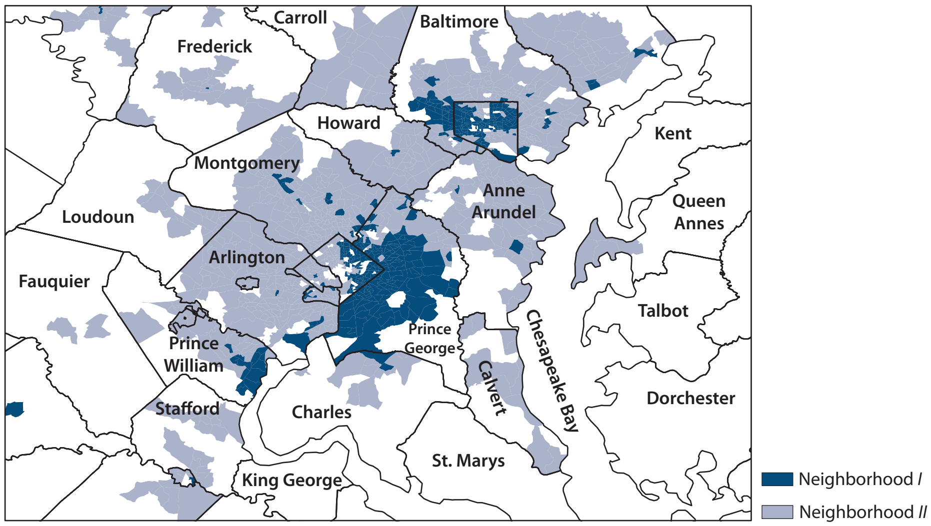
Detroit-Ann Arbor MSA: *K* = 3



NOTE: The figure shows selected neighborhoods of the corresponding MSA. Names of counties within the MSA are also listed.

Figure 7

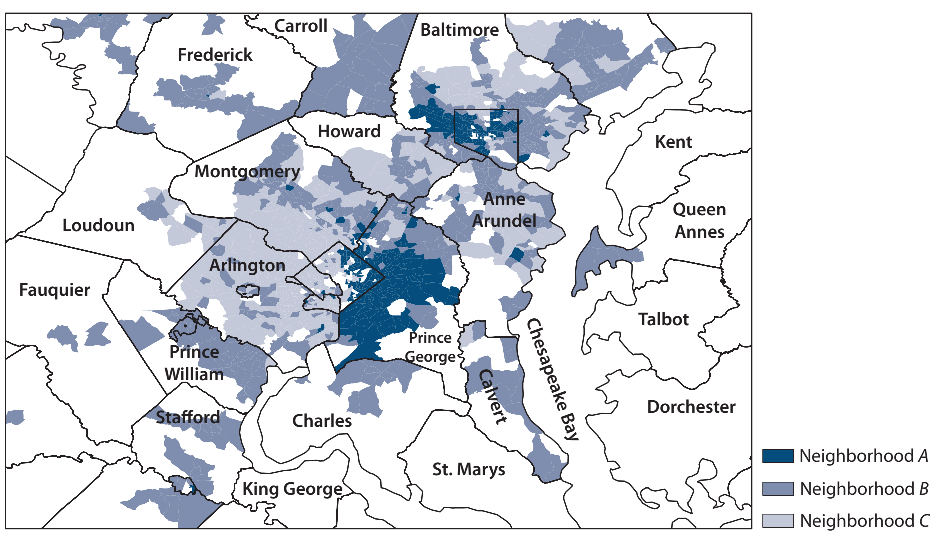
Washington-Baltimore-Arlington MSA: *K* = 2



NOTE: The figure shows selected neighborhoods of the corresponding MSA. Names of counties within the MSA are also listed.

Figure 8

Washington-Baltimore-Arlington MSA: *K* = 3



NOTE: The figure shows selected neighborhoods of the corresponding MSA. Names of counties within the MSA are also listed.

the same group for each MSA using MSA-by-MSA versus pooled clustering. The classifications are virtually identical for $K = 2$. The results for $K= 3$ are quite satisfactory with some exceptions: For example, for the Miami-Fort Lauderdale, the Norfolk-Virginia Beach-Newport News, and the Dallas-Fort Worth MSAs, clustering matches up for just 63, 61, and 65 percent of neighborhoods, respectively. I interpret these results as suggesting that the representative neighborhoods obtained reflect general economic and social forces common to most of the selected MSAs and specific regions or MSAs.

THE NATURE OF REPRESENTATIVE NEIGHBORHOODS

Two-Neighborhood Clustering

Two-neighborhood clustering provides the following characterization: Type *I* neighborhoods contain 27 percent of all households, have 4,800 residents per square kilometer, and cover about 4,600 square kilometers (Table 9). The population density of type *I* neighborhoods is about twice the density of type *II* neighborhoods, while the land area for type *I* neighborhoods is about one-fifth that of type *II* neighborhoods.

The $K = 2$ characterization reflects strong segregation by race. Of the households residing in type *I* neighborhoods, 32 percent are white, while 84 percent of households in type *II* neighborhoods are white (Table 10).

The $K = 2$ characterization also exhibits strong segregation by income. Household earnings average $33,591 in type *I* neighborhoods, representing 54 percent of average earnings in type *II* neighborhoods ($61,889). Household income averages $41,747 in type *I* neighborhoods, representing 56 percent of average income in type *II* neighborhoods ($74,577) (see Table 10).

Among black households, the average income for those in type *I* neighborhoods is 70 percent of the income for those in type *II* neighborhoods. For white households, type *I* neighborhood income is 58 percent of type *II* neighborhood income; for households in other racial categories the number is 62 percent. In type *I* neighborhoods, the average income of black households is $40,076, which is 90 percent of the average income of white households in that type of neighborhood ($44,727), while it is 74 percent in type *II* neighborhoods. Finally, the price of a unit of housing services is $10,405 in type *I* neighborhoods, representing 73 percent of the price in type *II* neighborhoods ($14,268) (see Table 10). This ratio is higher than the ratio observed for income, meaning that prices vary less than incomes across the two neighborhoods. This observation echoes a finding from the cross-MSA literature. Davis and Ortalo-Magne (2011) present evidence that the share of housing expenditures in income is constant in the United States. They show that a model with constant expenditure shares (i.e., with Cobb-Douglas preferences for housing and nonhousing consumption) and identical agents implies that prices should disproportionately reflect differences in incomes across MSAs. As in our two-neighborhood representation, the price measures provided by Davis and Ortalo-Magne vary less than incomes across MSAs. They find this observation puzzling viewed through the lens of their model.

Table 7

Percentage of Households by Neighborhood Class within Each MSA*

	K = 2		K = 3			
MSA	*I*	*II*	*A*	*B*	*C*	No. of tracts
Atlanta	32	68	27	53	20	568
Buffalo-Niagara Falls	34	66	16	79	5	250
Charlotte-Gastonia-Rock Hill	23	77	15	46	38	246
Chicago-Gary-Kenosha	22	78	19	35	47	1,658
Cincinnati-Hamilton	19	81	11	68	21	391
Cleveland-Akron	26	74	18	64	18	738
Columbus, OH	20	80	13	64	23	310
Dallas-Fort Worth	29	71	18	53	28	833
Detroit-Ann Arbor-Flint	24	76	20	29	50	1,335
Greensboro-Winston-Salem-High Point	29	71	20	58	23	196
Houston-Galveston-Brazoria	41	59	29	56	15	630
Indianapolis	22	78	13	66	21	278
Jacksonville, FL	32	68	15	65	20	162
Kansas City	24	76	15	66	19	400
Louisville	20	80	13	73	14	207
Memphis	49	51	46	32	22	203
Miami-Fort Lauderdale	49	51	33	42	25	409
Milwaukee-Racine	22	78	17	49	33	392
New York-Northern New Jersey-Long Island	23	77	19	24	57	3,850
Nashville	18	82	13	54	33	186
New Orleans	44	56	34	45	21	339
Norfolk-Virginia Beach-Newport News	36	64	26	66	8	309
Orlando	33	67	18	65	17	287
Philadelphia-Wilmington-Atlantic City	25	75	19	58	23	1,356
Raleigh-Durham-Chapel Hill	20	80	14	34	52	157
St. Louis	26	74	18	65	18	429
West Palm Beach-Boca Raton	33	67	16	57	27	243
Washington-Baltimore	29	71	23	46	31	1,453
Total	27	73	20	46	34	17,815
SD	8.4	8.4	7.6	13.7	12.4	
Tracts	5,649	12,166	4,458	7,456	5,901	

NOTE: *Benchmark variable configuration, *z* score normalization.

Table 8

Cluster Similarity: Pooled Versus MSA by MSA Clustering*

MSA	$K = 2$	$K = 3$	$K = 4$
Atlanta	96	89	80
Buffalo-Niagara Falls	91	74	79
Charlotte-Gastonia-Rock Hill	88	87	65
Chicago-Gary-Kenosha	98	79	84
Cincinnati-Hamilton	98	76	71
Cleveland-Akron	97	85	72
Columbus, OH	76	78	67
Dallas-Fort Worth	85	65	87
Detroit-Ann Arbor-Flint	99	91	65
Greensboro-Winston-Salem-High Point	98	75	61
Houston-Galveston-Brazoria	96	92	86
Indianapolis	82	71	62
Jacksonville, FL	98	78	70
Kansas City	99	82	64
Louisville	98	74	58
Memphis	97	77	76
Miami-Fort Lauderdale	93	63	72
Milwaukee-Racine	98	73	60
New York-Northern New Jersey-Long Island	83	91	54
Nashville	92	86	63
New Orleans	95	69	59
Norfolk-Virginia Beach-Newport News	93	61	63
Orlando	80	70	55
Philadelphia-Wilmington-Atlantic City	95	83	69
Raleigh-Durham-Chapel Hill	96	70	67
St. Louis	97	70	67
West Palm Beach-Boca Raton	95	86	93
Washington-Baltimore	86	66	55
Average	93	77	69

NOTE: The reported statistic corresponds to the percentage of neighborhoods classified in the same group by applying the clustering algorithm to the pooled dataset (all MSAs) versus applying it separately to each MSA. *Benchmark variable configuration, *z* score normalization.

Table 9

Population Density and Area*

	$K = 2$		$K = 3$		
Variable	***I***	***II***	***A***	***B***	***C***
Population density	4,837	2,251	5,333	2,070	2,674
Area (1,000 sq km)	4.59	25.32	3.19	17.07	10.0

NOTE: **z* Score normalization.

Table 10

Characteristics of Representative Neighborhoods: $K = 2$

Neighborhood	***I***	***II***	***I/(I + II)***
Number of households (thousands)			
Black	4,451	1,359	0.77
White	2,662	18,577	0.13
Other	1,152	2,150	0.35
Total	8,265	22,085	0.27
White/Total	0.32	0.84	

Neighborhood	***I***	***II***	***I/II***
Average income ($)			
Black	40,076	57,124	0.70
White	44,727	76,711	0.58
Other	41,320	67,166	0.62
Total	41,747	74,577	0.56
Average earnings ($)	33,591	61,889	0.54
Price of housing services*	10,405	14,268	0.73

NOTE: *Units are normalized to match the value of the original IRV measure (see the appendix).

Three-Neighborhood Clustering

The three representative neighborhoods are denoted by *A*, *B*, and *C*. The three-neighborhood clustering generates the following characterization. Type *A* neighborhoods cover 3,200 square kilometers, while type *B* neighborhoods cover 17,000 square kilometers and type *C* neighborhoods cover 10,000 square kilometers. The population density of type *A* neighborhoods is about 5,300 residents per square kilometer, while the density is much lower in the other two neighborhoods: 2,100 per square kilometer in type *B* and 2,700 per square kilometer in type *C* (see Table 9).

Table 11

Characteristics of Representative Neighborhoods: *K* = 3

Neighborhood	*A*	*B*	*C*	*A* / (*A* + *B* + *C*)	*B* / (*A* + *B* + *C*)
Number of households (thousands)					
Black	4,074	1,244	492	0.70	0.21
White	1,268	11,244	8,727	0.06	0.53
Other	821	1,396	1,084	0.25	0.42
Total	6,163	13,884	10,303	0.20	0.46
White/Total	0.24	0.90	0.95		
Neighborhood	***A***	***B***	***C***	***A/C***	***B/C***
Average income ($)					
Black	39,949	49,059	65,481	0.61	0.75
White	43,955	58,651	94,982	0.46	0.62
Other	40,363	53,483	77,620	0.52	0.69
Total	40,899	57,696	93,407	0.44	0.62
Average earnings ($)	33,142	47,106	76,303	0.43	0.62
Price of housing services*	10,715	11,238	17,377	0.62	0.65

NOTE: *Units are normalized to match the value of the original IRV measure (see the appendix).

In terms of racial configuration, there is a strong concentration of black households in type *A* neighborhoods, while type *B* and *C* neighborhoods contain similarly large percentages of white households. Only 21 percent of households in type *A* neighborhoods are white, while 81 percent and 85 percent of households in type *B* and *C* neighborhoods, respectively, are white (Table 11).

Percentage differences in income between type *A* and *B* neighborhoods and between type *B* and *C* neighborhoods are similar, generating three approximately equally spaced strata. Average earnings are $33,142, $47,106, and $76,303 in type *A*, *B*, and *C* neighborhoods, respectively. Thus, the ratio of average earnings of *A* with respect to *B* is 0.70, while the ratio of *B* to *C* is 0.62. The picture is similar for average income. Incomes in type *A*, *B*, and *C* neighborhoods average $40,899, $57,696, and $93,407, respectively (see Table 11).

For black households, the ratio of average income for those in type *A* neighborhoods to those in type *C* neighborhoods is 0.61. This between-neighborhoods ratio is 0.46 for white households and 0.52 for households in other racial categories. These ratios of average income by race in neighborhoods type *B* with respect to *C* are 0.75 for black households, 0.62 for white households, and 0.69 for households of other races. This shows that, in terms of averages, the sorting of households in ascending order of income into neighborhoods *A*, *B*, and *C* holds not only for aggregate populations but also for each race separately.

The black-to-white ratio of average income is 0.91, 0.84, and 0.69 in type A, B, and C, neighborhoods, respectively, while the overall ratio is 0.61.[10] The fact that the within-neighborhood ratios are above 0.61 suggests that within-neighborhood racial inequality is smaller than overall racial inequality for every neighborhood. This was also a feature of the two-neighborhood characterization (see Tables 10 and 11). Also, it is interesting to note that, as in the $K = 2$ case, cross-neighborhood differences are less marked for the price of housing services than for income; the ratio of the price of housing services in A with respect to B is 0.95, while the B to C ratio is 0.65.

ROBUSTNESS

This section determines the degree to which Census tracts in the sample are classified in the same way under (i) several variable configurations and variable normalization strategies and (ii) several variations of the sample selection criteria.

Variable Configuration/Normalization

First, the clustering procedure is applied under each possible (*variable configuration*, *normalization strategy*) combination. Then, the resulting clusterings are compared and a measure of clustering similarity is applied to determine whether the results are similar.

There is a natural measure of similarity in the literature that works well when the number of clusters K is small. The measure takes two different clusterings, say $C^1 = \{C_1^1 \ldots C_K^1\}$ and $C^2 = \{C_1^2 \ldots C_K^2\}$, and counts the fraction of objects that are classified into the same group in both clusterings.[11]

The results are striking for $K = 2$. In the worst case, 90 percent of neighborhoods are classified in the same group. On average, 94 percent of neighborhoods are classified in the same group. In many cases, the classification is identical. The results for $K = 3$ are less robust so they are provided in Table 12. In the worst case, 76 percent of neighborhoods are classified in the same group, but in most cases, more than 80 percent are classified in the same way.

The similarity across these clusterings suggests that racial configuration, income, and price of housing services provide a meaningful characterization of neighborhoods. Regardless of the diverse measures and normalizations used, the neighborhoods are similarly classified.

Sample Selection Criteria

Sample selection criteria are varied to examine the robustness of the representative-neighborhood characterizations presented in the previous two subsections. I consider the following four variations of sample selection criteria:

1. including MSAs with populations above 250,000 (versus 1 million in the baseline sample);
2. including MSAs with a 5 percent (or more) black population (versus 10 percent in the baseline sample);
3. including neighborhoods with 90 percent or less of "other race" households (versus 50 percent or less in the baseline sample); and

Table 12

Robustness to Variable Configuration/Normalization: Cluster Similarity (percent)*

	Normalization							
	z Score				Mahalanobis			
Configuration/normalization	**ben**	**inc**	**prent**	**pcost**	**ben**	**inc**	**prent**	**pcost**
z Score normalization								
ben	100	94	80	89	76	77	78	77
inc		100	81	87	76	78	78	77
prent			100	81	80	80	90	80
pcost				100	77	77	80	77
Mahalanobis normalization								
ben					100	92	88	98
inc						100	86	92
prent							100	88
pcost								100

NOTE: The reported statistic corresponds to the percentage of neighborhoods classified in the same group under two alternative variable configurations. Variable configurations are described in Table 2. *$K = 3$, all variable configurations and normalizations.

4. excluding neighborhoods with average earnings above $150,000 (versus no upper limit in the baseline sample).

The clustering procedure is applied to each sample variation. Table 13 presents the values of the centroids for (Y, R, P), as well as other important characteristics of the two-neighborhood characterization for the baseline sample and sample variations 1 through 4.

Sample variations 1 and 2 result in a dataset containing neighborhoods from 56 and 41 MSAs, respectively (compared with 28 MSAs in the baseline sample). Table 13 shows that sample variation 1 leaves the two-neighborhood characterization virtually unchanged with respect to the baseline sample.[12] Sample variation 2 implies changes in the racial composition of the sample. The overall fraction of black households in the sample falls from 0.19 to 0.16 with respect to the baseline. This change is reflected in the neighborhood characterization. The fraction of white households in type *I* neighborhoods moves from 0.32 to 0.40. This is the only appreciable change in the neighborhood characterizations imposed by the sample variation 2. Sample variation 3 implies the addition of 1,098 tracts to the sample (the number of MSAs remains 28). This change leaves the neighborhood characterization virtually unchanged. Finally, sample variation 4 implies the deletion of 212 observations, with no appreciable effects on the two-neighborhood characterization. Therefore, the results obtained in the baseline sample for the high-earnings neighborhood (type *II*) are not affected by the presence of a small fraction of neighborhoods with very large average earnings.

Table 13

Varying Sample Selection Criteria*

		Sample variation			
Statistic	**Baseline**	**1**	**2**	**3**	**4**
Neighborhood I					
Average earnings ($)	33,591	32,656	33,606	33,795	33,402
Fraction of white HHs	0.32	0.33	0.40	0.31	0.32
Price of housing services ($)	10,405	9,976	10,577	10,716	10,063
Neighborhood II					
Average earnings ($)	61,889	60,222	62,911	61,930	60,311
Fraction of white HHs	0.84	0.84	0.83	0.84	0.84
Price of housing services	14,268	13,604	16,562	14,228	13,768
Aggregate					
Fraction of HHs living in *II*	0.73	0.71	0.69	0.67	0.71
Overall fraction of white HHs	0.70	0.70	0.70	0.67	0.69
Overall fraction of black HHs	0.19	0.20	0.16	0.19	0.20
Number of MSAs	28	56	41	28	28
Number of tracts	17,815	20,148	24,054	18,913	17,603

NOTE: HH, household. Sample variations 1 through 4 correspond to the following sample selection criteria: (1) including MSAs with population above 250,000 (versus 1 million in baseline sample); (2) including MSAs with 5 percent or more black population (versus 10 percent in baseline sample); (3) including neighborhoods with 90 percent or less of "other race" households (versus 50 percent or less in baseline sample); (4) excluding neighborhoods with average earnings above $150,000 (versus no upper limit in the baseline sample). *Benchmark variable configuration, z score normalization, $K = 2$.

REMARKS AND CONCLUSION

This article explores the existence of a suitable representative-neighborhood characterization of metropolitan U.S. data. Such a characterization allows complex neighborhood-level data to be simplified. A simple characterization permits a transparent interpretation of data through models featuring a small number of neighborhoods with the advantage that the characterization has a direct geographic counterpart (see Figures 3 to 8).

One potential use of this characterization is to impose empirical discipline on quantitative models with a small number of locations. The main advantages for quantitative models calibrated to match a representative-neighborhood characterization are simplicity and clarity, yet such calibration has another appealing feature. The K-means clustering algorithm, as applied here, provides a partition of neighborhoods that minimizes a sum of squares criterion. Therefore, if the representative neighborhoods are reproduced by locations in a quantitative model, such a model will achieve the best possible fit to neighborhood-level data under the sum of squares criterion. This feature provides a rationale for fitting more-complex models to match aspects of the characterization developed in this article. ■

APPENDIX

Price of Housing Services

The data contain three sources of information regarding expenditures for housing services. The first source is the *median gross rent* variable. This is the median rent paid by renter households in a tract. The measure is designed to include the cost of utilities and fees, such as condo fees, when applicable, in addition to rent. The second source is the *median house value* variable. This measure is computed by the Census Bureau using market values of housing units reported by home-owning households. The housing literature uses these values to construct an expenditure measure or implicit rental value (IRV). The third source is the *median selected monthly owner costs* variable. This measure is constructed by the Census Bureau to estimate the monthly cost of housing for homeowners.[13]

Median tract house values are converted into median annual IRVs using a procedure based on Poterba (1992). This procedure consists of applying an annual user-cost factor to house values. A factor of 8.93 percent of the house value is used.[14]

NOTES

1 See Calabrese et al. (2006, footnote 4) for a list of examples.

2 The set of housing characteristics for each tract is composed of (i) the median number of rooms in the unit, (ii) a distribution of the number of units in the housing structure (10 categories), (iii) a distribution of the number of bedrooms (6 categories), (iv) the fraction of units with telephone service, (v) the fraction of units with complete plumbing facilities, (vi) the fraction of units with complete kitchen facilities, and (vii) the distribution of travel time to work (12 categories).

3 This is a standard threshold in the housing literature above which an area is considered urban.

4 Correctional institutions, nursing homes, juvenile detention facilities, college dormitories, military quarters, and group homes are considered group quarters.

5 Iterative relocation proceeds as follows: (i) Assign elements arbitrarily into an initial partition consisting of *K* clusters and calculate the centroid of each cluster. (ii) Generate a new partition by reassigning each element to the nearest cluster centroid. If no objects were reassigned, terminate. (iii) Compute new centroids using the partition obtained in step (ii). Return to step (ii).

6 The choice of normalization is not necessarily innocuous. For example, Jain and Dubes (1988, p. 25) provide a case in which *z* score normalization destroys the cluster structure in a particular dataset.

7 See Jain and Dubes, 1988, section 4.5, for an extensive discussion of the concept of cluster validity.

8 Since x_i and c are vectors, the "overall variability" is defined as the sum of each component's variation (see the denominator in the expression for R^2).

9 A branch of classification analysis deals with the clustering of objects that are described by a vector of variables (x_i) and also by their position on a plane. In some cases, it may be desirable that objects in the same class are also spatially contiguous. In the problem of digital image segmentation, it is usually desirable that adjacent pixels belong to the same class. See, for example, Theiler and Gisler (1997). In the extreme, one could restrict all elements in a given class to be contiguous. This constrained clustering problem is known as regionalization (see, for example, Duque, Church, and Middleton, 2006). A simpler approach (i) includes the spatial coordinates of each object in the vector of characteristics x_i and (ii) applies an unconstrained clustering algorithm. The algorithm will tend toward generating clusters that are "close" in the plane.

10 These ratios are not provided in the tables but are easily calculated from income entries in Table 11.

11 This task is complicated by the fact that the subindexes labeling each cluster can be assigned arbitrarily (i.e., there is no way to decide which cluster in C^1 corresponds to any particular cluster in C^2). Therefore, one should examine all possible permutations of the cluster subindexes and choose the one yielding the maximum fraction of coincidences. If *P* is the set of all possible permutations *p*(*k*) of the indexes (1,2,3...*k*...*K*), then the measure can be expressed as

$$\max_{p\in P} \frac{\sum_{k=1}^{K}\left|C_k^1 \cap C_{p(k)}^2\right|}{N},$$

where the "absolute value" denotes the number of elements in a cluster C_k.

12 The results shown in Table 13 compare only the (*Y*, *R*, *P*) averages across Census tracts (centroids). However, analysis of higher moments of (*Y*, *R*, *P*) in each representative neighborhood shows these are remarkably stable across different samples as well. Details are available from the author upon request.

13 The selected monthly owner costs variable includes reported payments of mortgages, deeds of trust, contracts to purchase, or similar debts on the property (including payments for the first mortgage, second mortgage, home equity loans, and other junior mortgages); real estate taxes; fire, hazard, and flood insurance on the property; utilities (electricity, gas, water, and sewer); and fuels (oil, coal, kerosene, wood, and so on). It also includes monthly condominium fees or mobile home costs (installment loan payments, personal property taxes, site rent, registration fees, and license fees).

14 Calabrese et al. (2006) use this approach. The user cost of housing for homeowners is calculated by letting implicit rental values *IRV* be given by $IRV = \kappa_p V$, where *V* is the market value of the home. The annual user-cost factor is given by

$$\kappa_p = (1 - t_y)(i + t_v) + \psi,$$

where t_y is the income tax rate, i is the nominal interest rate, t_v is the property tax rate, and ψ contains the risk premium for housing investments, maintenance and depreciation costs, and the inflation rate.

REFERENCES

Calabrese, Stephen; Epple, Dennis; Romer, Thomas and Sieg, Holger. "Local Public Good Provision: Voting, Peer Effects, and Mobility." *Journal of Public Economics*, August 2006, *90*(6-7), pp. 959-81.

Davis, Morris A. and Ortalo-Magne, Francois. "Household Expenditures, Wages, Rents." *Review of Economic Dynamics*, April 2011, *14*(2), pp. 248-61.

Duque, Juan Carlos; Church, Raúl and Middleton, Richard. "Exact Models for the Regionalization Problem," in Western Regional Science Association Annual Meetings, Santa Fe, 2006.

Fernandez, Raquel and Rogerson, Richard. "Public Education and Income Distribution: A Dynamic Quantitative Evaluation of Education-Finance Reform." *American Economic Review*, September 1998, *88*(4), pp. 813-33.

Iceland, John; Weinberg, Daniel H. and Steinmetz, Erika. "Racial and Ethnic Residential Segregation in the United States: 1980-2000." U.S. Census Bureau, Series CENSR-3. Washington, DC: U.S. Government Printing Office, 2002; http://www.census.gov/prod/2002pubs/censr-3.pdf.

Ioannides, Yannis M. "Neighborhood Income Distributions." *Journal of Urban Economics*, November 2004, *56*(3), pp. 435-57.

Jain, Anil K. and Dubes, Richard C. *Algorithms for Clustering Data*. Englewood Cliffs, NJ: Prentice Hall, 1988.

Poterba, James M. "Taxation and Housing: Old Questions, New Answers." *American Economic Review*, May 1992, *82*(2), pp. 237-42.

Theiler, James and Gisler, Galen. "A Contiguity-Enhanced K-Means Clustering Algorithm for Unsupervised Multispectral Image Segmentation." Proceedings of the Society of Optical Engineering, October 1997, *3159*, pp. 108-118.

U.S. Census Bureau. "2000 Census of Population and Housing—Summary File 3." Last revised October 13, 2011; http://www.census.gov/census2000/sumfile3.html.

Wheeler, Christopher H. and La Jeunesse, Elizabeth. A. "Neighborhood Income Inequality." Working Paper No. 2006-039B, Federal Reserve Bank of St. Louis, June 2006, revised February 2007; http://research.stlouisfed.org/wp/2006/2006-039.pdf.

Factor-Based Prediction of Industry-Wide Bank Stress

Sean Grover and Michael W. McCracken

This article investigates the use of factor-based methods for predicting industry-wide bank stress. Specifically, using the variables detailed in the Federal Reserve Board of Governors' bank stress scenarios, the authors construct a small collection of distinct factors. We then investigate the predictive content of these factors for net charge-offs and net interest margins at the bank industry level. The authors find that the factors do have significant predictive content, both in and out of sample, for net interest margins but significantly less predictive content for net charge-offs. Overall, it seems reasonable to conclude that the variables used in the Fed's bank stress tests are useful for identifying stress at the industry-wide level. The final section offers a simple factor-based analysis of the counterfactual bank stress testing scenarios. (JEL C12, C32, C52, C53)

Federal Reserve Bank of St. Louis *Review*, Second Quarter 2014, *96*(2), pp. 173-93.

The Dodd-Frank Wall Street Reform and Consumer Protection Act requires a specified group of large U.S. bank holding companies to submit to an annual Comprehensive Capital Analysis and Review (CCAR) of their capital adequacy should a hypothetical "severely adverse" economic environment arise. Of course, such an analysis requires defining the phrase "severely adverse," and one could imagine a variety of possible economic outcomes that would be considered bad for the banking industry. Moreover, what is stressful for one bank may not be stressful for another because of differences in their assets, liabilities, exposure to counterparty risk or particular markets, geography, and many other features associated with a specific bank.

The Dodd-Frank Act requires the Federal Reserve System to perform stress tests across a range of banks, so as a practical matter, it must choose a scenario that is likely to be severely adverse for all of the banks simultaneously. The current scenarios are provided in the "2014 Supervisory Scenarios for Annual Stress Tests Required under the Dodd-Frank Act Stress Testing Rules and the Capital Plan Rule" documentation.[1] The quarterly frequency series that characterize the "severely adverse" scenario include historical data back to 2001:Q1; but, more importantly, the series provide hypothetical outcomes moving forward from 2013:Q4 through

Sean Grover is a research associate and Michael W. McCracken is an assistant vice president and economist at the Federal Reserve Bank of St. Louis.
The authors thank Bill Dupor and Dan Thornton for helpful comments.

Table 1

Data Used

Variable name	Starting date	Source	Transformations
Charge-off rate on all loans, all commercial banks	1985:Q1	FRED	Percent
Net interest margin for all U.S. banks	1984:Q1	FRED	Percent
Real gross domestic product	1947:Q1	Haver Analytics	Annualized percent change
Gross domestic product	1947:Q1	Haver Analytics	Annualized percent change
Real disposable personal income	1959:Q1	Haver Analytics	Annualized percent change
Disposable personal income	1959:Q1	Haver Analytics	Annualized percent change
Unemployment rate, SA	1948:Q1	Haver Analytics	Percent
CPI, all items, SA	1977:Q1	Haver Analytics	Annualized percent change
3-Month Treasury yield, avg.	1981:Q3	Haver Analytics	Percent
5-Year Treasury yield, avg.	1953:Q2	Haver Analytics	Percent
10-Year Treasury yield, avg.	1953:Q2	Haver Analytics	Percent
Citigroup BBB-rated corporate bond yield index, EOP	1979:Q4	Haver Analytics	Percent
Conventional 30-year mortgage rate, avg.	1971:Q2	Haver Analytics	Percent
Bank prime loan rate, avg.	1948:Q1	Haver Analytics	Percent
Dow Jones U.S. total stock market index, avg.	1979:Q4	Haver Analytics	Annualized percent change
CoreLogic national house price index, NSA	1976:Q1	Haver Analytics	Annualized percent change
Commercial Real Estate Price Index, NSA	1945:Q4	Haver Analytics	Annualized percent change
Chicago Board Options Exchange market volatility index (VIX), avg.	1990:Q1	Haver Analytics	Levels

NOTE: Avg., average; CPI, consumer price index; EOP, end of period; NSA, non-seasonally adjusted; SA, seasonally adjusted.

2016:Q4 for a total of 13 quarters. There are 16 series $X = (x_1,\ldots,x_{16})'$ in the scenario. The real, nominal, monetary, and financial sides of the economy are all represented and include real gross domestic product (GDP) growth, consumer price index (CPI)-based inflation, various Treasury yields, and the Dow Jones Total Stock Market Index. Table 1 provides the complete list of variables. The paths of various international variables are also available in the documentation but are relevant for only a small subset of the banks undergoing stress testing. Hence, for brevity, they are omitted from the remainder of this discussion.

The path of the variables in this severely adverse scenario is safely described as indicative of those among the worst post-World War II recessionary conditions faced by the United States: The unemployment rate peaks at 11.25 percent in mid-2015, real GDP declines more than 4.5 percent in 2014, and housing prices decline roughly 25 percent. The other series are characterized by comparably dramatic shifts. Individually, each element of the scenario seems reasonable at an intuitive level. For example, recessions typically occur during periods when real GDP declines, and deeper recessions exhibit larger declines in real GDP. The given scenario is characterized by a large decline in real GDP; hence, the scenario could reasonably be described as a severe recession.

And yet, we are left with the deeper question of whether the severe recession is severely adverse for the banks. In this article, we attempt to address this question and provide indexes designed to measure and predict the scenario's degree of severity. The first step in our analysis requires choosing a measure of economic "badness" faced by a bank. We consider two measures: net charge-offs (NCOs) and net interest margins (NIMs). The NCO measure is the percent of debt that the bank believes it will be unlikely to collect. Banks write off poor-credit-quality loans as bad debts. Some of these debts are recovered, and the difference between the gross charge-offs and recoveries is NCOs. In a weakened economy, as loan defaults increase, NCOs would be expected to rise. The NIM measure is the difference in the rate of interest income paid to, say, depositors relative to the interest income earned from outstanding loans (in terms of assets). NIMs are a key driver of revenue in the traditional banking sector and, hence, movements in NIMs give an indication of the bank's overall health. Rather than attempt an analysis of individual banks, we instead use the charge-offs and interest margins aggregated across all commercial banks.

With these target variables (y) in hand, our goal is to construct and compare a few macroeconomic factors (rather than financial or banking industry-specific factors) that can be used to track the strength of the banking industry. In particular, the goal is to obtain macroeconomic factors that are useful as predictors of the future state of the banking sector. We consider three distinct approaches to extracting factors (f) from the predictors (X) with an eye toward predicting the target variable (y): principal components (PC), partial least squares (PLS), and the three-pass regression filter (3PRF) developed in Kelly and Pruitt (2011). We discuss each of these methods in turn in the following section.

We believe our results suggest that factor-based methods are useful as predictors of future bank stress. We reach these conclusions based on two major forms of evidence. First, based on pseudo out-of-sample forecasting exercises, it appears that factors extracted from these series in the CCAR scenarios can be fruitfully used to predict bank stress as measured using NIMs and, to a lesser extent, NCOs. In addition, when applied to the counterfactual scenarios, the factor models imply forecast paths that match our intuition on the degree of severity of the scenarios: The severely adverse scenarios tend to imply higher NCOs than do the adverse scenarios, which in turn tend to imply higher NCOs than do the baseline scenarios.

Not surprisingly given the increased attention to bank stress testing, a new literature strand focused on methodological best practices is quickly developing. Covas, Rump, and Zakrajsek (2013) use panel quantile methods to predict bank profitability and capital shortfalls. Acharya, Engle, and Pierret (2013) construct market-based metrics of bank-specific capital shortfalls. Our article is perhaps most closely related to those of Bolotnyy, Edge, and Guerrieri (2013) and Guerrieri and Welch (2012). Bolotnyy, Edge, and Guerrieri consider a variety of approaches to forecasting NIMs including factor-based methods, but they do so using level, slope, and curvature factors extracted from the yield curve rather than the macroeconomic aggregates specific to the stress testing scenarios. Guerrieri and Welch consider a variety of approaches to forecasting NIMs and NCOs using macroeconomic aggregates similar to those included in the scenarios, but they emphasize the role of model averaging across many bivariate vector autoregressions (VARs) rather than factor-based methods per se.

The next section offers a brief overview of the data used and estimation of the factors. We evaluate the predictive content of the indexes for NCOs and NIMs both in and out of sample. We then use the estimated factor weights along with the series provided in the hypothetical stress scenarios to construct implied counterfactual factors. The counterfactual factors provide a simple-to-interpret measure of the degree of stressfulness implied by the scenario.

DATA AND METHODS

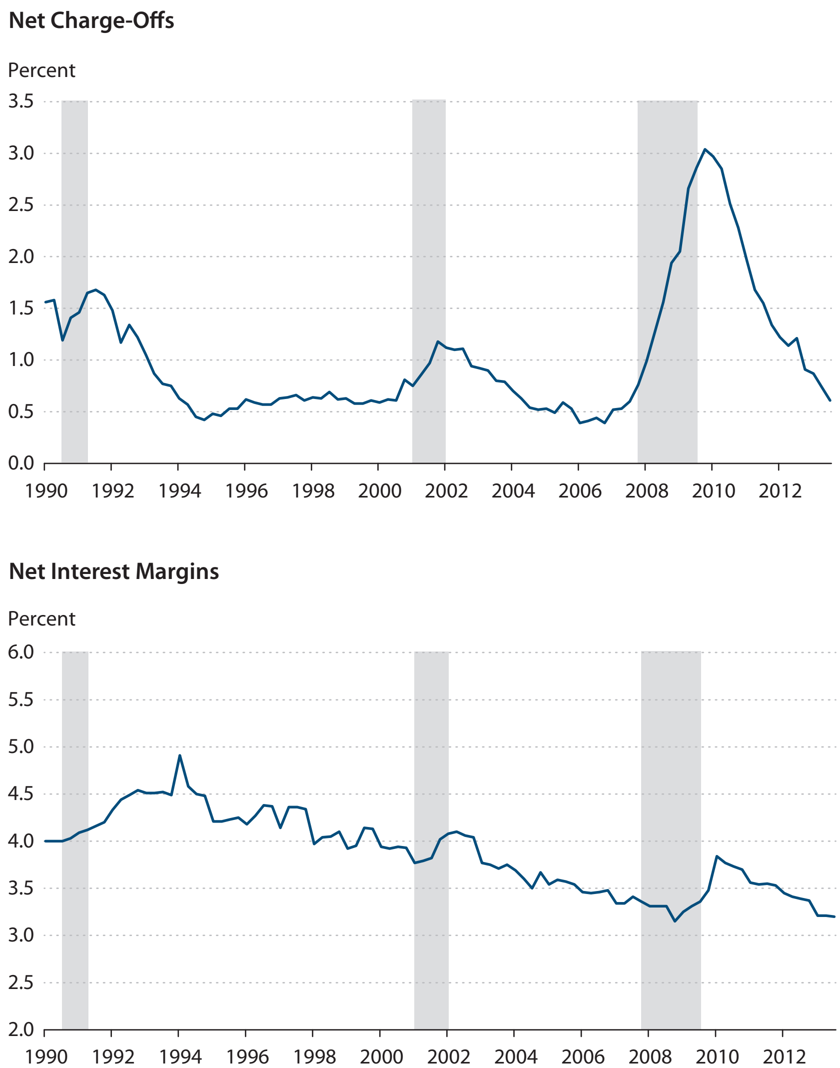
As we just noted, our goal is to develop simple-to-interpret factors that can be used to predict industry-wide bank stress when measured through the lens of either NCOs or NIMs. In much of the literature on factor-based forecasting, these factors are often extracted based on scores, if not hundreds, of macroeconomic or financial series. For example, Stock and Watson (2002) extract factors using two datasets, one consisting of 149 series and another consisting of 215 series. Ludvigson and Ng (2009) use a related dataset consisting of 132 series. Giannone, Reichlin, and Small (2008) use roughly 200 series to extract factors and nowcast current-quarter real GDP growth. Smaller datasets have also been used. Ciccarelli and Mojon (2010) extract a common global inflation factor using inflation series from only 22 countries. Engel, Mark, and West (2012) extract a common U.S. exchange rate factor using bilateral exchange rates between the United States and 17 member countries of the Organisation for Economic Co-operation and Development.

In this article, we focus on factors extracted from a collection of the 16 quarterly frequency variables (X) currently used by the Federal Reserve System to characterize the bank stress testing scenarios. We do so to facilitate a clean interpretation of our results as they pertain to bank stress testing conducted by the Federal Reserve System. See Table 1 for a complete list of these variables. The historical data for these series are from the Federal Reserve Economic Data (FRED) and Haver Analytics databases and consist of only the most recent vintages. The NCOs and NIMs data are available back to 1985:Q1 and 1984:Q1, respectively. Unfortunately, the Chicago Board Options Exchange Volatility Index (VIX) series used in the stress scenarios dates back to only 1990:Q1; hence, our analysis is based on observations from 1990:Q1 through 2013:Q3, providing a total of $T = 95$ historical observations. We considered substituting a proxy for the VIX to obtain a longer time series. Instead, we chose to restrict ourselves to the time series directly associated with the CCAR scenarios since these are the ones banks must use as part of their stress testing efforts.

Figure 1 plots both NCOs and NIMs. NCOs rose during the 1991, 2001, and 2007-09 recessions, with dramatic increases during the most recent recession. NIMs reacted less dramatically during these recessions and, more than anything, seemed to trend downward for much of the sample before rising during the recent recovery. The plots suggest that both target variables are highly persistent. Since both Dickey-Fuller (1979) and Elliott, Rothenberg, and Stock (1996) unit root tests fail to reject the null of a unit root in both series, we model the target variable in differences rather than in levels. Specifically, when the forecast horizon h is 1 quarter ahead, the dependent variable is the first difference $\Delta^1 = \Delta$, but when the horizon is four, we treat the dependent variable as the 4-quarter-ahead difference Δ^4. We should note,

Figure 1

History of Target Variables



NOTE: The shaded bars indicate recessions as determined by the National Bureau of Economic Research.

however, that while we consider the $h = 4$-quarter-ahead horizon, the majority of our results emphasize predictability at the $h = 1$-quarter-ahead horizon. As to the predictors x_i $i = 1,\ldots,16$, in most instances they are used in levels, but where necessary the data are transformed to induce stationarity. The relevant transformations are provided in Table 1.

We consider three distinct, but related, approaches to estimating the factors. By far, the PC approach is the most common in the literature. Stock and Watson (1999, 2002) show that PC-based factors (f^{PC}) are useful for forecasting macroeconomic variables, including nominal variables such as inflation, as well as real variables such as industrial production. See Stock and Watson (2011) for an overview of factor-based forecasting using PC and closely related methods. Theoretical results for factor-based forecasting can also be found in Bai and Ng (2006).

Although their use is standard, PC-based factors are not by construction useful for forecasting a given target variable $\Delta^h y$. To understand the issue, consider the simple h-step-ahead predictive model of the factors and target variables:

$$\Delta^h y_{t+h} = \beta_0 + \beta_1 f_t + \varepsilon_{t+h} \tag{1}$$

$$x_{i,t} = \alpha_{i0} + \alpha_{i1} f_t + \upsilon_{i,t} \quad i = 1, \ldots, N, \tag{2}$$

where N denotes the number of x series used to estimate the factor.

In equation (1), the factor is modeled as potentially useful for predicting the target h-steps into the future. It will be useful for forecasting if β_1 is nonzero and it is estimated sufficiently precise. Equation (2) simply states that each of the predictors x_i can be modeled as being determined by a single common factor f. When PC are used to estimate the factor, any link between equations (1) and (2) is completely ignored. Instead, the PC-based factors are constructed as follows: Let $\tilde{X}_t$ denote the $t \times N$ matrix $(\tilde{x}_{1t}, \ldots, \tilde{x}_{tt})'$ of $N \times 1$ time-t variance-standardized predictors $\tilde{x}_{st}$ $s - 1, \ldots, t$. If we define $\bar{x}_t$ as $t^{-1}\sum_{s=1}^{t} \tilde{x}_{st}$, the time-$s$ factor f_s^{PC} is estimated as $\hat{f}_{s,t}^{PC} = \hat{\Lambda}_t'(\tilde{x}_{st} = \bar{x}_t)$, where $\hat{\Lambda}_t$ denotes the eigenvector associated with the largest eigenvalue of $(\tilde{X}_t - \bar{X}_t)'(\tilde{X}_t - \bar{X}_t)$. Clearly the target variable $\Delta^h y$ plays no role whatsoever in construction of the factors, and specifically in the choice of $\hat{\Lambda}$, and hence there is no a priori reason to believe that β_1 will be nonzero.

It might therefore be surprising to find that PC-based factors are often useful for forecasting. Note, however, that the x variables used to estimate f^{PC} are never chosen randomly from a collection of all possible time series available. For example, the Stock and Watson dataset (2002) was compiled by the authors based on their years of experience in forecasting macroeconomic variables. As such, there undoubtedly was a bias toward including those variables in the dataset that had already shown some usefulness for forecasting. PC-based forecasting is therefore a reasonable choice provided the collection of predictors x is selected in a reasonable fashion.

This is a well-known issue in the factor literature; hence, methods—most notably PLS—have been developed to integrate knowledge of the target variable y into the estimation of Λ. One recently developed generalization of PLS is delineated in Kelly and Pruitt (2011). Their 3PRF estimates the factor weights Λ, taking advantage of the covariation between the predictors x and the target variable $\Delta^h y$, but does so while allowing for ancillary information z to be introduced into the estimation of the factors.[2] Effectively, this ancillary information introduces an additional equation of the form

$$z_t = \gamma_0 + \gamma_1 f_t + \nu_t \tag{3}$$

into the system above. Without going into the derivation here, if we define the vector $Z_t = (z_1,\ldots,z_t)'$, we obtain a closed-form solution for the factor loadings that takes the form

$$\hat{\Lambda}_t' = \hat{S}_{ZZ,t}\left(\hat{W}_{XZ,t}'\hat{S}_{XZ,t}\right)^{-1}\hat{W}_{XZ,t}', \tag{4}$$

where $J_t = I_t - \iota_t\iota_t'$ for I_t, the t-dimensional identity matrix ι_t is the t-vector of ones, $\hat{W}_{XZ,t} = J_N\tilde{X}_t'J_tZ_t$, $\hat{S}_{XZ,t} = \tilde{X}_t'J_tZ_t$, and $\hat{S}_{ZZ,t} = Z_t'J_tZ_t$. At first glance, it may appear that in this closed-form solution the estimated loadings do not account for the target variable; but as Kelly and Pruitt (2011) note, one can always choose to set z equal to the target variable, in which case the factor weights obviously do depend on the target variable. In fact, loosely speaking, if we set z equal to $\Delta^h y$, we obtain PLS as a special case. Kelly and Pruitt (2011) argue that there are applications in which using proxies instead of the target variable, perhaps driven by economic reasons, can lead to more-accurate predictions of the target variable. Drawing from this intuition and the results in Guerrieri and Welch (2012), in which the term spread is found to be a useful predictor of NCOs and NIMs, we use the first difference in the spread between a 10-year U.S. bond and a 3-month T-bill as our proxy when applying the 3PRF.

For each target variable we then have three sets of factors, $\hat{f}_{s,t}^{PC}, \hat{f}_{s,t}^{3PRF}$, and $\hat{f}_{s,t}^{PLS}$ $s = 1,\ldots,t$, that we can use to conduct pseudo out-of-sample forecasting exercises. Note that the PC and 3PRF factors are invariant to the target variable (and horizon) and, hence, we really have only four distinct factors. In the forecasting experiments we reestimate the factor loadings at each forecast origin. For example, if we have a full set of observations $s = 1,\ldots, T$ and forecast h-steps ahead starting with an initial forecast origin $t = R$, we obtain $P = T + h - R$ forecasts, which in turn can be used to construct forecast errors $\hat{\varepsilon}_{t+h}$ and corresponding mean squared errors (MSEs) as $MSEs = P^{-1}\sum_{t=R}^{T-h}\hat{\varepsilon}_{t+h}^2$.[3]

RESULTS

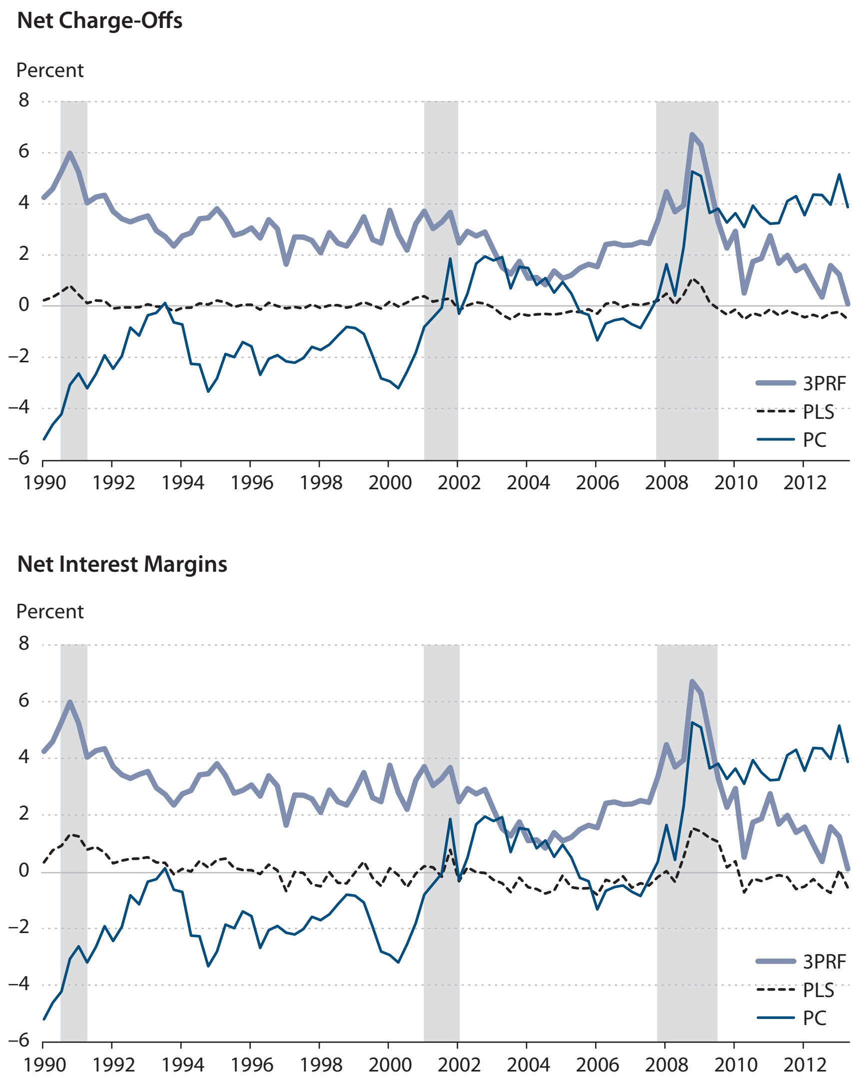
In the following, we provide a description of the factors and their usefulness for forecasting bank stress at the industry level.

The Factors

Figure 2 plots the three factors associated with the 1-step-ahead models estimated using the full sample $s = 1,\ldots,T$. The top panel plots the factors when NCOs are the target variable, while the lower panel plots the factors when NIMs are the target variable. For completeness we include the PC and 3PRF factors in both panels despite the fact they are identical in each panel. In both panels, the PLS and 3PRF factors appear stationary. Both sets of factors have peaks during both the 1991 and 2007-09 recessions, but there do not appear to be any such peaks associated with the 2001 recession. In contrast, the PC-based factors are quite interesting. Whereas the PLS and 3PRF factors appear stationary, the PC factor appears to be trending upward. In fact, one could argue that the PC factor rises sharply at the onset of each recession and then flattens but never reverts to its pre-recession level. Since stationarity of the factors is a primary assumption in much of the literature on PC-based forecasting, one might suspect

Figure 2

Estimated Historical Factors



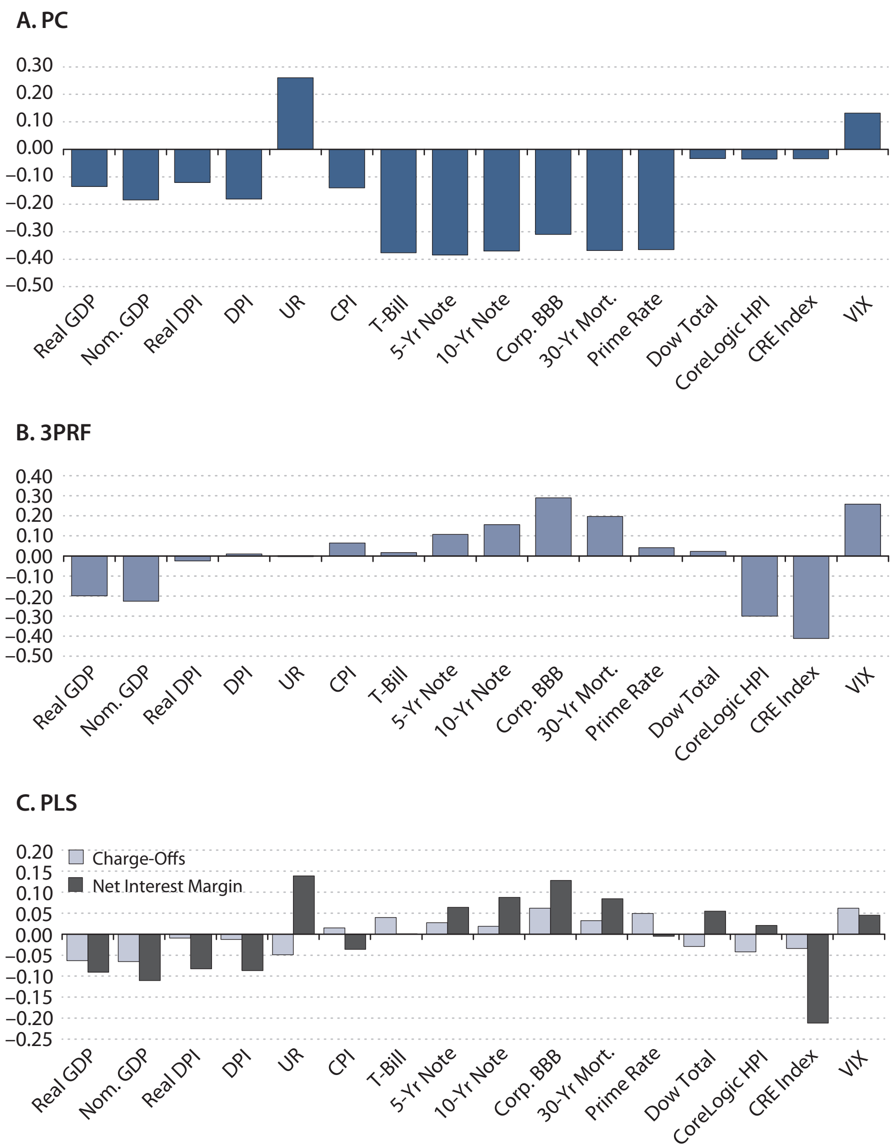
NOTE: The shaded bars indicate recessions as determined by the National Bureau of Economic Research.

that our PC-based factors will not prove useful for prediction. We return to this issue in the following section.

Figure 3 shows the factor weights $\hat{\Lambda}$ associated with the 1-step-ahead models estimated over the entire sample. There are three panels, one each for the three approaches to estimating the factors. Panel A depicts the PC weights. Since these weights are constructed on data that have been standardized, the magnitudes of the weights are comparable across variables. In addition, since the variables have all been demeaned, negative values of the weights indi-

Figure 3

Factor Weights



NOTE: BBB, Citigroup BBB-rated corporate bond; CRE, commercial real estate; DPI, disposable personal income; HPI, house price index; Mort., mortgage; Nom., nominal; UR, unemployment rate.

cate that if a variable is above its historical mean, all else equal, the factor will be smaller. With these caveats in mind, in most cases the weights are sizable, with values greater than 0.1 in absolute value. The exceptions are all financial indicators and include the Dow, the CoreLogic house price index, and the Commercial Real Estate price index. The largest weights are on the interest rate variables and are negative, sharing the same sign as that associated with real and nominal GDP as might be expected. In addition, the weights given to the unemployment rate and the VIX are positive, so they have the opposite sign of both nominal and real GDP as also might be expected.

Panel B of Figure 3 depicts the weights associated with the 3PRF-constructed factor. The weights are distinct from those associated with the PC-constructed factor, even though both sets of weights are fundamentally based on the same predictors. In particular, the weights in the financial and monetary variables are different. PC assign large negative weights on interest rates, while the 3PRF places positive weights on them. In addition, PC assign almost no weight on the Dow or either of the two price indexes, while the 3PRF assigns large negative weights on both the residential and commercial real estate price indexes. Both the PC and 3PRF assign significant, and comparably signed, weights on real and nominal GDP, as well as the VIX.

Panel C of Figure 3 has two sets of weights associated with PLS, one for each target variable. What is perhaps more interesting is how the magnitudes and signs of the weights change when the target variables are changed. As an example, note that the weight on the unemployment rate is large and positive when NIMs are the target but is modest and negative when NCOs are the target. And while PLS places a negative weight on the commercial real estate price index for both target variables, the weight is an order of magnitude larger for NIMs.

In some ways, however, the weights are similar across each of the three panels. All four sets of weights are negative for real and nominal GDP and all four are positive for the VIX. Perhaps the sharpest distinction among the three panels is the weight that PC place on the interest rate variables. Both PLS and the 3PRF place small or modestly positive weights on the interest variables, while PC place large negative weights on them.

Predictive Content of the Factors

We consider the predictive content of these factors based on both in-sample and out-of-sample evidence. Panel A of Table 2 shows the intercepts, slope coefficients, and *R*-squared values associated with the predictive regressions in equation (1) for each factor, target variable, and horizon estimated over the full sample. At each horizon and for each target variable the *R*-squared value associated with PC-based factors is smallest among the other two competitors. This is perhaps not surprising given the apparent nonstationarity of the PC-estimated factors as seen in Figure 2. When the 3PRF- and PLS-based factors are used, the *R*-squared values are typically neither large nor nontrivial. At the 1-quarter-ahead horizon, the *R*-squared values are roughly 0.20 and 0.30 for NCOs using 3PRF- and PLS-based factors, respectively. In all instances, the PLS-based factors are the largest among the three estimation methods. Even so, the PLS factors are constructed to maximize the correlation between the factor and the target variable; hence, it is not surprising that the *R*-squared values are larger for PLS. At the 1-quarter-ahead horizon, the *R*-squared values are much higher, particularly for PLS. Even so, the same

Table 2

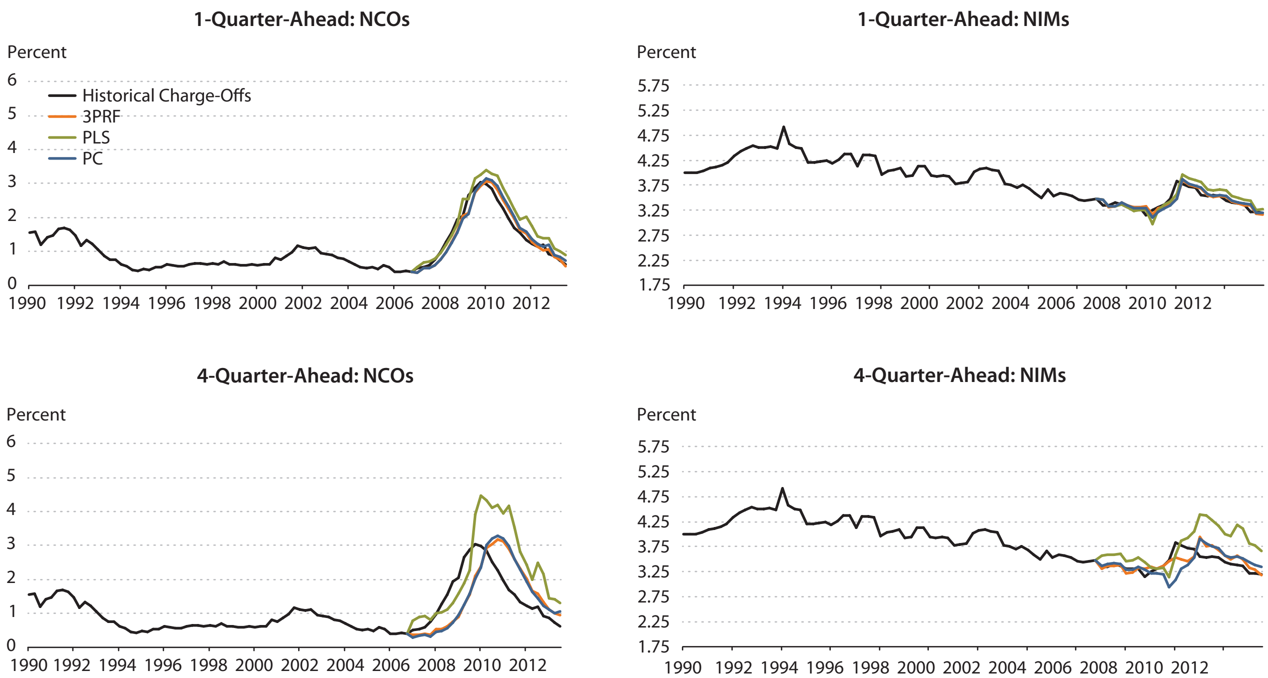
In-Sample and Out-of-Sample Evidence on Predictability

	1-Quarter-ahead horizon						4-Quarter-ahead horizon					
	NCO			NIM			NCO			NIM		
	3PRF	PLS	PC	3PRF	PLS	PC	3PRF	PLS	PC	3PRF	PLS	PC
Panel A												
Intercept	–0.165	–0.009	–0.009	–0.047	–0.008	–0.008	–0.442	–0.022	–0.022	–0.343	–0.032	–0.032
Slope	0.006	0.286	–0.004	0.014	0.049	0	0.070	0.430	–0.031	0.052	0.325	–0.001
R^2	0.192	0.287	0.003	0.020	0.049	0	0.101	0.431	0.027	0.277	0.333	0
Panel B												
RMSE, naive	0.190			0.135			0.819			0.325		
Ratio	1.096	1.988	1.277	0.723	1.027	0.778	0.966	1.420	0.994	0.490	1.430	0.782
Panel C												
PLS (MSE-*t*)	–3.120*			–2.139*			–1.407			–2.278*		
PC (MSE-*t*)	–3.365*	2.830*		–1.863*	2.025*		–1.366	1.368		–1.305	1.385	
Naive (MSE-*F*)	–4.54	–20.17	–10.45	24.65*	–1.39	17.66*	1.96	–13.62	0.32	85.37*	–13.79	17.20*

NOTE: Panel A reports OLS regression output associated with equation (1) in the text. Panel B reports RMSE for the naive, no-change model, as well as the ratio of the RMSEs from the naive and factor-based model. Numbers below 1 indicate nominally superior predictive content for the factor model. Panel C reports MSE-*t* statistics for pairwise comparisons of the factor models and MSE-*F* statistics for comparisons between the naive and factor-based models. *Indicates significance at the 5 percent level.

Figure 4

Factor-Based Forecasts of Target Variables



caveat applies and, hence, it is not clear how useful the estimated factors will be for actual predictions based solely on the in-sample *R*-squared values.

We perform a pseudo out-of-sample forecasting exercise to get a better grasp of the predictive content of these factors. Specifically, because we are most interested in predictive content during periods of bank stress, we perform a pseudo out-of-sample forecasting exercise that starts in 2006:Q4 and ends in 2013:Q3 so the forecast sample includes the Great Recession and the subsequent recovery. Figure 4 plots the path of the forecasts for each target variable, horizon, and factor estimation method. It appears that at the 1-quarter-ahead horizon the factor-based forecasts typically track the arc of the realized values of both NCOs and NIMs. Perhaps the biggest exceptions are those associated with PLS, which tend to be much higher for each target variable throughout the entire forecast period. At the 4-quarter-ahead horizon, the predictions generally track the arc of the realized values but do so with a substantial lag. As was the case for the 1-step-ahead forecasts, those associated with PLS tend to be much higher than either the PC- or 3PRF-based forecasts.

Visually, the factor-based predictions seem somewhat reasonable, but it is not clear which of the three factor-based methods performs the best. Moreover, it is even less clear whether any of the models performs better than a naive, "no-change" benchmark would. Therefore, in Panel B of Table 2 we report root mean squared errors (RMSEs) associated with forecasts made by the various factor-based models, as well as those associated with the naive benchmark. The first row reports the nominal RMSEs, while the second row reports ratios relative to the benchmark.

Particularly for NCOs, the naive benchmark tends to be a good predictor at both horizons and for both target variables. In all instances, the RMSE ratios are greater than or nearly indistinguishable from 1 and, hence, the naive model outpredicts the factor models. Predictions improve when NIMs are the target variable. At both horizons, 3PRF- and PC-based predictions are better than the naive benchmark. In some cases, the improvements are as large as 50 percent.

To better determine the magnitudes of the improvements and which factor estimation method performs best, we conduct tests of equal MSEs across each of the possible pairwise comparisons, holding the horizon and target variable constant. Specifically, we first construct the statistic

$$MSE\text{-}t = \frac{P^{-1/2}\sum_{t=R}^{T-h}\hat{d}_{ij,t+h}}{\hat{\sigma}_{i,j}} \tag{5}$$

for each of the pairwise comparisons across factor-type $i \neq j \in \{3PRF, PLS, PC\}$, where $\hat{d}_{ij,t+h} = \hat{\varepsilon}^2_{i,t+h} - \hat{\varepsilon}^2_{j,t+h}$. At the 1-quarter-ahead horizon the denominator is constructed using $\hat{\sigma}_{i,j} = \left(P^{-1}\sum_{t=R}^{T-h}\left(\hat{d}_{ij,t+h} - \overline{d}_{i,j}\right)^2\right)^{1/2}$, while at the 4-quarter-ahead horizon we use a Newey-West (1987) autocorrelation-consistent estimate of the standard error allowing for three lags. We use these statistics to test the null of equal population-level forecast accuracy $H_0 : E(\varepsilon^2_{i,t+h} - \varepsilon^2_{j,t+h}) = 0$ $t = R,\ldots,T-h$. In each instance, we consider the two-sided alternative $H_A : E(\varepsilon^2_{i,t+h} - \varepsilon^2_{j,t+h}) \neq 0$.

For the comparisons between the naive, no-change model and the factor-based models, we use the MSE-*F* statistic developed in Clark and McCracken (2001, 2005),

$$MSE\text{-}F = \frac{\sum_{t=R}^{T-h}\hat{d}_{ij,t+h}}{P^{-1}\sum_{t=R}^{T-h}\hat{\varepsilon}_{j,t+h}}, \tag{6}$$

for $i = Naive$ and $j \in \{3PRF, PLS, PC\}$. This statistic is explicitly designed to capture the fact that the two models are nested under the null of equal population-level predictive ability. For the comparisons between the naive and the factor-based models, we consider the one-sided alternative $H_A : E(\varepsilon^2_{naive,t+h} - \varepsilon^2_{j,t+h}) > 0$ and hence reject when the MSE-*F* statistic is sufficiently large.

For each of the pairwise factor-based comparisons, we treat the MSE-*t* statistic as asymptotically standard normal and hence conduct inference using the relevant critical values. As a technical matter, the appropriateness of standard normal critical values has not been proven in the literature. Neither the results in Diebold and Mariano (1995) nor those in West (1996) are directly applicable to situations where generated regressors are used for prediction. Even so, we follow the literature and use standard normal critical values when the comparison is between two non-nested models estimated by OLS and evaluated under quadratic loss.

Unfortunately, this approach to inference does not extend to the MSE-*F* statistic used for the nested model comparisons. Instead, we consider a fixed regressor wild bootstrap as described in Clark and McCracken (2012). The 1-step-ahead bootstrap algorithm proceeds as follows:

(i) Generate artificial samples of $\Delta y_t = y_t - y_{t-1}$ using the wild-form $\Delta y_t^* = \varepsilon_t^* = \varepsilon_t \eta_t = (y_t - y_{t-1})\eta_t$ for a simulated i.i.d. sequence of standard normal deviates η_t $t = 1,\ldots,T$.

(ii) Conduct the pseudo out-of-sample forecasting exercise precisely as was done in Table 2 but use ε_t^* as the dependent variable rather than $y_t - y_{t-1} = \varepsilon_t$. Construct the associated MSE-*F* statistic.

(iii) Repeat steps (i) and (ii) $J = 499$ times.

(iv) Estimate the $(100 - \alpha)$ percentile of the empirical distribution of the bootstrapped MSE-*F* statistics and use it to conduct inference.

For the 4-quarter-ahead case, the bootstrap differs only in how the dependent variable is simulated. Note that if $y_t = y_{t-1} + \varepsilon_t$, then $\Delta^4 y = y_t - y_{t-4} = (y_t - y_{t-1}) + (y_{t-1} - y_{t-2}) + (y_{t-2} - y_{t-3}) + (y_{t-3} - y_{t-4}) = \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \varepsilon_{t-3}$. If we define the dependent variable as $\Delta^4 y_t^* = \varepsilon_t^* + \varepsilon_{t-1}^* + \varepsilon_{t-2}^* + \varepsilon_{t-3}^* = \varepsilon_t \eta_t + \varepsilon_{t-1}\eta_{t-1} + \varepsilon_{t-2}\eta_{t-2} + \varepsilon_{t-3}\eta_{t-3}$ in step (i) of the algorithm, the remainder of the bootstrap proceeds as stated. Note, however, as was the case above, this bootstrap was developed to work in environments where the predictors were observed and not generated. Even so, we apply the bootstrap as designed in order to at least attempt to use critical values designed to accommodate a comparison between nested models.

The lower panel of Table 2 lists the results. Tests significant at the 5 percent level are indicated by an asterisk. First, consider the nested comparisons between the naive benchmark with each of the factor-based models. When NCOs are the target variable, we never reject the null of equal accuracy in favor of the alternative, suggesting that the factor-based models improve forecast accuracy. In contrast, when NIMs are the target variable, we find significant evidence that PC- and 3PRF-based forecasts improve on the naive model at both the 1- and 4-quarter-ahead horizons.

Now consider the pairwise comparisons across factor-based models. Here there is abundant evidence suggesting that the PLS-based approach to forecasting provides significantly less-accurate forecasts than those from PC- and 3PRF-based approaches. At each horizon for NIMs and at the 1-quarter-ahead horizon for NCOs, each of the MSE-*t* statistics is significantly different from zero for any comparison with PLS. Somewhat surprisingly, at the 4-quarter-ahead horizon and for NCOs, we fail to reject the null of equal accuracy in each comparison with PLS despite the fact that the PLS-based forecasts have much higher nominal MSEs.

Overall, a few conclusions on the predictive content of the factors seem reasonable. First, and perhaps most obvious, the PLS-based factors do not predict well at any horizon and for either target variable. In each instance, at least one of the PC- or 3PRF-based models forecasts much more accurately in an MSE sense. This is despite the fact that in sample, the PLS approach had the highest R-squared value, even though, as already stated, this is by construction and not particularly informative. Second, at the 1-quarter horizon, the 3PRF-based factors perform better than each of the competing factor-based forecast models. This is particularly true for NIMs. Even so, for NCOs this is small consolation given that the naive model does even better. Finally, at the longer horizons, the 3PRF-based factors are nominally the best but even so, the MSE-*t* statistics indicate that they are not significantly smaller than those associated with the PC-based factors for either target variable.

IMPLIED STRESS IN THE 2014 CCAR SCENARIOS

In the previous section, we established that the factors had varying degrees of predictive content for their corresponding target variable. In this section, we assess the degree to which the factors indicate stress in the context of the 2014 CCAR scenarios. That is, we attempt to measure the degree of stress faced by the banking industry in each of the baseline, adverse, and severely adverse scenarios when viewed through the lens of the implied counterfactual factors. We then conduct conditional forecasting exercises akin to those required by the Dodd-Frank Act. That is, we construct forecasts of NCOs and NIMs conditional on the counterfactual factors implied by the various scenarios and types of factor estimation methods.

The Counterfactual Factors

Figure 5 plots the factors associated with the 1-quarter-ahead horizon when the counterfactual scenarios are used to estimate the factors. To do so, we first estimate the four sets of factor weights through 2013:Q3, yielding $\hat{\Lambda}_{2013:Q3}$. Where necessary, we then demean and standardize the hypothetical variables in each of the baseline, adverse, and severely adverse scenarios using the historical sample means and standard deviations. Together these allow us to estimate the counterfactual factors $\hat{f}_{t,2013:Q3}$ for all t = 2013:Q4,...,2016:Q3.

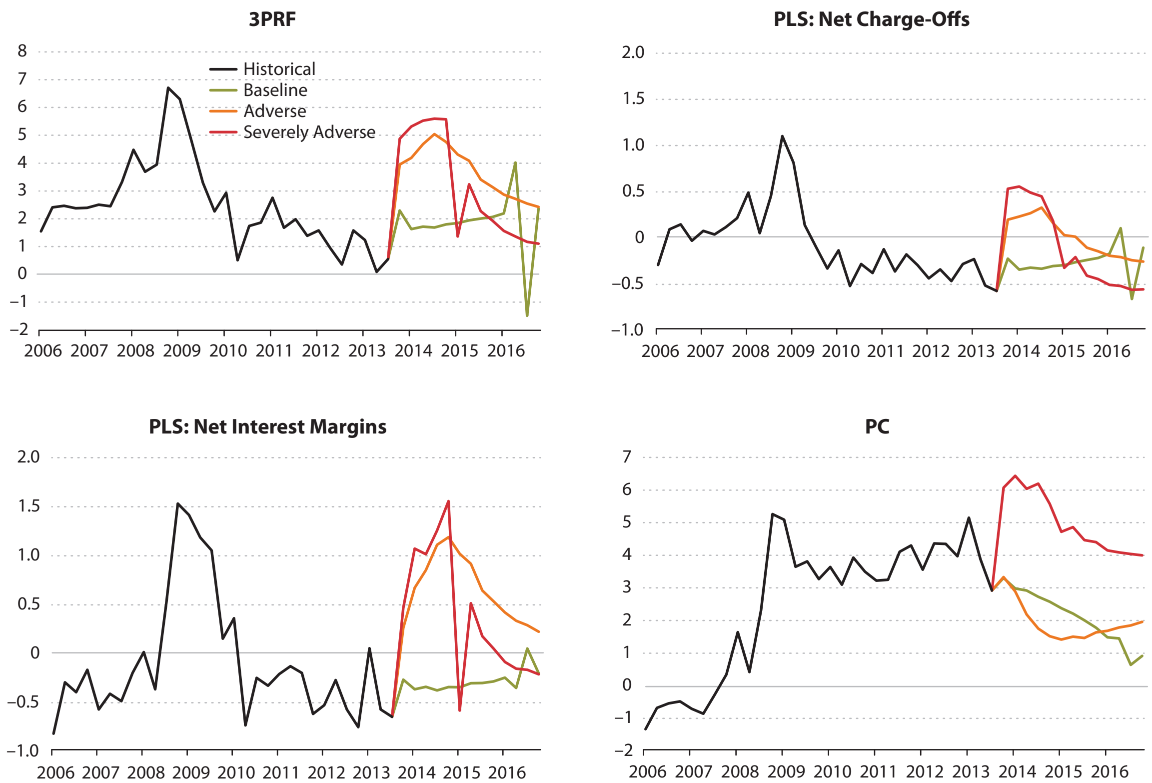
There are four panels in Figure 5, with one each for the 3PRF- and PC-based factors. There are two panels for PLS, one for each target variable. The black lines in the panels show the factors estimated over the historical sample and end in 2013:Q3. The remaining three lines correspond to a distinct scenario starting in 2013:Q4 and ending in 2016:Q4.

The three panels associated with the PLS and 3PRF-based factors show comparable results for the degree of severity of the scenarios. In each case, the hypothetical severely adverse factors reach the highest levels and remain elevated for at least the first full year of the scenario. Somewhat surprisingly, in each of these three cases the factors associated with the adverse scenario tend to be only modestly lower than those associated with the severely adverse scenario over the first year. In fact, the factors associated with the adverse scenario remain elevated throughout much of the horizon even after those associated with the severely adverse scenario have declined to levels more closely associated with the baseline levels, which appear to remain near the historical mean of (the first difference of) NCOs and NIMs. Finally, it is instructive to note that the levels of the factors in the severely adverse scenario reach maximum values near those observed during the Great Recession.

The final panel in Figure 5 shows the results for the PC-based factors. As before, the PC-based factors associated with the severely adverse scenario reach the highest levels of all three

Figure 5

Counterfactual Factors Based on Scenarios



scenarios. In contrast to the other three scenarios, here the hypothetical factors in the severely adverse scenario are always the largest of the three. Moreover, the levels are the same, if not a bit higher, than those observed during the Great Recession. Somewhat oddly, the factor associated with the adverse scenario is not particularly adverse and, in fact, often takes values lower than the factor associated with the baseline scenario.

Conditional Forecasting

In general, the paths of the counterfactual factors seem reasonable. The severely adverse factors are generally—albeit not always—higher than those associated with the adverse scenario, which in turn, are higher than those associated with the baseline scenario. Even so, this visual evaluation of the counterfactual factors does not by itself imply anything about the magnitude of stress faced by the banks over the scenario horizon. To get a better grasp of such stress, we conduct a conditional forecasting exercise akin to that required by the Dodd-Frank Act. Specifically, starting in 2013:Q3 we construct a path of 1-step-ahead through 4-step-ahead forecasts of both NCOs and NIMs, conditional on the counterfactual factors for each scenario.

In light of the apparent benefit of using the 3PRF-based factors, we focus exclusively on conditional forecasts using these factors.

Methodologically, our forecasts are constructed using the minimum MSE approach to conditional forecasting delineated in Doan, Litterman, and Sims (1984) designed for use with VARs. To implement this approach we first link the 1-step-ahead predictive equation from equation (1) with an AR(1) for the factors to form a restricted VAR of the form

$$\begin{pmatrix} \Delta y_t \\ f_t \end{pmatrix} = \begin{pmatrix} \alpha_y \\ \alpha_f \end{pmatrix} + \begin{pmatrix} 0 & \beta_y \\ 0 & \beta_f \end{pmatrix} \begin{pmatrix} \Delta y_{t-1} \\ f_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{yt} \\ \varepsilon_{ft} \end{pmatrix}. \tag{7}$$

We then estimate this system by OLS using the historical sample ranging from 1990:Q1 through 2013:Q3. The coefficients in the first equation line up with those already presented in Table 2. The coefficients in the second equation are not reported here but imply a root of the lag polynomial of roughly 0.80.[4] Using this system of equations, along with an assumed future path of the factors for all t = 2013:Q4,...,2016:Q3, we construct forecasts across the entire scenario horizon. We do so separately for each target variable and each scenario. For the sake of comparison, we also include the path of the unconditional forecast associated with equation (7) using an iterated multistep approach to forecasting.

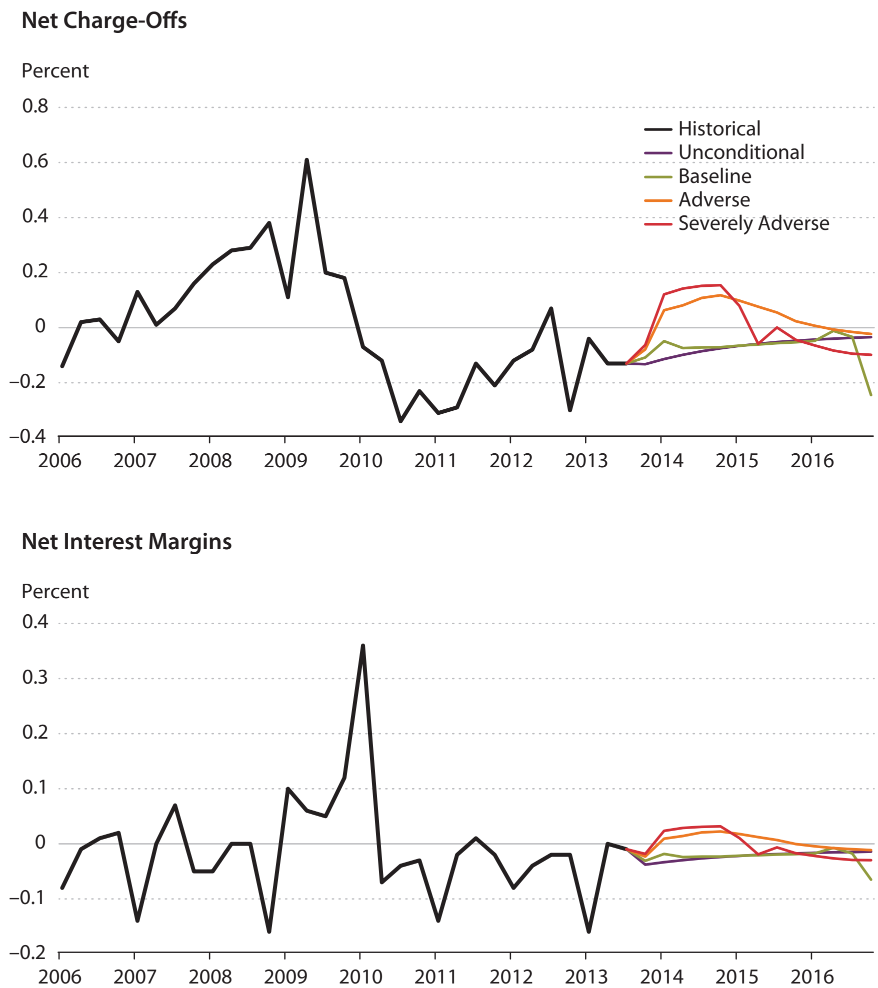
Figure 6 shows the conditional forecasts. The upper panel is associated with NCOs, while the latter is associated with NIMs. In each panel, we plot the historical path of the change in the target variable though 2013:Q3, at which point we plot four distinct forecast paths: one associated with the unconditional forecast and one associated with each of the three scenarios. The two plots are similar in many ways. The unconditional forecasts converge smoothly to the unconditional mean over the forecast horizon. As expected, the baseline forecasts differ little from those associated with the unconditional forecast.

In contrast, the adverse and severely adverse forecasts tend to be significantly higher than the unconditional forecast at least through mid-2015 for NCOs and early 2016 for NIMs. Through the end of 2014, the severely adverse predictions are higher than any of the other scenarios at any point over the forecast horizon. However, in early 2015 the forecasts associated with the severely adverse scenario decline sharply and thereafter track the baseline forecasts. The same path does not apply to the adverse scenario forecasts. They remain relatively higher until the end of 2015 before converging to those associated with the baseline forecasts.

The paths of the forecasts seem reasonable in the sense that the forecast for the severely adverse scenario is the most severe, followed by the adverse and then the baseline forecasts. Even so, the magnitudes of the forecasts, especially those associated with the severely adverse scenario, do not match the levels of either NCOs or NIMs observed during the Great Recession. On the one hand, this might be viewed as a failure of the model. But upon reflection, it is not so surprising that the forecasts do not achieve the levels observed during and immediately following the financial crisis. These forecasts are constructed based solely on a limited number of macroeconomic aggregates and do not include many of the series known to have been problematic during the crisis. These include, but are certainly not limited to, measures of financial market liquidity and counterparty risk, as well as measures related to monetary and fiscal

Figure 6

Forecasts of Target Variables Conditional on Future Path of Factors



policy. Variables such as money market rates and the TED spread (U.S. T-bill–eurodollar spread) would almost certainly accommodate these bank stress concerns but are not used here because they are not present in the CCAR scenarios. In short, while the conditional forecasts based on macroeconomic aggregates provide some information on the expected path of future NCOs and NIMs, they by no means paint a complete picture.

CONCLUSION

In this article, we investigate the usefulness of factor-based methods for assessing industry-wide bank stress. We consider standard PC-based approaches to constructing factors, as well as methods such as PLS and the 3PRF, that design the weights to relate the predictive content of the factor to a given target variable. As is common in the stress testing literature, we evaluate industry-wide bank stress based on the level of NCOs and NIMs.

We provide two types of evidence related to the usefulness of these indexes. Based on a pseudo out-of-sample forecasting exercise, we show that our 3PRF-based factor appears to typically provide the best or nearly the best forecasts for NCOs and NIMs. This finding holds at both the 1-quarter- and 4-quarter-ahead horizons. Admittedly, the best factor model performs worse than the naive, no-change model for NCOs, but for NIMs the factor models do appear to have marginal predictive content beyond that contained in the no-change model.

We then study how these factor-based methods might help policymakers evaluate the degree of stress present in the stress testing scenarios. To do so, we use the factor weights estimated using historical data, along with the hypothetical scenarios as presented in the official stress testing documentation, to construct hypothetical stress factors. At the 1-quarter-ahead horizon, these factors generate very intuitive graphical representations of the level of stress faced by the banking industry as measured through the lens of NCOs and NIMs. We conclude with a conditional forecasting exercise designed to provide some indication of the future path of NCOs and NIMs should one of the counterfactual scenarios occur. ■

NOTES

[1] The scenarios are available on the Federal Reserve Board of Governors website (http://www.federalreserve.gov/bankinforeg/bcreg20131101a1.pdf).

[2] We focus exclusively on the case of one factor throughout and hence consider only one proxy variable throughout. The theory developed in Kelly and Pruitt (2011) permits more than one proxy variable and thus more than one factor.

[3] For brevity, we let *P* denote the number of forecasts regardless of the horizon *h*.

[4] We also considered lags of Δy_t in the second equation. In all instances, they were insignificantly different from zero at conventional levels.

REFERENCES

Acharya, Viral V.; Engle, Robert and Pierret, Dianne. "Testing Macroprudential Stress Tests: The Risk of Regulatory Risk Weights." NBER Working Paper No. 18968, National Bureau of Economic Research, April 2013; http://www.nber.org/papers/w18968.pdf?new_window=1.

Bai, Jushan and Ng, Serena. "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions." *Econometrica*, July 2006, *74*(4), pp. 1133-50; http://www.jstor.org/stable/3805918.

Bolotnyy, Valentin; Edge, Rochelle M. and Guerrieri, Luca. "Stressing Bank Profitability for Interest Rate Risk." Unpublished manuscript, 2013, Board of Governors of the Federal Reserve System.

Ciccarelli, Matteo and Mojon, Benoit. "Global Inflation." *Review of Economics and Statistics*, August 2010, *92*(3), pp. 524-35; http://www.mitpressjournals.org/doi/pdf/10.1162/REST_a_00008.

Clark, Todd E. and McCracken, Michael W. "Tests of Equal Forecast Accuracy and Encompassing for Nested Models." *Journal of Econometrics*, November 2001, *105*(1), pp. 85-110; http://www.sciencedirect.com/science/article/pii/S0304407601000719?via=ihub.

Clark, Todd E. and McCracken, Michael W. "Evaluating Direct Multistep Forecasts." *Econometric Reviews*, 2005, *24*(4), pp. 369-404.

Clark, Todd E. and McCracken, Michael W. "Reality Checks and Comparisons of Nested Predictive Models." *Journal of Business and Economic Statistics*, January 2012, *30*(1), pp. 53-66.

Covas, Francisco B.; Rump, Ben and Zakrajsek, Egon. "Stress-Testing U.S. Bank Holding Companies: A Dynamic Panel Quantile Regression Approach." Finance and Economic Discussion 2013-155, Board of Governors of the Federal Reserve System, September 2013; http://www.federalreserve.gov/pubs/feds/2013/201355/201355pap.pdf.

Dickey, David A. and Fuller, Wayne A. "Distribution of the Estimators for Autoregressive Time Series with a Unit Root." *Journal of the American Statistical Association*, 1979, *74*(366), pp. 427-31; http://www.jstor.org/stable/2286348.

Diebold, Francis X. and Mariano, Roberto S. "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics*, July 1995, *13*(3), pp. 253-63; http://www.tandfonline.com/doi/abs/10.1080/07350015.1995.10524599#.UzRfSF8o6Uk.

Doan, Thomas; Litterman, Robert and Sims, Christopher. "Forecasting and Conditional Projection Using Realistic Prior Distributions." *Econometric Reviews*, 1984, *3*(1), pp. 1-100; http://www.tandfonline.com/doi/abs/10.1080/07474938408800053#.UzRdBF8o6Uk.

Elliott, Graham; Rothenberg, Thomas J. and Stock, James H. "Efficient Tests for an Autoregressive Root." *Econometrica*, July 1996, *64*(4), pp. 813-36; http://www.jstor.org/stable/2171846.

Engel, Charles; Mark, Nelson C. and West, Kenneth D. "Factor Model Forecasts of Exchange Rates." Working Paper No. 012, University of Notre Dame Department of Economics, January 2012; http://www3.nd.edu/~tjohns20/RePEc/deendus/wpaper/012_rates.pdf.

Giannone, Domenico; Reichlin, Lucrezia and Small, David. "Nowcasting: The Real-Time Informational Content of Macroeconomic Data." *Journal of Monetary Economics*, May 2008, *55*(4), pp. 665-76; http://www.sciencedirect.com/science/article/pii/S0304393208000652#.

Guerrieri, Luca and Welch, Michelle. "Can Macro Variables Used in Stress Testing Forecast the Performance of Banks?" Finance and Economics Discussion Series 2012-49, Board of Governors of the Federal Reserve System, July 2012; http://www.federalreserve.gov/pubs/feds/2012/201249/201249pap.pdf.

Kelly, Bryan T. and Pruitt, Seth. "The Three-Pass Regression Filter: A New Approach to Forecasting with Many Predictors." Research Paper No. 11-19, Fama-Miller Working Paper Series, University of Chicago Booth School of Business, January 2011; http://faculty.chicagobooth.edu/bryan.kelly/research/pdf/Forecasting_theory.pdf.

Ludvigson, Sydney C. and Ng, Serena. "Macro Factors in Bond Risk Premia." *Review of Financial Studies*, December 2009, *22*(12), pp. 5027-67.

Newey, Whitney K. and West, Kenneth D. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*, May 1987, *55*(3), pp. 703-08; http://www.jstor.org/stable/1913610.

Stock, James H. and Watson, Mark W. "Forecasting Inflation." *Journal of Monetary Economics*, October 1999, *44*(2), pp. 293-335; http://www.sciencedirect.com/science/article/pii/S0304393299000276?via=ihub.

Stock, James H. and Watson, Mark W. "Macroeconomic Forecasting Using Diffusion Indexes." *Journal of Business and Economic Statistics*, April 2002, *20*(2), pp. 147-62; http://dx.doi.org/10.1198/073500102317351921.

Stock, James H. and Watson, Mark W. "Dynamic Factor Models," in Michael P. Clements and David F. Hendry, eds., *Oxford Handbook of Economic Forecasting*. New York: Oxford University Press, 2011, pp. 35-59.

West, Kenneth D. "Asymptotic Inference About Predictive Ability." *Econometrica*, September 1996, *64*(5), pp. 1067-84; http://www.jstor.org/stable/2171956.

100 YEARS™
FEDERAL RESERVE SYSTEM

FRED®, the St. Louis Fed's Force of Data

Katrina Stierholz, Vice President and Chief Librarian, Research Division

No history of the St. Louis Fed would be complete without a chapter on its leadership in providing economic data for the public. Today, this long-standing commitment to data service is encapsulated in one name: FRED®.

FRED, Federal Reserve Economic Data, is the St. Louis Fed's main economic database. It is used by hundreds of thousands of economists, other researchers, market analysts, journalists, teachers, students, and members of the general public. FRED began about 20 years ago as a modest service provided by pre-Internet technology. Since then, it has continued to flourish and expand significantly in size and scope, usability and functionality, and number of users. It is known and valued around the world in many different ways by people who care about the numbers driving today's regional, national, and international economies. FRED data have been incorporated into textbooks, blogs, social media, and research papers.

Left to right: Leonall Andersen, St. Louis Fed economist, and Homer Jones, St. Louis Fed director of research, discuss the presentation of data, circa 1961.

FRED's Origins

FRED is a descendent of the data publications created by Homer Jones, the research director of the St. Louis Fed from 1958 to 1971. Jones believed in the benefits of making economic data widely available not just to policymakers but also to the general public, to allow them to judge for themselves the state of the economy and the outcome of policy. So he began publishing periodic reports on the latest economic data available.[1]

The "technology" available at the time was paper, so the data were printed and mailed through the U.S. postal system—at first to Jones's colleagues and then to a growing list of other interested subscribers. Eventually, the number of subscribers increased significantly, as did subscribers' reliance on the publications. St. Louis Fed employees from the 1970s and 1980s recall intense pressure to publish the Bank's main data publication, the weekly *USFD* (*U.S. Financial Data*). For example, reporters would repeatedly call the Bank each Thursday, the *USFD*'s release day, asking for the data so they could publish them in the next day's newspaper.

Pam Hauck, a Fed employee who worked on these data publications, described the stream of inquiries they received by phone on release days:

> The phone is ringing. "Got the data yet?" "No, we haven't got it yet." "Got the data yet?" "No, we haven't got it yet." Finally, the data became available and then for literally three hours it seemed the phone was non-stop ringing… Reporters, economists, college professors, students, a lot of students…

FRED's Progress

The St. Louis Fed's data publications evolved into a high-profile service that soon included multiple series of monetary data, national economic data, and international economic data. These paper data publications also translated well to online posting when that technology became available.

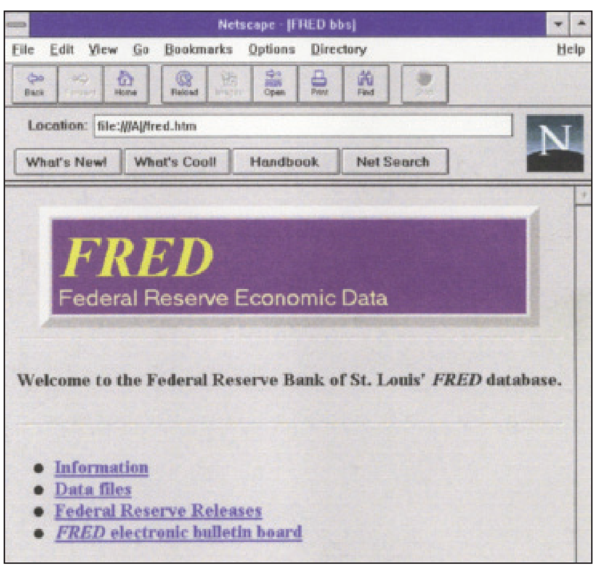
FRED got its start in 1991 as an electronic bulletin board—a precursor to the Internet. FRED offered "free up-to-the-minute economic data via modems connected to personal computers"; St. Louis Fed employees described the public's response as "staggering" and "overwhelming," with usage throughout the St. Louis area and places as distant as Vancouver, London, and Taipei.[2]

Reuters has noted that FRED is "among the more popular pages with economists and consultants"; Barron's has called it "a terrific resource."

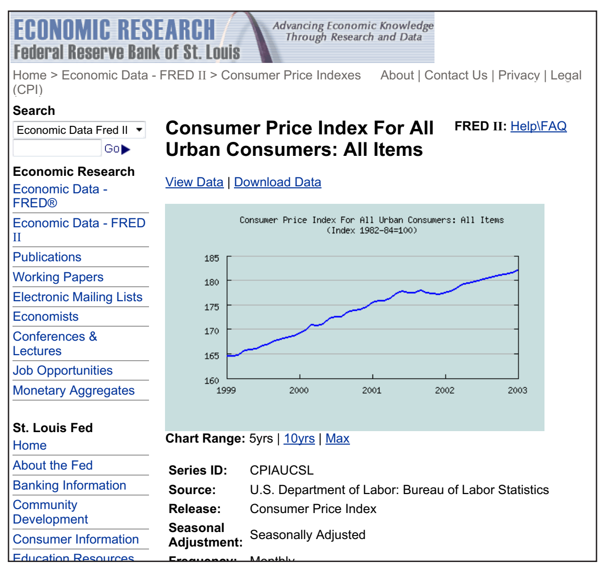
Initially, FRED had 620 users who were given access to 30 data series that could be downloaded in ASCII (an early system of electronic code) at a modem speed of up to 14.4 kilobits per second. Users were allowed one hour of bulletin board use per day. Over time, data series from other St. Louis Fed publications were added, with more than 300 series available in 1993.

In the mid-1990s, FRED made the transition to the Internet. It contained 865 data series by then and was accessed an average of 6,000 times per week. (At the time, there were only an estimated 12 million people using the Internet. The St. Louis Fed provided a shared computer station for employees to access the Internet, but only those who had a business need to do so.)

FRED has evolved dramatically over the past 20 years: By 2004, FRED offered more than 2,900 economic data series and the ability to download data in Excel



FRED website screenshot, circa 1996.



FRED website screenshot, circa 2003.

FRED website screenshot, June 1, 2014.

and text formats. Users were also given tools to create graphs of the data. By November 2010, FRED had expanded to more than 24,000 data series, which included more than 21,000 regional series.

A November 28, 2012, post on the Economist.com website went beyond the numbers and summed up the mission behind FRED: "The Federal Reserve Bank of St. Louis provides a valuable public good for observers of the American economic scene..."

FRED Today

At the time of this writing, FRED contains more than 217,000 regional, national, and international economic data series, with data coming in from 67 reporting agencies around the world. It operates on a high-speed Ethernet service (with download speeds in the millions of bits per second), provides state-of-the-art graphing software, is available through apps on smartphones and tablets, can be mapped, is accessible through Excel, and is used in thousands of classrooms. What began as a simple, printed data publication has grown into a sophisticated and successful vehicle for freely sharing important economic data with anyone around the world. ■

NOTES

1 Jones's first "data memo," sent to a few colleagues in 1961, included only three tables of data: rates of change of (i) the money supply, (ii) the money supply plus time deposits, and (iii) the money supply plus time deposits plus short-term government securities.

2 "Introducing FRED..." Federal Reserve Bank of St. Louis *Eighth Note*, May/June 1991.

FRED Quotes

National newspapers, periodicals, and sources of economic and financial analysis have followed and touted FRED's progress over the years.

"FRED…offers a mix of information as diverse as consumer credit and the money supply [and] allows users looking for data to have it E-mailed directly to them…"
New York Times, 7/15/1996

"The St. Louis Federal Reserve [is] one of the pioneers in setting up a home page… Among the more popular pages with economists and consultants is FRED."
Reuters News, 10/8/1996

"Another great tool is the Federal Reserve Economic Data database on the St. Louis Fed site…easy to navigate and free… This is a terrific resource."
Barron's, 5/3/2004

"The St. Louis Fed's massive online collection of data and analysis—called Federal Reserve Economic Data or FRED—is drawing notice and praise from numbers geeks all over."
St. Louis Business Journal, 11/1/2012

"The Federal Reserve Bank of St. Louis provides a valuable public good for observers of the American economic scene..."
Economist.com, 11/28/2012

"FRED is a leading economics data system recognized the world over. In 2011 alone, FRED was accessed by nearly 2 million individuals from over 200 countries who created over 600,000 custom graphs using its data series…"
Best of the Best Business Websites, 12/22/2012

"FRED Economic Data, free from the Federal Reserve Bank of St. Louis, is available for Apple and Android devices…FRED displays the numbers, turns them into simple charts, and supplies enough of a narrative to make some sense of it all."
Philadelphia Inquirer, 3/6/2013

FRED Stats

FRED by the numbers, as of June 4, 2014

223,000 economic time series

88,531 international series

37,170,431 data points

67 sources

213 countries' data contained in FRED

227 countries' residents using FRED

715 published data lists

3 months of FRED blog posts

2,600+ FRED graphs embedded in other websites

500+ user dashboards created

6 additional doors to FRED data:
GeoFRED, ALFRED, Excel Add-in, API, FRED app, mobile FRED

William Poole was president of the St. Louis Fed from 1998 to 2008. He often spoke to civic and professional organizations, universities, and the public about economic conditions and monetary policy, which often included reference to economic data. The article that follows was originally published in the March/April 2007 issue of the Federal Reserve Bank of St. Louis *Review*. In this article, Poole provides an especially detailed perspective on both the facts and the philosophies behind the St. Louis Fed's tradition of providing economic data to the public. As the Federal Reserve System commemorates its first 100 years of service, the Research Division of the St. Louis Fed hopes that our subscribers continue to find our services relevant and useful today.

Data, Data, and Yet More Data

William Poole

This article was originally presented as a speech at the Association for University Business and Economic Research (AUBER) Annual Meeting, University of Memphis, Memphis, Tennessee, October 16, 2006.

Federal Reserve Bank of St. Louis *Review*, March/April 2007, *89*(2), pp. 85-89.

I've long had an interest in data, and I think that this topic is a good one for this conference. The topic is also one I've not addressed in a speech.

A personal recollection might be a good place to begin. In the early 1960s, in my Ph.D. studies at the University of Chicago, I was fortunate to be a member of Milton Friedman's Money Workshop. Friedman stoked my interest in flexible exchange rates, in an era when mainstream thinking was focused on the advantages of fixed exchange rates and central banks everywhere were committed to maintaining the gold standard. Well, I should say central banks almost everywhere, given that Canada had a floating-rate system from 1950 to 1962. Friedman got me interested in doing my Ph.D. dissertation on the Canadian experience with a floating exchange rate, and later I did a paper on nine other floating rate regimes in the 1920s. For this paper I collected daily data on exchange rates from musty paper records at the Board of Governors in Washington.

What was striking about the debates over floating rates in the 1950s is that economists were so willing to speculate about how currency speculators would destabilize foreign exchange markets without presenting any evidence to support those views. In this and many other areas, careful empirical research has resolved many disputes. Our profession has come a long way in institutionalizing empirical approaches to resolving empirical disputes. The enterprise requires data, and what I will discuss is some of the history of the role of the Federal Reserve Bank of St. Louis in providing the data.

Before proceeding, I want to emphasize that the views I express here are mine and do not necessarily reflect official positions of the Federal Reserve System. I thank my colleagues at the Federal Reserve Bank of St. Louis for their comments. Robert H. Rasche, senior vice president and director of research, provided special assistance.

ORIGINS

The distribution of economic data by the Research Division of the Federal Reserve Bank of St. Louis can be traced back at least to May 1961. At that time, Homer Jones, then director of research, sent out a memo with three tables attached showing rates of change of the money supply (M1), money supply plus time deposits, and money supply plus time deposits plus short-term government securities. His memo indicated that he "would be glad to hear from anyone who

William Poole is the president of the Federal Reserve Bank of St. Louis. The author thanks colleagues at the Federal Reserve Bank of St. Louis. Robert H. Rasche, senior vice president and director of research, provided special assistance. The views expressed are the author's and do not necessarily reflect official positions of the Federal Reserve System.

thinks such time series have value, concerning promising applications or interpretations." Recollections of department employees from that time were that the mailing list was about 100 addressees.

Apparently Homer received significant positive feedback, since various statistical releases emerged from this initial effort. Among these were *Weekly Financial Data*, subsequently *U.S. Financial Data*; *Bank Reserves and Money*, subsequently *Monetary Trends*; *National Economic Trends* (1967) and *International Economic Trends* (1978), all of which continue to this date. In April 1989, before a subscription price was imposed, the circulation of *U.S. Financial Data* had reached almost 45,000. A *Business Week* article published in 1967 commented about Homer that "while most leading monetary economists don't buy his theories, they eagerly subscribe to his numbers." As an aside, as a Chicago Ph.D., I both bought the theories and subscribed to the data publications. By the late 1980s, according to Beryl Sprinkel (1987, p. 6), a prominent business economist of the time, "weekly and monthly publications of the Research Division, which have now become standard references for everyone from undergraduates to White House officials, were initially Homer's products."

Why should a central bank distribute data as a public service? Legend has it that Homer Jones viewed as an important part of his mission providing the general public with timely information about the stance of monetary policy. In this sense he was an early proponent, perhaps the earliest proponent, of central bank accountability and transparency. While Homer was a dedicated monetarist, and data on monetary aggregates have always figured prominently in St. Louis Fed data publications, data on other variables prominent in the monetary policy debates at the time, including short-term interest rates, excess reserves, and borrowings, were included in the data releases.

Early on, the various St. Louis Fed data publications incorporated "growth triangles," which tracked growth rates of monetary aggregates over varying horizons. Accompanying graphs of the aggregates included broken trend lines that illustrated rises and falls in growth rates. This information featured prominently in monetarist critiques of "stop-go" and procyclical characteristics of monetary policy during the Great Inflation period.

Does the tradition of data distribution initiated by Homer Jones remain a valuable public service? I certainly believe so. But I will also note that the St. Louis Fed's data resources are widely used within the Federal Reserve System. This information is required for Fed research and policy analysis; the extra cost of making the information available also to the general public is modest.

RATIONAL EXPECTATIONS MACROECONOMIC EQUILIBRIUM

The case for making data readily available is simple. Most macroeconomists today adhere to a model based on the idea of a rational expectations equilibrium. Policymakers are assumed to have a set of goals, a conception of how the economy works, and information about the current state and history of the economy. The private sector understands, to the extent possible, policymakers' views and has access to the same information about the state and history of the economy as policymakers have.

An equilibrium requires a situation in which (i) the private sector has a clear understanding of policy goals and the policymakers' model of the economy and (ii) the policy model of the economy is as accurate as economic science permits. Based on this understanding, market behavior depends centrally on expectations concerning monetary policy and the effects of monetary policy on the economy, including effects on inflation, employment, and financial stability. If the policymakers and private market participants do not have views that converge, no stable equilibrium is possible because expectations as to the behavior of others will be constantly changing.

The economy evolves in response to stochastic disturbances of all sorts. The continuous flow of new information includes everything that happens—weather disturbances, technological developments, routine economic data reports, and the like. The core of my policy model is that market responses and policy responses to new information are both maximizing—households maximize utility, firms maximize profits, and policymakers maximize their policy welfare function.

A critical assumption in this model is the symmetry of the information that is available to both policymakers and private market participants. In cases where the policymakers have an informational advantage over market participants, policy likely will not unfold in the way that markets expect, and the equilibrium that I have characterized here will not emerge. Hence, public access to current information on the economy at low cost is a prerequisite to good policy outcomes.

THE EVOLUTION OF ST. LOUIS FED DATA SERVICES

Data services provided by the Federal Reserve Bank of St. Louis have evolved significantly from the paper publications initiated by Homer Jones. The initial phase of this evolution began in April 1991 when FRED®, Federal Reserve Economic Data, was introduced as a dial-up electronic bulletin board. This service was not necessarily low cost. For users in the St. Louis area, access was available through a local phone call. For everyone else, long-distance phone charges were incurred. Nevertheless, within the first month of service, usage was recorded from places as wide ranging as Taipei, London, and Vancouver.[1] FRED was relatively small scale. The initial implementation included only the data published in *U.S. Financial Data* and a few other time series. Subsequently, it was expanded to include the data published in *Monetary Trends*, *National Economic Trends*, and *International Economic Trends*. At the end of 1995, the print versions of these four statistical publications contained short histories on approximately 200 national and international variables; initially FRED was of comparable scope.

The next step occurred in 1996 when FRED migrated to the World Wide Web. At that point, 403 national time series became available instantaneously to anyone who had a personal computer with a Web browser. An additional 70 series for the Eighth Federal Reserve District were also available. The data series were in text format and had to be copied and pasted into the user's PC. In July 2002, FRED became a true database and the user was offered a wider range of options. Data can be downloaded in either text or Excel format. Shortly thereafter, user accounts were introduced so that multiple data series can be downloaded into a single Excel workbook, and data lists can be stored for repeated downloads of updated information. In the first six months after this version of FRED was released, 3.8 million hits were recorded to the web site. In a recent six-month period, FRED received 21 million hits from over 109 countries around the world. FRED currently contains 1,175 national time series and 1,881 regional series. FRED data are updated on a real-time basis as information is released from various statistical agencies.

After 45 years, Homer Jones's modest initiative to distribute data on three variables has developed into a broad-based data resource on the U.S. economy that is available around the globe at the click of a mouse. Through this resource, researchers, students, market participants, and the general public can reach informed decisions based on information that is comparable to the information policymakers have.

In the past year, we have introduced a number of additional data services. One of these, ALFRED® (Archival Federal Reserve Economic Data), adds a vintage (or real-time) dimension to FRED. The ALFRED database stores revision histories of the FRED data series. Since 1996, we have maintained monthly or weekly archives of the FRED database. All the information in these archives has been populated to the ALFRED database, and the user can access point-in-time revisions of these data.[2] We have also extended the revision histories of many series back in time using data that were

[1] *Eighth Note* (1991, p. 1).

[2] We do not maintain histories of daily data series in ALFRED. Interest rates and exchange rates appear at daily frequencies in FRED. In principle, these data are not revised, though occasional recording errors do slip into the initial data releases. Such reporting errors are corrected in subsequent publications, so there is sometimes a vintage dimension to one of these series.

recorded in *U.S. Financial Data*, *Monetary Trends*, and *National Economic Trends*. For selected quarterly national income and product data, we have complete revision histories back to 1959 for real data and 1947 for nominal data. Revision histories are available on household and payroll employment data back to 1960. A similar history for industrial production is available back to 1927.

Preserving such information is crucial to understanding historical monetary policy. For example, Orphanides (2001, p. 964) shows "that real-time policy recommendations differ considerably from those obtained with ex-post revised data. Further, estimated policy reaction functions based on ex-post revised data provide misleading descriptions of historical policy and obscure the behavior suggested by information available to the Federal Reserve in real time." Orphanides concludes that "reliance on the information actually available to policy makers in real time is essential for the analysis of monetary policy rules."

Such vintage information also is essential for analysis of conditions at subnational levels. For example, in January 2005 the Bureau of Labor Statistics estimated that nonfarm employment in the St. Louis MSA had increased by 38.8 thousand between December 2003 and December 2004. This increase was widely cited as evidence that the MSA had returned to strong employment growth after four years of negative job growth. However, these data from the Current Employment Statistics were not benchmarked to more comprehensive labor market information that is available only with a lag.[3] The current estimate of nonfarm employment growth in the St. Louis MSA for this period, after several revisions, is only 11.6 thousand, less than 30 percent of the increase originally reported.

Another data initiative that we launched several years ago is FRASER®—the Federal Reserve Archival System for Economic Research. The objective of this initiative is to digitize and distribute the monetary and economic record of the U.S. economy. FRASER is a repository of image files of important historical documents and serial publications. At present we have posted the entire history of *The Economic Report of the President*, *Economic Indicators*, and *Business Conditions Digest*. We have also posted images of most issues of the *Survey of Current Business* from 1925 through 1990 and are working on filling in images of the remaining volumes. The collection also includes *Banking and Monetary Statistics* and the *Annual Statistical Digests* published by the Board of Governors, as well as the *Business Statistics* supplements to the *Survey of Current Business* published by the Department of Commerce. We are currently working, in a joint project with the Board of Governors, to create digital images of the entire history of the *Federal Reserve Bulletin*. Finally, we are posting images of historical statistical releases that we have collected in the process of extending the vintage histories in ALFRED back in time. These images should allow scholars, analysts, and students of economic history to reconstruct vintage data on many series in addition to those we are maintaining on ALFRED.

TRANSPARENCY, ACCOUNTABILITY, AND INFORMATION DISTRIBUTION

As just indicated, the scope of the archival information in FRASER extends beyond numeric data. Ready access to a wide variety of information is essential for transparency and accountability of monetary authorities and the public's full understanding of policy actions. Since 1994, the Federal Reserve System and the FOMC have improved the scope and timeliness of information releases. I have discussed this progress in previous speeches.[4] Currently, the FOMC releases a press statement at the conclusion of each scheduled meeting and three weeks later follows up with the release of minutes of the meeting. The press release and the minutes of the meetings record the vote on the policy action. The policy statement and minutes give the public a clear understanding of the action taken and insight into the rationale for the action.

Contrast the current situation with the one in 1979. At that time, actions by the Board of

[3] Wall and Wheeler (2005).

[4] See, for example, Poole (2005).

Governors on discount rate changes were reported promptly, but there was no press release subsequent to an FOMC policy action and FOMC meeting minutes were released with a 90-day delay. On September 19, 1979, the Board of Governors voted by the narrow margin of four to three to approve a ½-percentage-point increase in the discount rate, with all three dissents against the increase. This information generated the public perception that Fed officials were sharply divided and, therefore, that the Fed was not prepared to act decisively against inflation. John Berry (1979, p. A1), a knowledgeable reporter at the *Washington Post*, observed that "the split vote, with its clear signal that from the Fed's own point of view interest rates are at or close to their peak for this business cycle, might forestall any more increases in market interest rates." However, the interpretation of the "clear signal" was erroneous. On that same day, the FOMC had voted eight to four to raise the range for the intended funds rate to 11¼ to 11¾ percent. More importantly, three of the four dissents were in favor of a more forceful action to restrain inflation (see Lindsey, Orphanides, and Rasche, 2005, pp. 195-96). Neither the FOMC's action, the dissents, nor the rationale for the dissents were revealed to the public under the disclosure policies then in effect. The result was to destabilize markets, with commodity markets, in particular, exhibiting extreme volatility.

CONCLUSION

The tradition of data services was well established when I arrived in St. Louis in 1998, and I must say that I am proud that leadership in the Bank's Research Division has extended that tradition. Data are the lifeblood of empirical research in economics and of policy analysis. Our rational expectations conception of how the macroeconomy works requires that the markets and general public understand what the Fed is doing and why. Of all the things on which we spend money in the Federal Reserve, surely the return on our data services is among the highest.

REFERENCES

Berry, John. "Fed Lists Discount Rate to Peak of 11% on Close Vote." *Washington Post*, September 19, 1979, p. A1.

Business Week. "Maverick in the Fed System." November 18, 1967.

Eighth Note. "Introducing FRED." Federal Reserve Bank of St. Louis, May/June 1991, p. 1.

Orphanides, Athanasios. "Monetary Policy Rules Based on Real-Time Data." *American Economic Review*, September 2001, *91*(4), pp. 964.

Poole, William. "FOMC Transparency." Federal Reserve Bank of St. Louis *Review*, January/February 2005, *87*(1), pp. 1-9.

Sprinkel, Beryl W. "Confronting Monetary Policy Dilemmas: The Legacy of Homer Jones." Federal Reserve Bank of St. Louis *Review*, March 1987, *69*(3), p. 6.

Wall, Howard J. and Wheeler, Christopher H. "St. Louis Employment in 2004: A Tale of Two Surveys." CRE8 Occasional Report No. 2005-1, Federal Reserve Bank of St. Louis, February 2005.