Monetary Policy Normalization in the United States
Stephen D. Williamson

Common Fluctuations in OECD Budget Balances
Christopher J. Neely and David E. Rapach

Does the United States Lead Foreign Business Cycles?
Neville Francis, Michael T. Owyang, and Daniel Soques

How Does Informal Employment Affect the Design of Unemployment Insurance and Employment Protection?
Emilio Espino and Juan M. Sánchez
87
Monetary Policy Normalization in the United States
Stephen D. Williamson

109
Common Fluctuations in OECD Budget Balances
Christopher J. Neely and David E. Rapach

133
Does the United States Lead Foreign Business Cycles?
Neville Francis, Michael T. Owyang, and Daniel Soques

159
How Does Informal Employment Affect the Design of Unemployment Insurance and Employment Protection?
Emilio Espino and Juan M. Sánchez
Review
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, The 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
Articles may be reprinted, reproduced, republished, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. In these cases, there is no need to request written permission or approval. Please send a copy of any reprinted or republished materials to Review, Research Division of the Federal Reserve Bank of St. Louis, PO. Box 442, St. Louis, MO 63166-0442; STLS.Research.Publications@stls.frb.org.

Please note that any abstracts, synopses, translations, or other derivative work based on content published in Review may be made only with prior written permission of the Federal Reserve Bank of St. Louis. Please contact the Review editor at the above address to request this permission.

Economic Data
General economic data can be obtained through FRED® (Federal Reserve Economic Data), our free database with nearly a quarter of a million 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.

© 2015, Federal Reserve Bank of St. Louis.
ISSN 0014-9187
The Great Recession, which began in late 2007 and continued until mid-2009, demarcates some key changes in U.S. monetary policy. In 2015, the Federal Reserve's balance sheet is much larger than before the Great Recession. From December 2007 to October 2014, the Fed's balance sheet (either total assets or total liabilities, which are equal) increased more than fourfold. Since late 2008, the Fed's target interest rate, the federal funds rate, has been close to zero. This long period of a zero interest rate policy, or ZIRP, is unprecedented since the Treasury-Federal Reserve Accord of 1951 modernized the approach to central banking in the United States.

In its “Policy Normalization Principles and Plans” (Board of Governors, 2014), the Federal Open Market Committee (FOMC) proposed a program that should ultimately return the Fed’s balance sheet to a state similar to that of December 2007. This does not literally mean a return to a balance sheet of the same nominal size (in U.S. dollars) as before the financial crisis, but to a balance sheet that allows the Fed to implement monetary policy in the same way it did in December 2007, as discussed later in this article. In addition, the FOMC’s principles and plans outline a sequence of actions by which “liftoff”—departure from a ZIRP—would be achieved, followed by a reduction in the size of the Fed’s balance sheet.

The purpose of this article is first to take stock of the state of U.S. monetary policy and how that state was reached. This discussion is followed by an analysis of how the FOMC envisions normalization and the pitfalls that may arise on the road to normalization.
UNCONVENTIONAL MONETARY POLICY AND ITS LEGACY

There are two key aspects of the Fed’s monetary policy that should be considered: (i) the size and composition of the Fed’s balance sheet and (ii) the Fed’s interest rate policy. Each of these is analyzed in turn.

The Size and Composition of the Fed’s Balance Sheet

Figure 1 summarizes the time series of items on the assets and liabilities sides of the Fed’s balance sheet from 2007 to the present. Clearly, there has been a remarkable increase in the size of the balance sheet since the period prior to the financial crisis. (The critical period of the financial crisis period is defined here as the fall of 2008 through early 2009). For example, total assets held by the Fed increased from $891 billion in December 2007 to $4,498 billion in December 2014, a more than fourfold increase—or, alternatively, an increase from 6.1 percent of annual gross domestic product (GDP) to 25.3 percent of GDP.

While the size of the Fed’s balance sheet has increased greatly since the Great Recession began in December 2007, such an increase is by no means unprecedented, nor is it a world record. For example, at the end of September 2014, the Swiss National Bank, the Bank of Japan,
the People's Bank of China, and the National Bank of Denmark held assets that were, respec-
tively, 80.1 percent, 57.0 percent, 52.3 percent, and 26.3 percent of GDP, while the Fed held
assets that were only 25.3 percent of GDP. Of course, at that time, the Fed’s balance sheet was
larger than those of some central banks: Bank of England (22.5 percent of GDP), the European
Central Bank (20.2 percent), the Swedish Riksbank (12.1 percent), the Reserve Bank of Australia
(8.9 percent), and the Bank of Canada (4.6 percent).

It is important to note that much of the increase in the size of the Fed’s balance sheet has
occurred since the Great Recession ended in June 2009, when the Fed’s assets stood at $2,026
billion, or 14.1 percent of GDP. Further, the sources of the increases on the asset side of the
Fed’s balance sheet during the Great Recession were quite different from the sources after the
Great Recession. As shown in Figure 1, much of the increase in the Fed’s assets during the Great
Recession resulted from its lending programs—both conventional and unconventional—and
purchases of troubled assets and commercial paper. These interventions were responses to the
global financial crisis. However, the post-recession increases in the Fed’s assets were mainly
purchases of Treasury securities, mortgage-backed securities (MBS), and agency securities.
Further, a key element of the Fed’s post-Great Recession policy was a lengthening in the average
maturity of assets in the Fed’s portfolio. In December 2007, the Fed’s securities were entirely
Treasury securities, and of these 32.1 percent were Treasury bills (short-term government debt).
In December 2014, the Fed held no short-term Treasury securities; 41 percent of its security
holdings were long-maturity MBS and 58.1 percent of its security holdings were long-maturity
Treasury notes and bonds.

The primary liabilities on the Fed’s balance sheet before the Great Recession were currency
outstanding and reserves. Currency outstanding consists of the Federal Reserve notes and
coins held outside the banking system, while reserves are essentially the transaction accounts
of financial institutions held with the Fed, plus a small amount of currency held in bank vaults
and ATMs. In December 2007, $792 billion in currency outstanding financed most of the Fed’s
asset portfolio of $891 billion. Reserves were a relatively small amount, at $4.2 billion. From
December 2007 to December 2014, the stock of currency outstanding increased from $792
billion to $1,299 billion, or from 5.4 percent of annual GDP to 7.3 percent of GDP. Thus, there
has been a substantial increase in the demand for currency since the beginning of the Great
Recession, but that increase was far from sufficiently large to finance the expansion in the Fed’s
assets. From December 2007 to December 2014, the quantity of reserves outstanding increased
from 0.5 percent of total Fed liabilities to 52.9 percent. As well, reverse repurchase (repo) agree-
ments (an interest-bearing component of Fed liabilities) accounted for 11.3 percent of Fed
liabilities in December 2014.

**Interest Rate Policy**

Figure 2 shows the federal funds rate, the Fed’s policy target, from 1954 to the present.
Since its meeting December 16, 2008, the FOMC’s target federal funds rate has been in the
range of 0 to 0.25 percent. As Figure 2 shows, the federal funds rate was near zero at times
during the 1950s and early 1960s, but the period from December 2008 to the present—more
than six years of a ZIRP—is unprecedented in the United States.
A key change in the Fed’s interest rate policy was the switch to a regime that pays interest on reserves. With passage of the Financial Services Regulatory Relief Act of 2006, Congress authorized the Fed to pay interest on reserves effective October 1, 2011, but implementation was moved up to October 1, 2008, as part of the Fed’s response to the financial crisis.

The argument made by economists for paying interest on reserves appeals in part to economic efficiency. As first pointed out by Friedman (1969), when the nominal interest rate on safe assets is positive, economic agents will tend to hold too little money (in real terms) for transaction purposes. One solution to this economic inefficiency problem is for the government, or the central bank, to pay interest on all money balances—including holdings of currency and reserves, which together constitute the stock of outside money. Paying interest on currency is impractical: Currency is held by the public in widely dispersed locations, so payment of interest on this currency is extremely costly, if not infeasible, for the Fed. However, paying interest on reserves, which are account balances held by financial institutions with the Fed, is straightforward—as easy as the payment of interest by a private bank on a transaction account. If interest is not paid on reserves, there is effectively a tax on financial intermediation, because reserve requirements are still imposed in the United States on specific deposit liabilities. Thus, if reserves earn zero interest and banks are required to hold such reserves, this introduces a distortion and a loss in economic welfare.

A second rationale for interest on reserves (see, e.g., Goodfriend, 2002) is that if there is a sufficiently large stock of excess reserves outstanding overnight, then in a frictionless overnight financial market, the interest rate on reserves will determine the overnight interest rate. As an example of this mechanism, consider the following crude model of the federal funds market.
Figure 3
Federal Funds Market Model: Channel System

shown in Figure 3. In the figure, the nominal federal funds rate \( R^f \) is on the vertical axis, and the quantity of reserves \( Q \) is on the horizontal axis. The demand for reserves is given by the curve \( D \), which slopes downward due to a liquidity effect. Thus, to induce financial institutions to hold a higher stock of overnight reserves, the federal funds rate must be lower, as the federal funds rate is the opportunity cost of holding overnight reserves. But, the federal funds rate is bounded by the discount rate \( R_d \) and the interest rate on reserves (\( IOER \)). The upper bound is determined by the fact that financial institutions with reserve accounts can borrow from the Fed at the discount rate, which is sometimes treated as a separate policy instrument by the Fed; these institutions would not borrow on the federal funds market at a rate higher than that rate. Similarly, the lower bound is determined by the ability of financial institutions to lend to the Fed at the IOER; these institutions would not wish to lend on the federal funds market at a lower rate.

Note that “IOER” typically is used to denote “interest on excess reserves,” where excess reserves are the quantity of reserves held in excess of a bank’s reserve requirement. U.S. law specifies that the interest rates on required reserves and excess reserves could be different, but for practical purposes the Fed has been setting the interest rates on both required reserves and excess reserves at 0.25 percent.

Before the financial crisis, when financial institutions in the United States were not paid interest on reserves and the quantity of reserves held in excess of reserve requirements was tiny, the federal funds market worked as captured in Figure 3, with \( IOER = 0 \). The position of the demand curve \( D \) on any given day was determined by activity in the financial payments system. For example, large shocks can be associated with the overnight lending and borrowing activities of large financial institutions, the day of the week, the month in the year, or the day...
within a particular reserve-averaging period (the period over which average reserves held for a bank must exceed the average required reserves on average deposit liabilities). Some of these large shocks are predictable, whereas others are not. Thus, on a given day in the pre-financial crisis period, the New York Fed would need to predict the position of the demand curve in Figure 3. The supply curve $S$ is vertical, determined by the intervention of the New York Fed in the overnight financial market. Typically, the New York Fed would determine the quantity of reserves outstanding overnight by varying the quantity of repo agreements (collateralized lending by the Fed to financial institutions). Standard open market operations—purchases and sales of Treasury securities by the Fed—determined the medium- and long-run quantity of reserves in the system. In terms of Figure 3, $R^T$ is the Fed’s target fed funds rate, and the goal of the Fed on a given day was to intervene so that the supply curve $S$ intersects the demand curve $D$ at the target interest rate.

In the planning stages for payment of interest on reserves in the United States, it was thought that, with a sufficiently large stock of interest-bearing reserves outstanding, the federal funds market would work as in Figure 4. That is, if the supply curve $S$ lies sufficiently far to the right, the IOER should determine the federal funds rate. Such a system of central bank intervention is called a **floor system** (as the floor on the overnight rate, the IOER, determines the overnight rate), while the system depicted in Figure 3 is a **channel system (or corridor system)** under which the overnight rate falls in a channel defined by the central bank’s lending rate (the upper bound on the channel) and IOER (the lower bound on the channel). In principle, a floor system should greatly simplify interest rate targeting for the Fed. As an intermediate step to hitting a given target for the federal funds rate, instead of having to forecast the daily demand for reserves, the New York Fed could simply ensure that a sufficiently large quantity...
of reserves was outstanding each night. Then the administered IOER, set at the target federal funds rate, should determine the federal funds rate.

Economists such as Goodfriend (2002) had good reasons to think that the payment of interest on reserves in the United States would lead to the potential for a smoothly operating floor system of central bank intervention to target overnight interest rates. Indeed, some floor systems have worked well in practice, essentially as depicted in Figure 4. For example, the Bank of Canada operated a floor system from April 2009 until May 2010. The Bank of Canada usually intervenes through a channel system, whereby the target overnight interest rate is bounded by the rate at which the Bank of Canada lends and the interest rate the Bank pays on reserves, much as depicted in Figure 3. Standard operating procedure is for the Bank to set its lending rate (the “Bank Rate”) at 0.25 percent above the target overnight rate and the interest rate on reserves at 0.25 percent below the target overnight rate. But from April 2009 to May 2010, the overnight rate and the interest rate on reserves were both at 0.25 percent, with the Bank Rate at 0.50 percent. As shown in Figure 5, the Bank accomplished its goal by ensuring that about $3 billion (Canadian dollars) in reserves was outstanding each night. Note in the figure that the Bank of Canada successfully pegged the overnight rate at the interest rate on reserves (0.25 percent) during this period.

In contrast to the 2009-10 Canadian experience, the floor system in place in the United States since late 2008 has not been successful, if success is measured by the ability of the central bank to hit a particular overnight target rate. Despite an extremely large quantity of reserves outstanding overnight since the financial crisis, the federal funds rate has not been pegged by the IOER (Figure 6). The difference between the IOER and the federal funds rate has been variable and surprisingly large—sometimes as much as 20 basis points. This explains why the
FOMC has chosen to state its target federal funds rate as a range of 0 to 0.25 percent over this period rather than quoting a target rate.

**HOW DID WE GET HERE? THE RATIONALE FOR A ZIRP AND QUANTITATIVE EASING**

Understanding the issues involved in policy normalization requires delving into the reasons for the Fed’s unconventional steps of expanding its balance sheet and lowering its interest rate target to near zero for such a long period. As noted in the first section, much of the balance-sheet increase occurred after the Great Recession, so these unusual interventions were associated not just with the financial crisis but with its aftermath as well.

**ZIRP and Forward Guidance**

The Fed has kept its funds rate target near zero for more than six years. Why has the ZIRP lasted so long? At least part of the rationale for a ZIRP came from New Keynesian (NK) macroeconomic theory, popularized by Woodford (2003) and others. The argument is articulated in more technical terms in Werning (2011). Though the members of the FOMC have diverse views on how (or perhaps whether) to model the macroeconomy, their actions appear to be consistent with the basic NK model. Here I sidestep the issue of whether NK theory is a useful guide for monetary policy, a matter subject to some debate.

In a basic NK model, sticky prices and wages cause an inefficiency that monetary policy can correct. Figure 7 provides a simple representation of what happens in the NK model. The real interest rate $R - \pi$ is on the vertical axis; that is, the nominal interest rate $R$ on short-term government debt minus the anticipated inflation rate $\pi$. $Y$, which denotes real GDP, is on the
horizontal axis. The curve IS is the NK IS curve, which certainly looks like the familiar IS curve from traditional Keynesian analysis but is constructed differently (see, e.g., Clarida, Gali, and Gertler, 1999). Basically, the IS curve denotes a set of combinations of real interest rates and aggregate real income for which the market for government debt is in equilibrium. The supply of government debt outstanding is determined by the fiscal authority—the Treasury in the United States—and the demand for government debt is increasing in the real interest rate on that debt, increasing in current aggregate real income $Y$, and decreasing in future aggregate real income $Y_f$. These effects arise because (i) the demand for an asset increases with its rate of return and (ii) assets are held to smooth consumption over time. Thus, if current income is higher, everything else held constant, economic agents will want to smooth consumption by purchasing assets, including government debt, to allow more consumption in the future (as well as more in the present). If future income is expected to be higher, economic agents will want to sell assets to allow more consumption in the present (as well as more in the future).

In this simplified NK model, monetary policy works as follows. In Figure 7, the quantity $r^*$ denotes the “natural real rate of interest,” which is what the real interest rate would be in equilibrium if all prices and wages were flexible. To simplify things, I assume that $Y'$ and $\pi$ are exogenous in the short run. Monetary policy sets the nominal interest rate at $R_0$ in Figure 7; I also suppose that the economy is initially in long-run equilibrium with the actual real interest rate $R_0 - \pi_0$ equal to $r^*$, the natural real rate of interest.

Suppose that the central bank lowers the nominal interest rate to $R_1$. Given the exogenous anticipated rate of inflation $\pi_0$, the real interest rate then falls to $R_1 - \pi_0$ (see Figure 7), which implies that real income must rise from $Y_0$ to $Y_1$ equilibrate to the market for government debt. Therefore, there is a short-run non-neutrality of money in that monetary policy easing—a
reduction in the nominal interest rate—increases aggregate output. In the long run, the natural real rate of interest must be the actual rate of interest, so if the nominal interest rate stays at $R_1$ forever, then the anticipated inflation rate must fall to $\pi_1$, with $R_1 - \pi_1 = r^*$. This is just the long-run Fisher effect: Monetary policy is neutral in the long run and has no long-run effect on real output or the real interest rate.

In this simplification of the basic NK framework, one way to represent the effects of the financial crisis is as a shift in the $IS$ curve. This is not a “demand shock” in the sense of an old-style Keynesian model, but an increase in the demand for government debt. Such an effect could arise for at least two reasons. First, credit constraints could become tighter for households, causing them to value government debt more as an asset to self-insure against shocks that would otherwise cause them to borrow. Second, an effect of the financial crisis was to effectively destroy part of the stock of assets used as collateral in financial markets—principally, asset-backed securities that were no longer viewed as safe collateral. This reduction in assets would tend to increase the demand for government debt used as collateral.

In Figure 8, the $IS$ curve shifts to the left from $IS_0$ to $IS_1$ as the result of a financial crisis shock. I also assume that the financial crisis shock is persistent, which implies that the natural real rate of interest also falls from $r_0^*$ to $r_1^*$. Initially the nominal interest rate is $R_0$ and the anticipated inflation rate is $\pi_0$, with the real interest rate equal to $R_0 - \pi_0 = r_0^*$. When the shock occurs, if the central bank keeps the nominal interest rate at $R_0$, then aggregate output will fall from $Y_0$ to $Y_1$. The goal of monetary policy in this model is essentially to remove inefficiency in the market for government debt. To remove inefficiency in these circumstances, the central bank would like to lower the nominal interest rate in the short run, with the goal of lowering the real interest rate to the new natural real rate of interest $r_1^*$.\[\text{Figure 8} \quad \text{A Financial Crisis Shock in the Simplified New Keynesian Model}\]
However, the nominal interest rate cannot go below zero—that is, there is a zero-lower-bound problem. If the financial crisis shock is a very large one, then a situation as in Figure 8 could develop in which the nominal interest rate is lowered to zero by the central bank, so the actual real rate of interest is \( -\pi_t \). But in Figure 8 \( -\pi_t > r^*_1 \), so the economic inefficiency cannot be removed by the central bank, given the zero lower bound on the nominal interest rate. Even at the zero lower bound on the nominal interest rate, the real rate of interest is still inefficiently high.

In NK models, the central bank is not powerless at the zero lower bound, however, as it can resort to forward guidance. That is, the central bank can make promises about future monetary policy, as outlined in Werning (2011) or Woodford (2012). For example, the central bank could promise higher future inflation and thus lower the real interest rate, which is \( -\pi \), given the central bank’s zero-lower-bound policy. If the central bank is convincing enough, it may be possible to achieve an equilibrium in which the anticipated inflation rate is \( \pi_t = -r^*_1 \), so the central bank can completely remove the economic inefficiency through forward guidance and achieve a level of aggregate output \( Y_3 \) in Figure 8.

Though simplified, this captures the basic NK theory behind forward guidance, which in principle rationalizes the forward guidance policies of the Fed following the financial crisis. In the FOMC statement, which follows regular FOMC meetings, the Committee at first expressed its forward guidance in terms of the future of the ZIRP, and this forward guidance changed over time. In January 2009, the Committee stated that the ZIRP would continue for “some time.” Then on March 18, this language changed to “extended period.” On August 9, 2011, this became “at least through mid-2013,” followed on January 25, 2012, by “at least through late 2014” and on September 13, 2012, by “at least through mid-2015.” In the statement on December 12, 2012, the forward guidance language became considerably longer and more specific:

[The Committee] currently anticipates that this exceptionally low range for the federal funds rate will be appropriate at least as long as the unemployment rate remains above 6½ percent, inflation between one and two years ahead is projected to be no more than a half percentage point above the Committee’s 2 percent longer-run goal, and longer-term inflation expectations continue to be well anchored. The Committee views these thresholds as consistent with its earlier date-based guidance. In determining how long to maintain a highly accommodative stance of monetary policy, the Committee will also consider other information, including additional measures of labor market conditions, indicators of inflation pressures and inflation expectations, and readings on financial developments. When the Committee decides to begin to remove policy accommodation, it will take a balanced approach consistent with its longer-run goals of maximum employment and inflation of 2 percent.

This language seems to be at least broadly consistent with what Woodford (2012) appeared to have in mind, and so in principle is in line with a clearly articulated and consistent economic theory. The forward guidance language in the quoted material expresses what the Committee cares about—the unemployment rate and the inflation rate—and gives specific numerical thresholds. However, the statement leaves some questions unanswered: for example, (i) What
happens when the unemployment rate falls below 6.5 percent? and (ii) What happens if the forecast for the inflation rate is below 2 percent?

On March 19, 2014, the Committee’s forward guidance language changed again, to

In determining how long to maintain the current 0 to ¼ percent target range for the federal funds rate, the Committee will assess progress—both realized and expected—toward its objectives of maximum employment and 2 percent inflation. This assessment will take into account a wide range of information, including measures of labor market conditions, indicators of inflation pressures and inflation expectations, and readings on financial developments. The Committee continues to anticipate, based on its assessment of these factors, that it likely will be appropriate to maintain the current target range for the federal funds rate for a considerable time after the asset purchase program ends, especially if projected inflation continues to run below the Committee’s 2 percent longer-run goal, and provided that longer-term inflation expectations remain well anchored.

When the Committee decides to begin to remove policy accommodation, it will take a balanced approach consistent with its longer-run goals of maximum employment and inflation of 2 percent. The Committee currently anticipates that, even after employment and inflation are near mandate-consistent levels, economic conditions may, for some time, warrant keeping the target federal funds rate below levels the Committee views as normal in the longer run.

With the unemployment rate nearing 6½ percent, the Committee has updated its forward guidance. The change in the Committee’s guidance does not indicate any change in the Committee’s policy intentions as set forth in its recent statements. This statement contains more information about the Committee’s intentions beyond the ZIRP. For example, the language concerning “keeping the target federal funds rate below levels” indicates that the Committee expects that policy may continue to be more accommodative than usual, even after its target interest rate starts to rise.

Therefore, one explanation for the long duration of the ZIRP policy is that the extended time at the zero lower bound was a necessary part of meeting earlier commitments made by the Fed. In line with Werning (2011) and Woodford (2012), it could be argued that to mitigate the large inefficiency immediately after the Great Recession, it was important for the Fed to promise to stay with the ZIRP for a long time. If that was the idea, then one could criticize the FOMC for being vague and inconsistent in its forward guidance. Explicit commitment is important in the Werning (2011) and Woodford (2012) frameworks. With regard to explicitness, the FOMC’s forward guidance used language such as “some time,” “extended period,” and “considerable time”—phrases open to interpretation. On commitment, clearly changes in the calendar dates for the forward guidance and the failure of 6½ percent unemployment to trigger any policy action indicate, if anything, a lack of commitment. Thus, on explicitness and commitment, the Fed’s forward guidance appears to have left something to be desired.

The Size of the Fed’s Balance Sheet: Quantitative Easing

The post-Great Recession increase in the size of the Fed’s balance sheet was driven by asset purchases—quantitative easing (QE). There were four different components to QE after the Great Recession: (i) purchases of long-maturity Treasury debt, (ii) purchases of MBS and
agency securities, (iii) increases in the average maturity of the Fed’s security holdings through exchanges of short-maturity Treasury debt for long-maturity Treasury debt, and (iv) reinvestment—that is, purchases of long-maturity Treasury debt and MBS as those securities mature to hold constant the size of the balance sheet.

When the Great Recession ended in June 2009, the Fed was in the midst of purchasing a substantial quantity of MBS and agency securities, and these purchases continued well into 2010. This program was part of what is sometimes called QE1. Another element of QE—reinvestment of maturing securities—began in August 2010 and continues today. From November 2010 until June 2011, the Fed purchased a stock of about $600 billion in long-maturity Treasury securities in an asset purchase program sometimes called QE2. Then from September 2011 until December 2012, the Fed engaged in “Operation Twist,” which involved swaps by the Fed of short-maturity Treasury securities for long-maturity Treasury securities.1 Finally, purchases of MBS resumed in September 2012: From January 2013 to October 2014, the Fed purchased a large quantity of MBS and long-maturity Treasury securities in what is sometimes called QE3.

What is the rationale for QE? In principle, QE is another unconventional monetary policy instrument—like forward guidance—that allows the central bank to ease monetary conditions when a ZIRP is in place and the short-term nominal interest rate cannot go any lower. But how does it work, exactly? To address this question one needs to think about the outstanding debt of the consolidated government, which is defined as including the Treasury and the Fed. Thus, the consolidated government debt for the United States includes currency and reserves (Fed liabilities) plus debt issued by the U.S. Treasury and held by the public (not including any Treasury debt held by the Fed).

When the nominal interest rate is zero, there is a liquidity trap in the sense that reserves and short-term government debt are roughly identical assets. So, if the central bank swaps reserves for short-term government debt, this should not matter at all—the central bank’s intervention will be neutral, affecting no prices or quantities of consequence in the economy. But QE is not a swap of reserves for short-term government debt. For example, the most straightforward QE intervention by the central bank is a Fed purchase of long-maturity government debt with reserves.

Even if the short-term nominal interest rate is zero, long bond yields can be positive, as they have been in the United States in the post-Great Recession period. This reflects the fact that long-maturity Treasury securities and reserves are different in ways that short-maturity Treasuries and reserves are not at the zero lower bound. This might mean that QE will not be neutral at the zero lower bound—possibly, purchases of long-maturity Treasuries by the Fed will lower long-term bond yields.

To the extent that QE is justified by any existing economic theory, policymakers typically appeal to segmented markets financial theory (see, e.g., Bernanke, 2012). Some segmented markets theories of the term structure of interest rates fall under the umbrella of “preferred habitat theory” (see Vayanos and Vila, 2009). Preferred habitat theory posits that different financial market participants have different preferences for assets of particular maturities. For example, life insurance companies have long-maturity liabilities and these institutions might prefer long-maturity assets to manage maturity risk. Preferred habitat preferences tend to seg-
ment the market for assets by maturity, according to the theory. Alternatively, policymakers sometimes articulate a segmented asset markets theory of QE in terms of “portfolio balance” models (see, e.g., Tobin, 1969). The idea here is a close cousin of preferred habitat theory, as portfolio balance models specify demand and supply functions for assets as functions of the rates of return and wealth. Then, the critical parameters in such a theory are those determining whether, and to what degree, assets are substitutes or complements.

Whether economists are appealing to preferred habitat or portfolio balance models to rationalize QE, the mechanism by which QE is imagined to work is similar. According to the story, if the Fed purchases long-maturity assets, arbitrage will not be complete, and this will tend to increase the prices of long-maturity assets and reduce their yields. Then, lower long-term bond yields might act to increase investment, consumption, and output through a conventional Keynesian mechanism. Policymakers also envision an effect of QE on inflation, though this mechanism is perhaps less clear.

What is notable is an absence of modern economic theory that provides a role for QE. Indeed, some monetary and financial theory shows how particular central bank asset purchases can have no effect at all. For example, Wallace (1981) provides a general theorem for the neutrality of open market operations. Wallace’s theorem is akin to the Modigliani-Miller theorem in corporate finance, which gives conditions under which the financial structure of a firm can be irrelevant. In Wallace’s theorem, under certain conditions the financial structure of the central bank is irrelevant for economic activity, asset prices, and goods prices.

There may be strong reasons to believe that theorems such as Wallace’s are not relevant for thinking about reality, as the conditions required for the theorems to hold are far too stringent. However, the case can be made that monetary policy matters—and can potentially increase economic welfare—because the central bank has some special powers that private financial intermediaries do not. For example, the Fed has a monopoly on issuing currency, and many of the nation’s daytime payments involve the transfer of Fed liabilities—reserves—among financial institutions. But QE may not be an activity for which the Fed has a special advantage. For example, turning long-maturity Treasury securities into overnight assets (reserves) is very close to the private sector activity of turning Treasury securities into overnight repos. If the Fed engages in QE, why would there not be an equal and offsetting effect in the private sector, yielding no net change in any aggregate economic variables?

In some widely used macroeconomic models QE is irrelevant. For example, in basic NK models (as Woodford, 2012, notes) there are complete markets and Ricardian equivalence. Financial arbitrage is perfect, and the quantity and composition of government debt are irrelevant. Further, while some macroeconomists (e.g., Gertler and Karadi, 2013) have studied the effects of QE, they do so without explicitly integrating a central bank and monetary policy into their models. Thus, there is no broader theory integrating mainstream NK theory—which motivates forward guidance—with a theory of QE.

In previous work (Williamson, 2014a), I develop an alternative theory of QE. In the theory, collateral is important in the banking sector for supporting financial intermediation activity, and collateral can be in short supply because the fiscal authority will not expand the quantity of government debt, even though that might be appropriate. Binding collateral constraints then tend to lower the real interest rate on government debt. At the zero lower bound on the
nominal interest rate, this tends to make the inflation rate high, given the Fisher effect. In the model, QE increases the effective stock of eligible collateral, as QE is a swap of short-maturity assets for long-maturity assets, and short-maturity assets are better collateral. Therefore, QE relaxes collateral constraints, decreases nominal long bond yields, increases the short-term real interest rate, and decreases the inflation rate. The latter two effects (the real interest rate increase and the inflation rate decrease) are certainly not what central bankers typically imagine will happen when they resort to QE interventions, though these effects are consistent in the model with an increase in economic welfare.

In general, one can conclude there should be little confidence in our knowledge of the quantitative effects of QE, let alone the qualitative effects. The existing theories are not well developed and have not been confronted by the empirical evidence in ways that shed light on how and to what extent QE works. There is some evidence that QE has important announcement effects on asset prices (see Neely, 2015, and Williamson, 2014b), though such evidence is consistent with a signaling role for QE. That is, QE may matter only to the extent that it acts as forward guidance—signaling the future path of market interest rates. There is little solid evidence about the effects of QE on economic quantities of interest—for example, employment, unemployment, GDP, and inflation.

**POLICY NORMALIZATION**

The FOMC announced its “Policy Normalization Principles and Plans” on September 17, 2014 (Board of Governors, 2014). There are two key aspects to the Fed’s normalization plans. First, the FOMC will raise its overnight interest rate target when it deems there is no longer as great a need for monetary accommodation. Second, the FOMC wishes to ultimately reduce the Fed’s balance sheet to a size commensurate with operating a channel system—that is, to a range such that the quantity of reserves outstanding will be on the order of pre-financial crisis levels. However, this balance-sheet reduction will occur slowly, with no plans by the FOMC to sell assets, though that option has not been completely ruled out. Primarily, the Fed’s assets will decline as Treasury securities and MBS mature. Treasury securities mature in a predictable manner, but the rate at which MBS decline is uncertain, as this depends in part on refinancing and default on the private mortgages backing the MBS.

QE ended in October 2014, so the next steps by the Fed toward normalizing monetary policy are (i) taking actions to increase short-term nominal interest rates—deciding on “liftoff” from the zero lower bound—and (ii) reducing the size of the balance sheet. The FOMC has made clear that liftoff will occur before any steps are taken toward balance-sheet reduction, so liftoff will happen in the context of a large stock of excess reserves in the financial system. This timing of policy actions (liftoff before balance-sheet reduction) is perhaps puzzling. Indeed, since the Fed first reduced its interest rate target and then resorted to balance-sheet-expanding asset purchases, one might think these actions would occur in reverse as the Fed exits its accommodative stance, with balance-sheet reduction occurring before liftoff. In any case, no rationale appears to have been provided by the FOMC for the sequence of actions laid out in its normalization plans.
As indicated earlier, the U.S. financial system presents some unique problems for central bank interest rate control under a floor system. While the IOER should effectively determine the federal funds rate, it has already been shown that this has not happened (see Figure 6). That is, the margin between the IOER and the federal funds rate is significant—and variable. Why is this? Under the Financial Services Regulatory Relief Act of 2006, the Fed was not authorized to pay interest on the reserve balances of government-sponsored enterprises (GSEs), including Fannie Mae, Freddie Mac, and the Federal Home Loan Banks (FHLBs). In principle, this should not cause a problem for interest rate control, as the GSEs could earn overnight interest by lending on the federal funds market. That is, a financial institution with a reserve account earning interest on reserves at 0.25 percent could borrow on the federal funds market overnight from a GSE and hold the funds overnight as reserves. In the absence of risk or costs of lending and borrowing, and under perfect competition, the interest rate on such a federal funds loan should be 0.25 percent. But since the federal funds rate has been substantially below IOER, there must be some friction or frictions that inhibit arbitrage in the federal funds market.

At least two factors potentially explain the margin between the IOER and the federal funds rate (see Martin et al., 2013). The first is balance-sheet costs for the financial institutions that can earn interest on reserves. These balance-sheet costs currently arise for two reasons. First, under the Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010, the rules for deposit insurance premiums paid by banks to the Federal Deposit Insurance Corporation changed in February 2011. Formerly, a bank’s deposit insurance premium depended only on domestic deposits, but now the premium depends instead on total assets. This means that if a bank borrows on the federal funds market overnight and invests the proceeds in its reserve account with the Fed, its total deposit insurance premium will be higher. This regulatory change was instituted as a means to correct incentive problems for larger banks, which depend more on nondeposit funding than do small banks. But the change has the side effect of increasing a bank’s balance-sheet costs when it borrows on the federal funds market and holds the proceeds as reserves.

Second, given how capital requirements are imposed on banks, if a bank holds more reserves over a long period, its capital requirement will increase. That is, some of the increase in reserves must be financed by an increase in capital and cannot all be financed by borrowing on the federal funds market. To the extent that capital is a more costly source of funding than borrowing on the federal funds market, capital requirements then add another balance-sheet cost—over and above the cost of additional deposit insurance premiums—to the cost of holding interest-bearing reserves for a financial institution that cannot receive interest on reserves.

Finally, in addition to balance-sheet costs, it is possible that imperfect competition could explain part of the margin between IOER and the federal funds rate. For example, because GSEs have few counterparties in the federal funds market, it may be the case that these counterparties extract some rent in federal funds transactions from the GSEs. However, even if the GSEs have few counterparties, it is not clear why the threat of entry from other counterparties would not enforce competitive pricing. Thus, balance-sheet costs seem the more likely source...
of the interest rate margin, though we do not have a good explanation for why the interest rate margin in Figure 6 is so variable.

As explained in Afonso, Entz, and LeSueur (2013a,b), because most financial institutions that can hold reserve accounts in the United States are flush with reserves, the majority of lending and borrowing on the federal funds market consists of lending by FHLBs to the branches of foreign banks in the United States. Thus, activity on the federal funds market is much different than it was before the financial crisis—it is now mostly an attempt by GSEs to earn interest on overnight balances.

Given our imperfect understanding of the determinants of the margin between the IOER and the federal funds rate, it is uncertain what will happen to that margin as the IOER rises. And given that the FOMC wishes to retain its focus on the federal funds rate as the relevant target overnight rate, it could be problematic if the IOER/federal funds rate margin increases or decreases as IOER rises. A potential solution to this problem involves use of the overnight reverse repurchase agreement (ON RRP) facility at the Fed. A reverse repo is a loan to the Fed with Fed assets posted as collateral. An ON RRP is thus another interest-bearing Fed liability—effectively reserves by another name. The ON RRP facility at the Fed has an expanded list of counterparties, including GSEs and money market mutual funds (MMMFs). First, the GSEs can lend directly to the Fed by way of the ON RRP facility as a means of earning overnight interest. Second, institutions such as the MMMFs, which cannot hold reserves, now have access to interest-bearing Fed liabilities.

As shown in Figure 9, the ON RRP facility was in use prior to the financial crisis, but it was not used as a policy tool. Daily intervention to target the federal funds rate typically consisted of varying the quantity of repos (collateralized loans by the Fed) rather than ON RRPs.
In the figure, the increased recent activity at the ON RRP facility is due to experimentation by the New York Fed in an attempt to understand the mechanics of the market before liftoff occurs. The ON RRP facility should, in principle, set a floor under the federal funds rate. If the Fed fixes an ON RRP interest rate and then borrows whatever quantity is forthcoming from its counterparties at that rate (a procedure called “fixed rate, full allotment”), then no market participant should wish to lend on the federal funds market at a rate below the ON RRP rate. But two key questions then arise: (i) What should the ON RRP rate be relative to the IOER? (ii) What interest rate should the Fed treat as its overnight target? These two questions are closely related.

Clearly, an ON RRP rate greater than the IOER makes no sense, as the Fed should not want to pay a higher interest rate on its overnight liabilities than the rate it pays to financial institutions with interest-bearing reserve accounts. But suppose that the Fed sets the ON RRP rate equal to the IOER. Since MMMFs do not have the balance-sheet costs of commercial banks (no deposit insurance and no capital requirements), MMMFs should be able to attract depositors away from commercial banks and cause a migration of Fed liabilities from reserves to ON RRPs held by MMMFs. Ultimately, reserves would shrink to a small amount—on the order of what was outstanding before the financial crisis, and ON RRPs would account for most of the Fed’s outstanding interest-bearing liabilities. Relative to a regime without the ON RRP facility, this would (i) reduce the Fed’s interest costs of a given quantity of outstanding interest-bearing liabilities and (ii) peg the overnight rate to the IOER. However, federal funds market activity would dwindle to a very small amount and, given the riskiness of unsecured federal funds lending, the federal funds rate would be greater than the IOER. Under these circumstances, the Fed could focus on the ON RRP rate or the IOER (the same thing in this case) as the relevant policy rate and ignore the federal funds rate.

Alternatively, the FOMC could set the ON RRP rate below the IOER. The wider the margin between the IOER and the ON RRP rate, the smaller would be the quantity of ON RRPs outstanding relative to reserves and the larger would be the quantity of federal funds market activity (because of lending by GSEs to banks earning interest on reserves). If under these circumstances the Fed were to focus on the federal funds rate as the FOMC’s policy rate, then this procedure should bound the federal funds rate between the ON RRP rate and the IOER. Where the federal funds rate would reside in that corridor is unpredictable, in part because future changes in regulation could affect the balance-sheet costs of financial institutions borrowing on the federal funds market.

There are other potential solutions to the problem of setting a floor under the federal funds rate. For example, Congress could authorize the Fed to pay interest on the reserve balances of the GSEs. Another possibility is that banks could be permitted to hold segregated accounts with the Fed, which are not subject to balance-sheet costs and would therefore allow more efficient arbitrage and a federal funds rate closer to the IOER. However, neither of these alternatives appears to be under consideration.

Reducing the Size of the Balance Sheet

In “Policy Normalization Principles and Plans” (Board of Governors, 2014), the FOMC indicates that it plans to reduce the size of the balance sheet slowly, primarily by not reinvest-
ing as securities mature. There are no plans for outright sales of assets in the Fed’s portfolio. Further, the reinvestment program will not end until after liftoff occurs.

After reinvestment ends, how long will it take to shrink the balance sheet to a size commensurate with a low level of reserves—in line with the level outstanding prior to the financial crisis? The assets in the Fed’s portfolio are financed primarily by currency outstanding, reserves, and now ON RRPs as well. So, as government debt and MBS mature and assets fall, this must be met by a reduction in the quantity of one of the primary liabilities—currency, reserves, or ON RRPs. One can think of the quantity of currency outstanding as being driven by demand, so as Fed assets mature, reserves and ON RRPs must fall. Note also that reserves and ON RRPs will fall as the demand for currency rises, everything else held constant. Therefore, the speed of normalization for the balance sheet depends mainly on the rate at which the Fed’s assets mature and the rate of increase in currency demand over time.

Carpenter et al. (2013) provide projections that estimate 2020 or so as when the Fed balance sheet normalizes—that is, the date at which reserves are at a level similar to what was outstanding before the financial crisis. Their projections were made in early 2013 under the assumptions that QE3 would end in June 2014, that liftoff would occur in the third quarter of 2015, and that the reinvestment program would end prior to liftoff. If liftoff occurs as they projected, then the facts that QE3 continued until October 2014 and that the Board of Governors (2014) normalization plan includes a plan to stop reinvestment after liftoff occurs imply that the date of balance-sheet normalization would be later than 2020. This is because the process of balance sheet-normalization is expected to start later and with a larger balance sheet than what Carpenter et al. (2013) assumed.

The projections of Carpenter et al. (2013) also assume there are no future reinvestment or QE programs on the path to normalization. There is no guarantee of that. Much depends on how soon the Fed achieves liftoff, how rapidly market interest rates rise after liftoff, and the size and timing of future shocks to the economy. It is conceivable that liftoff could be postponed or proceed very slowly, increasing the chance of a large negative macroeconomic shock to the U.S. economy before short-term market interest rates reach normal levels. Thus, the Fed could possibly find itself following a ZIRP once again, with the real economy performing poorly. Then, more QE could be a possibility.

However, suppose liftoff occurs sooner rather than later and short-term interest rates increase at a moderate pace, with no large unanticipated negative macroeconomic shocks. What then are the implications of a large balance sheet, other than the differences in the Fed’s operating procedure for targeting interest rates as already discussed? The Fed’s balance sheet is currently structured with a level of maturity mismatch that is far different from before the financial crisis, as already noted. The Fed currently borrows short—primarily overnight—and lends long. At the end of 2012, the average maturity of the Fed’s asset portfolio was calculated as 10.4 years (see Bukhari et al., 2013) and has increased since then. If the Fed were a private financial institution, this mismatch would put it in a precarious position. Maturity mismatch is risky for private financial intermediaries because, if short-term interest rates rise, the financial institution’s profit margin can disappear and default can ensue.
But the Fed is not a private financial institution. Essentially, it cannot default on its liabilities, except for ON RRP (which are identical to the repo agreements traded by private repo market participants). Currency and reserves are not promises to deliver anything in the future, so there are no promises concerning these liabilities that the Fed could fail to meet.

What about the Fed’s profits? Typically, the Fed earns interest on the securities in its portfolio; it pays the interest on its liabilities (except for currency, which of course does not bear interest); it pays its costs, including wages to Fed employees; and it remits the remainder to the U.S. Treasury. So, if short-term interest rates rise, then Fed profits will be lower, and it is feasible for short-term interest rates to rise sufficiently that remittances to the Treasury fall to zero.

But is this scenario likely to happen over the normalization period? The projections of Carpenter et al. (2013) suggest not. According to their estimates, remittances to the Treasury by the Fed will fall, at worst, to pre-financial crisis levels over the normalization period. Even if remittances did fall to zero, in what sense would that matter? A period with zero remittances to the Treasury by the Fed is one in which the Fed has effectively increased the interest payments on the consolidated government debt. That is, by purchasing long-maturity government debt with short-term liabilities and suffering a loss, the Fed has created a situation in which the consolidated government debt has shorter maturity than it otherwise would (if the Fed had not made the asset purchases) and the consolidated government is making larger interest payments than if the long-term Treasury debt had not been purchased by the Fed. But QE has also reduced interest payments on the consolidated government debt up to the present, and that is projected to continue for some period. Indeed, the Fed has been making record profits since the onset of the financial crisis in 2008. Further, Carpenter et al. (2013) argue that the Fed’s QE programs will ultimately produce a larger cumulative flow of remittances to the Treasury than if QE had not happened.

Perhaps one should not be so sanguine, though. Goodfriend (2014) argues that the expansion of the Fed’s balance sheet is a kind of “carry trade,” whereby the Fed lends long and borrows short. Goodfriend’s point is that it would have been preferable for the Fed to (i) retain interest earnings while it was making large profits on this carry trade and (ii) then later draw down the accumulated surplus on the back end of the carry trade when short-term interest rates rise. By not doing so, Goodfriend argues that the Fed has compromised its monetary policy goals.

**CONCLUSION**

The Fed is currently faced with an unprecedented problem: how to normalize policy in the face of a large balance-sheet expansion and a very long period with short-term nominal interest rates close to zero. The FOMC has laid out a set of plans (Board of Governors, 2014) that specify a gradualist program for normalization. Balance-sheet reduction is expected to occur through the maturing of long-term Treasury debt and MBS in the Fed’s portfolio after the Fed’s reinvestment program ends. The end of reinvestment will occur sometime after liftoff occurs—the date at which the Fed ends its zero interest rate policy and raises short-term market interest rates.
NOTES

1 The first Operation Twist occurred in 1961. The Fed wanted to “twist,” or flatten, the yield curve on U.S. Treasury securities, thereby increasing short-term interest rates while reducing long-term bond yields. This objective is slightly different from its objective in 2011-12, when the Fed wanted to flatten the yield curve but keep short-term interest rates pegged essentially at zero.

2 This pattern of borrowing and lending on the federal funds market occurs because (i) FHLBs cannot receive interest on reserves and (ii) the branches of foreign banks in the United States have interest-bearing reserve accounts and low balance-sheet costs.

REFERENCES


Gertler, Mark and Karadi, Peter. “QE 1 vs. 2 vs. 3…. A Framework for Analyzing Large-Scale Asset Purchases as a Monetary Policy Tool.” International Journal of Central Banking, January 2013, 9(Suppl. 1), pp. 5-53; http://www.ijcb.org/journal/ijcb13q0a1.pdf.


Williamson


Recent huge U.S. budget deficits—12.8 percent and 12.3 percent of gross domestic product (GDP) in 2009 and 2010, respectively—have returned fiscal issues to the front burner (Calmes, 2009). Analysts typically credit or blame the government for a country’s fiscal situation. Leonhardt (2009), for example, apportions blame for prospective U.S. deficits to current and past presidents. Although Leonhardt (2009) more or less ignores the legislative branch, such assignments are appropriate in some sense: Governments decide how much to tax and spend and therefore are ultimately responsible for fiscal outcomes.

When analyzing fiscal balances, however, it is important to consider economic circumstances, because such circumstances determine the welfare implications and sustainability of fiscal policy. In this article, we analyze the effects of international circumstances on fiscal balances. Two observations motivate our analysis. First, the growth in economic and financial interdependence over the postwar era increases the potential for international circumstances to influence national fiscal policies. Second, Neely’s (2003) casual examination of international circumstances suggests that global budget surpluses are influenced by international economic factors.

The authors use a dynamic latent factor model to analyze comovements in OECD surpluses. The world factor underlying common fluctuations in budget surpluses across countries explains an average of 28 to 44 percent of the variation in individual country surpluses. The world factor, which can be interpreted as a global budget surplus index, declines substantially in the 1980s, rises throughout much of the 1990s, peaks in 2000, and declines again after the financial crisis of 2008. The authors then estimate similar world factors in national output gaps, dividend-to-price ratios, and military spending that significantly explain the variation in the world budget surplus factor. Idiosyncratic components of national budget surpluses correlate with well-known “unusual” country circumstances, such as the Swedish banking crisis of the early 1990s. (JEL C32, E62, F42, H62)
comovements in fiscal balances illustrates the relevance of international influences in such matters.

We begin our analysis by estimating a dynamic factor model to identify the latent world factor underlying fiscal surpluses in 18 industrialized countries for 1980-2013. A latent factor model is a method to summarize common movements in related variables. For example, economists often use latent factor models to characterize the movement of interest rates on bonds of different maturities. The determinants or factors are called latent (hidden) because economists cannot directly observe them. But one can infer the behavior of these unobserved factors from the common movement across variables in the data. Factor analysis estimates these hidden factors to explain as much of the variance of the dependent variables as possible; this effectively condenses the information from many related variables into one or a few underlying influences. When the factors are allowed to be autocorrelated over time, then they are called “dynamic” factors. For interest rates, the first three factors can be readily interpreted as the level, slope, and curvature of the yield curve.

In this article, we use factor analysis to summarize and analyze the comovements of national fiscal balances and investigate the determinants of those balances. Why do we study such comovements? We start by observing that net lending is strongly positively correlated across countries. One might think that some international factor or factors drive this comovement, but their identity and quantitative importance are not clear. Potentially, international budget measures might be related for many reasons. For example, net lending tends to be correlated internationally because trade and capital flows link business conditions between countries. Asset market conditions, such as equity valuations and interest rates, are also linked internationally and can affect fiscal measures through capital gains tax revenues and interest payments on debt. Noneconomic factors, such as common trends in age demographics (e.g., the baby boom) or military expenditures (the peace dividend in the 1990s) can also affect fiscal balances.

A factor method captures covariation among many variables in a unified framework and has major advantages over alternative procedures for measuring comovements in national budget surpluses. For example, the performance of a few large countries will dominate a GDP-weighted average of national surpluses. Similarly, pairwise correlations or related statistics are unwieldy, difficult to summarize, and fail to provide a unified framework.

Our estimated world budget surplus factor explains a substantial portion of the variability in individual budget surpluses in member countries of the Organisation for Economic Co-operation and Development (OECD). Furthermore, the world factor, which can be interpreted as a global budget surplus index, varies markedly over our sample: It declines during the early 1980s and early 1990s, rises sharply for much of the 1990s, peaks in 2000, and declines again after the financial crisis of 2008. Reassuringly, although our procedure does not weight countries by output, it still explains a substantial part of the variability in the U.S. surplus over the sample, which suggests that international factors are relevant even for a very large and relatively closed economy.

We then examine the relation between the world budget surplus factor and estimated world factors in national output gaps, equity valuation ratios, unexpected inflation, and military spending. These variables are potentially important determinants of national budget sur-
pluses and can be viewed as nearly predetermined with respect to fiscal balances. Estimated world factors in national output gaps, price-to-dividend ratios, and military spending significantly explain fluctuations in the world budget surplus factor. Surprisingly, the world output gap factor even significantly explains the world factor in cyclically adjusted surplus measures, which indicates that OECD cyclical adjustments do not remove all business cycle variation in such measures. The fact that the world dividend-to-price ratio factor explains movements in the world budget surplus factor highlights the importance of swings in international equity markets in determining common trends in national budget balances. Finally, the significant relation between world military spending and world budget surplus factors points to the relevance of geopolitical events, such as the end of the Cold War.

In addition to discerning international trends in fiscal situations, the dynamic factor model decomposes national budget surpluses into common and idiosyncratic components. We interpret the common component as the impact of international conditions on a country’s budget surplus. This allows one to evaluate whether the government’s fiscal position is unusual compared with its historical record of budget comovement with similar countries. The common component thus provides a useful benchmark against which to gauge government policies and highlight the importance of particular national circumstances—for example, a war, tax changes, a domestic financial crisis, or atypical terms of trade as distinguished from common reactions to international economic conditions in determining fiscal balances and their sustainability. Substantial fluctuations in the idiosyncratic components of the national budget surpluses often readily relate to well-known “unusual” country circumstances. For example, a sharp decline in the idiosyncratic component of Sweden’s budget surplus in the early 1990s clearly corresponds to the Swedish banking crisis.

While there is a vast literature on topics such as fiscal sustainability and the relation between deficits and growth, there is little work that characterizes international determinants of deficits in industrialized countries. Neely (2003) casually studies recent correlations among national budget deficits and speculates that common shocks to technology, demographics, commodity prices, and political uncertainty drive this covariance. Aside from Neely’s (2003) very short study, two segments of the literature study the causes of deficits and therefore are tangentially related to the present issue of international influences on budget deficits. First, Roubini and Sachs’s (1989) seminal empirical work, related to the theoretical study of Alesina and Tabellini (1990), presents evidence that OECD countries with short-tenure governments and coalition governments are more likely to experience deficits, although Edin and Ohlsson (1991) and de Haan and Sturm (1997) challenge the Roubini-Sachs findings. Second, Lane (2003) finds that OECD countries with volatile output and dispersed political power are more likely to exhibit procyclical fiscal policies; Strawczynski and Zeira (2009) determine that expenditures and deficits react countercyclically to transitory shocks, while government investment reacts procyclically to permanent shocks.

The remainder of the article is organized as follows. The next section outlines the dynamic factor model and its estimation. The third section describes the data and reports dynamic factor model estimation results for national budget surpluses, output gaps, equity valuation ratios, unexpected inflation, and military spending. The fourth section analyzes the relation-
ships between world factors in national budget surpluses and the other variables, and the fifth section examines idiosyncratic components in national budget surpluses.

**ECONOMETRIC METHODOLOGY**

The dynamic latent factor model is given by

\[ y_{i,t} = \beta_i f_t + \epsilon_{i,t}, \]

where \( y_{i,t} \) is the demeaned budget surplus as a share of GDP for country \( i \) (\( i = 1, \ldots, N \)) in year \( t \) (\( t = 1, \ldots, T \)). The world factor, \( f_t \), is common across all of the \( N = 18 \) OECD countries we consider and captures the global comovements in national budget surpluses. \( \beta_i \) is a loading that measures the response of an individual country’s budget surplus to fluctuations in the world factor. The final term in equation (1), \( \epsilon_{i,t} \), is an idiosyncratic component or country-specific factor.

To make equation (1) a dynamic latent factor model, we permit \( f_t \) and \( \epsilon_{i,t} \) to follow auto-regressive (AR) processes. Each idiosyncratic component follows an AR(\( p \)) process, while the world factor obeys an AR(\( q \)) process:

\[ \epsilon_{i,t} = \rho_{i,1} \epsilon_{i,t-1} + \cdots + \rho_{i,p} \epsilon_{i,t-p} + u_{i,t}, \]

\[ f_t = \rho_{f,1} f_{t-1} + \cdots + \rho_{f,q} f_{t-q} + u_{f,t}, \]

where \( u_{i,t} \sim N(0, \sigma_i^2) \), \( u_{f,t} \sim N(0, \sigma_f^2) \), and \( E(u_{i,t} u_{i,t-s}) = E(u_{f,t} u_{f,t-s}) = 0 \) for \( s \neq 0 \). We set \( p = q = 1 \) when estimating the dynamic factor model in the next section; the results are not sensitive to other nonzero values for \( p \) or \( q \). We make the standard assumption that the shocks in equations (2) and (3), \( u_{i,t} \) and \( u_{f,t} \), respectively, are uncorrelated contemporaneously and at all leads and lags, implying that the world and country-specific factors are orthogonal.

Note that neither the signs nor scales of the world factor and factor loadings are separately identified in equation (1). For example, multiplying the world factor by \(-2\) and the loadings by \(-\frac{1}{2}\) produces exactly the same model. To normalize the signs of the factor and loadings, we restrict the loading on the world factor for Australia—the first country (alphabetically) in our sample—to be positive. To normalize the scales, we assume that \( \sigma_f^2 = 1 \) (e.g., see Sargent and Sims, 1977, and Stock and Watson, 1989, 1993). The sign and scale normalizations lack economic content and do not affect any economic inference. Nevertheless, the factor loadings in the next section are typically positive, which means that national budget surpluses are nearly all positively related to the world factor.

The dynamic factor model attributes comovements in national budget surpluses solely to the world factor, \( f_t \), through the factor loadings, \( \beta_i \). That is, \( f_t \) tracks common fluctuations in national budget surpluses. To provide further intuition, consider two extremes. First, if \( \sigma_i^2 = 0 \) and \( \beta_i \neq 0 \) for all \( i \), then \( y_{i,t} = \beta_i f_t \) for all \( i \), so that national budget surpluses are perfectly correlated. At the other extreme, if \( \beta_i = 0 \) and \( \sigma_i^2 \neq 0 \) for all \( i \), then \( y_{i,t} = \epsilon_{i,t} \) for all \( i \), so that national...
budget surpluses are completely uncorrelated. Of course, the patterns in the data are likely to fall between these extremes.

More formally, we can decompose the variation in a country’s budget surplus into the share attributable to the world factor, \( f_i \), and the country-specific factor, \( e_{i,t} \). Given that the factors are orthogonal, this variance decomposition is straightforward to compute for country \( i \):

\[
\theta_i^{\text{world}} = \beta_i^2 \frac{\text{var} \left( f_i \right)}{\text{var} \left( y_{i,t} \right)},
\]

(4)

\[
\theta_i^{\text{country}} = \frac{\text{var} \left( e_{i,t} \right)}{\text{var} \left( y_{i,t} \right)},
\]

(5)

where

\[
\text{var} \left( y_{i,t} \right) = \beta_i^2 \text{var} \left( f_i \right) + \text{var} \left( e_{i,t} \right).
\]

(6)

\( \theta_i^{\text{world}} (\theta_i^{\text{country}}) \) is the proportion of the total variability in country \( i \)’s budget surplus attributable to the world (country-specific) factor. As discussed earlier, the world factor will explain a larger proportion of the variation in countries with high \( \beta_i \) and low \( \text{var} \left( e_{i,t} \right) \) values. That is, these countries will have a higher \( \theta_i^{\text{world}} \) (and lower \( \theta_i^{\text{country}} \)) value and thus will be more closely tied to global fluctuations in national budget surpluses.

Because the world factor is unobservable (latent), we cannot simply estimate it with conventional regression methods. Therefore, we follow Otrok and Whiteman (1998) and Kose, Otrok, and Whiteman (2003, 2008) in estimating the model with a Bayesian approach based on a Markov chain Monte Carlo (MCMC) algorithm to simulate draws from the relevant posterior distributions. We compute posterior distribution properties for the world factor and model parameters based on 10,000 MCMC replications after 2,000 burn-in replications. Otrok and Whiteman (1998) detail the estimation procedure. Because \( \theta_i^{\text{world}} \) and \( \theta_i^{\text{country}} \) are functions of the model parameters and data, we can generate these statistics for each MCMC replication, thereby building up their posterior distributions.\(^6\)

We use the following diffuse conjugate priors, which are similar to those used in Otrok and Whiteman (1998) and Kose, Otrok, and Whiteman (2003, 2008), to implement Bayesian analysis:

\[
\beta_i \sim N(0,1) \quad (i = 1, \ldots, N),
\]

(7)

\[
\left( \rho_{f,1}, \ldots, \rho_{f,d} \right) \sim N \left[ 0, \text{diag} \left( 1,0.5,\ldots,0.5^{d-1} \right) \right] \quad (i = 1, \ldots, N),
\]

(8)

\[
\left( \rho_{f,1}, \ldots, \rho_{f,d} \right) \sim N \left[ 0, \text{diag} \left( 1,0.5,\ldots,0.5^{d-1} \right) \right],
\]

(9)

\[
\sigma_i^2 \sim IG \left( 6, 0.001 \right) \quad (i = 1, \ldots, N),
\]

(10)
where IG denotes the inverse-gamma distribution. Equations (8) and (9) imply that the prior distributions for the AR parameters become more tightly centered on zero as the lag length increases. The prior for the idiosyncratic shock variances given by equation (10) is very diffuse; Otrok and Whiteman (1998) point out that only the first two moments exist for this proper prior. The results reported here are not sensitive to reasonable perturbations of these priors.

We also assume that the AR processes in equations (2) and (3) are stationary, which implies that national budget surpluses are also stationary. This sensible assumption is consistent with the fact that an intertemporal government budget constraint implies a mean-reverting budget deficit.

**DYNAMIC LATENT FACTOR MODEL ESTIMATION RESULTS**

**Data**

We consider four OECD measures of a country’s fiscal position: (i) net lending as a share of GDP; (ii) primary balance as a share of GDP; (iii) cyclically adjusted net lending as a share of potential GDP; and (iv) cyclically adjusted primary balance as a share of potential GDP. Net lending is the most common measure of a country’s fiscal situation—it is the general government budget surplus. The primary balance excludes interest payments from net lending. Cyclically adjusted net lending and primary balances are attempts by the OECD to measure the fiscal balance if the output gap were zero. We use annual data from all 18 OECD countries with full data samples for each of the four measures for the period 1980-2013.

We wish to explain the common variation in our budget surplus measures with other variables that can reasonably be viewed as predetermined with respect to the budget surplus. The output gap is an obvious candidate to explain cyclically unadjusted surpluses. Another candidate is the dividend-to-price ratio, a proxy for transitory but potentially persistent fluctuations in equity prices that provide temporary revenues through capital gains taxes. For example, the U.S. dividend yield and U.S. capital gains taxes as a share of GDP have a correlation of −0.62 from 1970 to 2008. Unexpected inflation can potentially affect debt financing. Finally, we consider whether trends in military spending might explain budget balances. Governments might treat defense spending variation as they typically treat wars—that is, as a temporary change in expenditures to be mostly accommodated by deficit financing rather than by suboptimally large discrete changes in taxation.

We use output gap and consumer price index (CPI) price-level data from the OECD and dividend-to-price ratio data from the Global Financial Data database. We measure unexpected inflation as the first difference in the CPI inflation rate (Atkeson and Ohanian, 2001). We obtain data on military spending from various issues of *World Military Expenditures and Arms Transfers*, which is compiled by the U.S. Department of State, Bureau of Arms Control, Verification and Compliance and obtained from the Inter-University Consortium for Political and Social Research. Military spending is measured as a percentage of GDP.
Summary Statistics

Table 1 displays summary statistics for the fiscal surplus measures from 1980 to 2013. The average fiscal surplus (net lending) was –2.9 percent of GDP, and the average standard deviation was 3.6 percentage points. Extreme deficits or surpluses were fairly common: Eleven of the 18 countries had deficits exceeding 10 percent of GDP, while 5 had surpluses of at least 5 percent of GDP. Cyclically adjusted deficits were somewhat less variable than the unadjusted deficits, with a standard deviation of 3.0 percentage points. The average primary balances were near zero, indicating that government revenues nearly matched expenditures during this sample when one excludes interest payments on previously accumulated debt.

Figure 1 shows the time series of annual fiscal surpluses for the 18 OECD countries during the 1980-2013 sample. The (dashed) solid blue lines indicate (cyclically adjusted) net lending, while the (dashed) solid red lines indicate the (cyclically adjusted) primary balance. A few patterns are noteworthy. First, the values for red lines (primary balances) are generally more positive than those for the corresponding blue lines, reflecting the fact that almost all countries paid interest on existing debt during the sample. Second, the cyclically adjusted balances (dashed lines) do not appear substantially different from the unadjusted balances (solid lines), suggesting that the cyclical adjustments have had little effect. Norway is an exception to both these observations. Norway is unusual in that its primary balances are visibly lower than the full balances because the government receives significant revenues from asset holdings purchased by its sovereign wealth fund, which is funded by oil exports. In addition, the cyclical adjustment significantly changes Norway’s fiscal balances because the OECD’s cyclically adjusted net lending and GDP variables pertain to “mainland” net lending and GDP, which excludes oil production and shipping, rather than to total GDP and net lending.12

Figure 1 suggests that fiscal balances tend to move together internationally; for example, fiscal situations improve in the late 1990s across countries and deteriorate very substantially after the 2008 financial crisis. We use the dynamic factor model to formally measure the common component in national budget surpluses.

Estimation Results for National Budget Surpluses

For each budget surplus measure, Figure 2 shows the mean as well as the 0.10 and 0.90 quantiles of the posterior distribution for the country loadings on the world budget surplus factor. The estimated point loadings are always positive for all four measures, and the interior 80 percent of the posterior distribution generally excludes zero for almost all countries for all measures.13 Increases in the world factor thus imply rising budget surpluses for nearly every country. Japan’s atypically low loadings are unsurprising in light of the particular macroeconomic challenges faced by Japan over much of the sample, including the “lost decade” of the 1990s. Norway has very low loadings for the cyclically adjusted balances, which indicates that non-business-cycle international influences have little effect on its fiscal balances. This lack of effect probably reflects Norway’s position as an oil exporter, which cushions international influences on its economy and budget.14 Italy also tends to have low loadings for all four
Table 1

<table>
<thead>
<tr>
<th>Country</th>
<th>Net lending as a share of GDP</th>
<th>Cyclically adjusted net lending as a share of potential GDP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Australia</td>
<td>-0.015</td>
<td>0.025</td>
</tr>
<tr>
<td>Austria</td>
<td>-0.029</td>
<td>0.014</td>
</tr>
<tr>
<td>Belgium</td>
<td>-0.052</td>
<td>0.045</td>
</tr>
<tr>
<td>Canada</td>
<td>-0.033</td>
<td>0.037</td>
</tr>
<tr>
<td>Denmark</td>
<td>-0.011</td>
<td>0.038</td>
</tr>
<tr>
<td>Finland</td>
<td>0.012</td>
<td>0.041</td>
</tr>
<tr>
<td>France</td>
<td>-0.035</td>
<td>0.016</td>
</tr>
<tr>
<td>Greece</td>
<td>-0.082</td>
<td>0.031</td>
</tr>
<tr>
<td>Iceland</td>
<td>-0.019</td>
<td>0.041</td>
</tr>
<tr>
<td>Ireland</td>
<td>-0.051</td>
<td>0.072</td>
</tr>
<tr>
<td>Italy</td>
<td>-0.067</td>
<td>0.059</td>
</tr>
<tr>
<td>Japan</td>
<td>-0.041</td>
<td>0.072</td>
</tr>
<tr>
<td>Netherlands</td>
<td>-0.031</td>
<td>0.024</td>
</tr>
<tr>
<td>Norway</td>
<td>0.078</td>
<td>0.057</td>
</tr>
<tr>
<td>Spain</td>
<td>-0.047</td>
<td>0.036</td>
</tr>
<tr>
<td>Sweden</td>
<td>-0.013</td>
<td>0.040</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>-0.036</td>
<td>0.033</td>
</tr>
<tr>
<td>United States</td>
<td>-0.051</td>
<td>0.030</td>
</tr>
<tr>
<td>Average</td>
<td>-0.029</td>
<td>0.036</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Country</th>
<th>Primary balance as a share of GDP</th>
<th>Cyclically adjusted primary balance as a share of potential GDP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Australia</td>
<td>0.000</td>
<td>0.022</td>
</tr>
<tr>
<td>Austria</td>
<td>-0.004</td>
<td>0.013</td>
</tr>
<tr>
<td>Belgium</td>
<td>0.017</td>
<td>0.037</td>
</tr>
<tr>
<td>Canada</td>
<td>-0.006</td>
<td>0.035</td>
</tr>
<tr>
<td>Denmark</td>
<td>0.016</td>
<td>0.038</td>
</tr>
<tr>
<td>Finland</td>
<td>0.009</td>
<td>0.040</td>
</tr>
<tr>
<td>France</td>
<td>-0.012</td>
<td>0.015</td>
</tr>
<tr>
<td>Greece</td>
<td>-0.023</td>
<td>0.035</td>
</tr>
<tr>
<td>Iceland</td>
<td>-0.003</td>
<td>0.038</td>
</tr>
<tr>
<td>Ireland</td>
<td>-0.012</td>
<td>0.069</td>
</tr>
<tr>
<td>Italy</td>
<td>0.003</td>
<td>0.032</td>
</tr>
<tr>
<td>Japan</td>
<td>-0.031</td>
<td>0.037</td>
</tr>
<tr>
<td>Netherlands</td>
<td>0.000</td>
<td>0.023</td>
</tr>
<tr>
<td>Norway</td>
<td>0.059</td>
<td>0.057</td>
</tr>
<tr>
<td>Spain</td>
<td>-0.024</td>
<td>0.038</td>
</tr>
<tr>
<td>Sweden</td>
<td>0.000</td>
<td>0.039</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>-0.010</td>
<td>0.033</td>
</tr>
<tr>
<td>United States</td>
<td>-0.017</td>
<td>0.031</td>
</tr>
<tr>
<td>Average</td>
<td>-0.002</td>
<td>0.035</td>
</tr>
</tbody>
</table>

NOTE: “Average” is the average across all countries. SD, standard deviation.
Figure 1
Annual Budget Surpluses: 18 OECD Countries (1980-2013)
measures, which means that its high deficits over the sample period have only a very modest positive relation to the primary international factor that affects deficits in other countries. Figure 3 displays the mean and the 0.10, 0.33, 0.66, and 0.90 quantiles of the posterior distribution for the world factor in each of the four budget surplus measures. \(^{15}\) We note that removing interest payments from the budget balances makes little difference to the patterns in the world factors; compare the world factor for net lending with that for the primary balance.
Figure 3 illustrates significant fluctuations in the world factor for each of the fiscal measures: a fall in the early 1980s, a rise to a local maximum in 1989, another downturn to a trough in 1993, a subsequent rise leading to a global maximum in 2000, a relative plateau from 2000 to 2006, and then a precipitous decline after the financial crisis of 2007-09. The 80 percent posterior coverage regions generally exclude zero. Overall, Figure 3 points to substantial common fluctuations in national budget surpluses.

Figure 4 illustrates $\Theta_{\text{world}}$ variance decompositions, which measure the extent to which international influences affect national fiscal balances. As in Figure 2, the blue circles correspond to the mean of the posterior distribution for each country, while the vertical blue bars delineate the 0.10 and 0.90 quantiles. On average across the 18 countries, the point estimates indicate that the world factor explains 44 percent of total variance for net lending, 28 percent for cyclically adjusted net lending, 44 percent for the primary balance, and 35 percent for the cyclically adjusted primary balance. The variance decompositions are precisely measured. The difference between the cyclically adjusted and unadjusted measures suggests that the world business cycle explains at least part of the global influence on deficits. The variation in the
cyclically adjusted measures indicates that there are other important global (non-business-cycle) influences or that the cyclical adjustment is inadequate. In sum, Figures 2 through 4 illustrate that common fluctuations in OECD national budget surpluses represent a significant portion of the variability in these surpluses. These common movements in cyclically adjusted and primary balances indicate that global influences on fiscal balances extend beyond business cycle and interest rate effects.

**Estimation Results for Predetermined Variables**

To explain the variation in the four measures of fiscal balances, we first compute world factors for national output gaps, dividend-to-price ratios, unexpected inflation, and military spending, which we treat as nearly predetermined driving variables. These variables are not...
Figure 5
Loadings on the World Factor for Predetermined Variables (1980-2013)

NOTE: The blue circles indicate the mean, and the vertical blue bars delineate the 0.10 and 0.90 quantiles for the posterior distribution of each country. See Figure 2 for definitions of country abbreviations.
truly exogenous, of course, but it seems reasonable to treat them as predetermined because we do not think the factor in global fiscal balances has a strong contemporaneous influence on them. To test the sensitivity of our results to this quasi-exogeneity assumption, we estimate instrumental variable regressions with lagged regressors as instruments and compare the results with those of ordinary least squares (OLS). The two sets of results are similar; this similarity supports the quasi-exogeneity assumption.

We compute the world factors in these variables in the same way that we computed the world factors for the fiscal balances. Figure 5 displays the mean and 0.10 and 0.90 quantiles for each country’s loading on the world factor for each quasi-exogenous variable. The point estimates of the loadings indicate that each variable for each country is positively related to the world factor, with the exception of military spending for Japan.

Figure 6
World Factors for Predetermined Variables (1980-2013)

NOTE: The dashed black line delineates the mean of the posterior distribution. The blue (red) lines delineate the 0.33 and 0.66 (0.10 and 0.90) quantiles for the posterior distribution. The world factor for military spending is estimated for 1980-2010.
Figure 6 portrays the estimated world factor for each of the predetermined variables. The world factor for the output gap displays a temporal pattern similar to that in net lending and the primary balance. The 1990s bull market in global equities is clearly evident in the dividend yield world factor (high equity prices and thus low dividend-to-price ratios), as well as the 2008 plunge in prices. The world factor in unexpected inflation visibly covaries positively with the world output gap factor—the correlation between the series is 0.61—which is in line with an expectations-augmented Phillips curve. World factors in output gaps, dividend-to-price ratios, unexpected inflation, and military spending fluctuate substantially from 1980 to 2013 and, with the exception of the military spending factor, are estimated reasonably precisely. The world factor in military spending is estimated very imprecisely, but there is a notable decrease in the early 1990s, shortly after the fall of the Berlin Wall. In the next section, we use the world factors for the four predetermined variables to explain the world fiscal surplus factors.

### RELATING PREDETERMINED VARIABLES TO BUDGET SURPLUSES

A priori, we expect that the output gap significantly explains net lending and primary balances but does not explain the cyclically adjusted versions of those measures. We also conjecture that the dividend-to-price ratio is negatively related to all fiscal balances through capital gains taxes because, as stock prices exceed fundamental values, government revenues will rise above typical levels. Examination of U.S. capital gains tax receipt data (omitted for brevity) indicates that such receipts can vary by almost 1 percent of GDP within a few years. Unexpected inflation could influence fiscal deficits in either direction. On the one hand, if higher unexpected inflation signals an adverse aggregate supply shock, then one would expect it to reduce fiscal surpluses. Similarly, higher unexpected inflation could increase the cost of financing the short-term portion of the debt. On the other hand, if monetary stimulus produces unexpected inflation, one might expect a larger fiscal surplus. Finally, we expect that defense spending would be negatively related to all fiscal balances. That is, we expect that taxes would not always be immediately adjusted for changes in defense spending.

To explore the determinants of budget balances, we regress the world fiscal balance factors on world factors for the output gaps, dividend-to-price ratios, unexpected inflation, and military spending. We estimate both bivariate and multivariate regression models to contrast the results and highlight the dependencies in the explanatory variables. The bivariate regression model takes the following form:

\[
    f_t^{\text{surplus}} = a + b_j f_t^j + e_t^{\text{surplus}},
\]

where \(f_t^{\text{surplus}}\) is the world factor for the fiscal surplus in year \(t\) and \(f_t^j\) is the world factor for one of the four explanatory variables, indexed by \(j\): output gaps, dividend-to-price ratios, unexpected inflation, and military spending. The multivariate regression is as follows:

\[
    f_t^{\text{surplus}} = a + \sum_{j=1}^4 b_j f_t^j + e_t^{\text{surplus}}.
\]
We estimate equations (11) and (12) using OLS, accounting for autocorrelation with Newey-West (1987) standard errors.

We present the regression results with two caveats. First, the factors on both the left- and right-hand sides of the regressions are generated variables. The error in the left-hand-side variables (i.e., the world budget surplus factors) will decrease the apparent amount of predictability in the relations. The estimated $R^2$ value will then understate the $R^2$ value that is theoretically expected in the absence of measurement error because the estimated total sum of squares will exceed the total sum of squares without measurement error. Likewise, the error in the predetermined variables on the right-hand side will attenuate their estimated coefficients toward zero and thus inflate their $p$-values. Therefore, the error in the factor estimation will cause our regressions to present a conservative picture of the relation between the fiscal surpluses and predetermined variables.

Second, we view the right-hand-side variables in equations (11) and (12) as nearly predetermined. Strictly speaking, these variables are endogenous, which means the coefficients will be subject to simultaneity bias. We believe the explanatory variables are largely predetermined, however, and unlikely to exhibit strong contemporaneous reactions to fiscal balances. Therefore, we do not believe that simultaneity bias will strongly influence our results.

Table 2 presents the bivariate and multiple regression results for all four fiscal surplus measures. The sample period is 1980-2013, except for regressions including military spending, for which the sample period is 1980-2010. Given that including military spending reduces the sample length, is imprecisely estimated, and is very persistent, we estimate multiple regression models both with and without this variable.

In the bivariate regressions, the output gap factor is positive and significant at the 1 percent level for all four fiscal measures, presumably through the familiar taxing and spending channels. International business cycle fluctuations have the most explanatory power for the unadjusted surpluses: net lending and the primary balance, with $R^2$ statistics of 73 percent and 59 percent, respectively. This is not surprising, as cyclical adjustment would be expected to remove some or all international influences. The explanatory power of the output gap factor for the cyclically adjusted surpluses is surprisingly large, however, with very sizable $R^2$ statistics of 41 percent and 24 percent for the cyclically adjusted net lending and cyclically adjusted primary balance, respectively. The OECD’s cyclical adjustments apparently do not completely capture international business cycle effects on budgets.

Consistent with the idea that higher equity prices increase capital gains tax revenues, the dividend-to-price ratio factor is significantly negatively related to cyclically adjusted net lending, the primary balance, and cyclically adjusted primary balance factors in the bivariate regressions. The $R^2$ statistics are sizable: 15 percent, 18 percent, and 28 percent for cyclically adjusted net lending, the primary balance, and cyclically adjusted primary balance, respectively. Our results indicate that global bull (bear) equity markets significantly raise (lower) the primary balance in industrialized countries. The dividend-to-price ratio factor is not significantly related to the net lending factor, although the relationship is nearly significant at the 10 percent level. The dividend-to-price factor explains more of the variability in the primary balances than in the non-primary surpluses. A systematic relationship between global equity valuations and interest rates could create this difference.
### Table 2

**OLS Estimation Results: Bivariate and Multivariate Regression Models (1980-2013)**

<table>
<thead>
<tr>
<th>Regressor</th>
<th>Bivariate regression</th>
<th></th>
<th></th>
<th></th>
<th>Multivariate regression</th>
<th></th>
<th></th>
<th>Multivariate regression, excluding military spending</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>t-Statistic</td>
<td>$R^2$</td>
<td>Coefficient</td>
<td>t-Statistic</td>
<td>$R^2$</td>
<td>Coefficient</td>
<td>t-Statistic</td>
<td>$R^2$</td>
<td></td>
</tr>
<tr>
<td><strong>Panel A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regressand = Net lending, world factor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Output gap, world factor</td>
<td>0.99</td>
<td>(7.72)</td>
<td>0.73</td>
<td>1.17</td>
<td>(9.07)</td>
<td>0.87</td>
<td>1.01</td>
<td>(7.62)</td>
<td>0.81</td>
<td></td>
</tr>
<tr>
<td>Dividend-to-price ratio, world factor</td>
<td>-0.56</td>
<td>-(1.59)</td>
<td>0.09</td>
<td>-0.69</td>
<td>-(4.09)</td>
<td></td>
<td>-0.52</td>
<td>-(4.17)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unexpected inflation, world factor</td>
<td>2.53</td>
<td>(2.76)</td>
<td>0.16</td>
<td>-0.25</td>
<td>-(0.56)</td>
<td></td>
<td>-0.26</td>
<td>-(0.56)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Military spending, world factor</td>
<td>0.08</td>
<td>(2.78)</td>
<td>0.11</td>
<td>-0.04</td>
<td>-(2.11)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Panel B</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regressand = Cyclically adjusted net lending, world factor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Output gap, world factor</td>
<td>0.61</td>
<td>(3.99)</td>
<td>0.41</td>
<td>0.81</td>
<td>(5.27)</td>
<td>0.69</td>
<td>0.63</td>
<td>(4.08)</td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td>Dividend-to-price ratio, world factor</td>
<td>-0.57</td>
<td>-(2.12)</td>
<td>0.15</td>
<td>-0.86</td>
<td>-(4.26)</td>
<td></td>
<td>-0.55</td>
<td>-(2.87)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unexpected inflation, world factor</td>
<td>1.43</td>
<td>(1.90)</td>
<td>0.08</td>
<td>-0.06</td>
<td>-(0.10)</td>
<td></td>
<td>-0.30</td>
<td>-(0.43)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Military spending, world factor</td>
<td>0.04</td>
<td>(1.47)</td>
<td>0.04</td>
<td>-0.08</td>
<td>-(2.54)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Panel C</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regressand = Primary balance, world factor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Output gap, world factor</td>
<td>0.86</td>
<td>(5.80)</td>
<td>0.59</td>
<td>0.93</td>
<td>(4.25)</td>
<td>0.78</td>
<td>0.87</td>
<td>(5.08)</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>Dividend-to-price ratio, world factor</td>
<td>-0.74</td>
<td>-(2.53)</td>
<td>0.18</td>
<td>-0.93</td>
<td>-(6.24)</td>
<td></td>
<td>-0.70</td>
<td>-(5.38)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unexpected inflation, world factor</td>
<td>2.11</td>
<td>(2.02)</td>
<td>0.12</td>
<td>0.15</td>
<td>(0.19)</td>
<td></td>
<td>-0.29</td>
<td>-(0.37)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Military spending, world factor</td>
<td>0.07</td>
<td>(2.21)</td>
<td>0.11</td>
<td>-0.06</td>
<td>-(2.89)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Panel D</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regressand = Cyclically adjusted primary balance, world factor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Output gap, world factor</td>
<td>0.48</td>
<td>(2.69)</td>
<td>0.24</td>
<td>0.55</td>
<td>(2.18)</td>
<td>0.58</td>
<td>0.49</td>
<td>(2.44)</td>
<td>0.51</td>
<td></td>
</tr>
<tr>
<td>Dividend-to-price ratio, world factor</td>
<td>-0.81</td>
<td>-(3.71)</td>
<td>0.28</td>
<td>-1.11</td>
<td>-(6.34)</td>
<td></td>
<td>-0.79</td>
<td>-(4.45)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unexpected inflation, world factor</td>
<td>1.05</td>
<td>(1.10)</td>
<td>0.04</td>
<td>0.32</td>
<td>(0.34)</td>
<td></td>
<td>-0.30</td>
<td>-(0.30)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Military spending, world factor</td>
<td>0.04</td>
<td>(1.49)</td>
<td>0.05</td>
<td>-0.08</td>
<td>-(2.78)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**NOTE:** The t-statistics are based on Newey-West standard errors. The sample period is 1980-2010 for regression models that include the world factor for military spending as a regressor.
The unexpected inflation factor significantly explains net lending, cyclically adjusted net lending, and the primary balance in the bivariate regressions. The $R$-squared values are modest, ranging from 8 to 16 percent. As noted in the section “Estimation Results for Predetermined Variables,” the unexpected inflation factor is positively correlated with the output gap factor, so the significantly positive coefficients on the unexpected inflation factor likely capture similar business cycle effects.

The military spending factor is significant at the 1 percent level in the bivariate regression model for net lending and the primary balance but not the cyclically adjusted measures. The $R$-squared statistics are modest: 11 percent for both the net lending and primary balance measures. The estimated positive coefficients are counterintuitive; they likely reflect long-term upward trends in deficits as military spending as a percentage of GDP declines. Thus, they are a spurious product of a regression with a very persistent variable.

In the multiple regressions, the output gap, the dividend-to-price ratio, and the military spending factors are significant at conventional levels for all four measures. The significance of the output gap factor confirms that the cyclical adjustments do not completely capture international business cycle effects. Unexpected inflation coefficients are no longer significant in any of the multiple regressions, probably because the unexpected inflation factor is strongly correlated with the output gap factor. The military spending factor significantly explains all of the fiscal surplus factors at conventional levels. Importantly, the signs of the coefficient on the military spending factor become reliably and significantly negative, as one would expect, when the other variables are controlled for. The $R$-squared statistics in the sixth column of Table 2 show that world factors in the four predetermined variables collectively explain most of the variability in the global budget surplus factors, especially for net lending and the primary balance, where the $R$-squared statistics are 87 percent and 78 percent, respectively.

The imprecise estimation and strong persistence in the military spending variable are causes for concern. Therefore, we also estimate the multiple regression without the military variable and use a 1980-2013 sample. In this specification, the output gap and dividend-to-price ratio factors remain significant at the 1 percent level for each of the four surplus factors. The unexpected inflation factor remains insignificant at conventional levels in each of the four regressions. The $R$-squared statistics continue to be substantial in the final column of Table 2, ranging from 51 to 81 percent.

In summary, Table 2 indicates that the output gap, the price-to-divided ratio, and the military spending world factors substantially determine fluctuations in fiscal surplus world factors. Unexpected inflation also has predictive value when considered by itself but not in conjunction with the other variables. Global expansions, bullish equity markets, and reduced military spending improve fiscal balances across industrialized countries.\(^{19}\)

**IDIOSYNCRATIC COMPONENTS**

Our method of investigating international influences on fiscal balances permits us to isolate the effect of domestic events on fiscal balances. That is, we can examine the common and idiosyncratic components of budget surpluses to determine the effect of domestic events or
policies. Figure 7 displays common and idiosyncratic components for net lending of selected countries.20

The top-left panel of Figure 7 shows demeaned U.S. net lending and its two components: the common component (the product of the world factor and its loading) and the U.S. idiosyncratic component. Demeaned net lending is the sum of the common and idiosyncratic components, of course. Note that because net lending is demeaned and the sample mean for U.S. net lending was −5.1 percent, values of demeaned net lending near zero still indicate fairly high deficits. The figure illustrates that both global and idiosyncratic components contributed to all the major movements in U.S. net lending. For example, both components contributed to the increase in deficits—that is, the decrease in net lending—in the early 1980s and the movement from substantial deficits to surpluses in the 1990s. The substantial deterioration in the U.S. fiscal balance in 2001 partly reflected the common component but was mostly due to the U.S. idiosyncratic component, however. That is, U.S. factors—such as the 2001 tax cuts, the September 11th attacks, and the wars in Afghanistan and Iraq—bore the lion’s share of the

---

**Figure 7**

**Common and Idiosyncratic Components for Demeaned Net Lending: Selected Countries (1980-2013)**

---

20 Neely and Rapach

Federal Reserve Bank of St. Louis REVIEW  
Second Quarter 2015  
127
blame for the decline in the fiscal situation during that period. The fastest changes in U.S. net lending occurred during the 2008 financial crisis, however—which, again, were driven by large declines in both the common and idiosyncratic components.

The upper-right panel of Figure 7 portrays the idiosyncratic components for a pair of highly indebted European countries, Belgium and Greece. The idiosyncratic components were quite different in these two countries during the 1980s. Both countries, however, faced pressure in the 1990s to reduce their debt and deficits to levels required by the Treaty on European Union (informally, the Maastricht treaty) for entry into the European Economic and Monetary Union on January 1, 1999. This regional influence is clearly evident during the rise in the idiosyncratic components for these countries during the 1990s.

The lower-left panel of Figure 7 shows the common and idiosyncratic components for Sweden and highlights the important fiscal effect of the Swedish banking crisis of 1990-94. During the late 1980s, the idiosyncratic component contributed to a marked improvement in Sweden’s fiscal surplus. With the advent of the banking crisis in 1990, however, Sweden was forced to spend relatively large sums to recapitalize its banking systems, which resulted in a sharp decrease in the idiosyncratic component of net lending during the early 1990s. The common component also decreased in the early 1990s; thus, the early 1990s were characterized by a steep decline in overall Swedish net lending. As one might expect, the resolution of the banking crisis led to a sizable increase in the idiosyncratic component during the late 1990s.

Finally, the lower-right panel of Figure 7 illustrates the importance of the oil market for Norway. In addition to the Norwegian idiosyncratic component, the figure shows the value of Norwegian oil exports as a share of GDP. The two variables generally move together, indicating that oil revenues are especially important for improving the fiscal situation in Norway. Observe, however, that oil revenues increased while the idiosyncratic component decreased around 1990. This likely reflects the influence of the Scandinavian banking crisis that affected Norway and started earlier than the Swedish crisis (Vale, 2004). The increase in Norway’s oil revenues during this time helped to cushion the negative budgetary impulse of the banking crisis.

In summary, decomposing net lending into common and idiosyncratic components allows us to more easily evaluate the effects of domestic events and policies on a country’s fiscal situation.

CONCLUSION

The emergence of the prospect of unprecedented deficits in the United States has rekindled interest in the causes of such imbalances and the question of responsibility for them. Properly addressing these imbalances requires understanding their sources and influences, including international influences.

While researchers, such as Roubini and Sachs (1989), have examined how political polarization might affect deficits, and others, such as Lane (2003), have evaluated the cyclicity of deficits, there has been no significant previous work on internationally driven comovements in deficits. In this article, we use a dynamic latent factor model to identify substantial international comovements in four budget surplus measures for 18 OECD countries for 1980-2013.
with a dynamic latent factor model. Depending on the measure of the fiscal surplus, the world factor explains 28 to 44 percent of surplus variability, on average, across countries. The world factor explains 47 percent of the variation in U.S. net lending, for example.

World factors in national output gaps, dividend-to-price ratios, and military spending usually significantly explain variation in the four world fiscal surplus factors. Surprisingly, the output gap factor significantly explains not only the net lending and primary balance factors, but also the cyclically adjusted versions of those measures. This indicates that the OECD cyclical adjustments do not completely remove the contribution of the international business cycle on fiscal balances. The importance of the world dividend-to-price ratio factor highlights the role of global equity market conditions in affecting fiscal balances, while the significance of the military spending factor points to the effect of an international peace dividend in the 1990s.

Our results show that trends in international business cycles, equity markets, and military spending create common fluctuations in national budget surpluses. The discovery of a significant global factor in international budget deficits suggests avenues for future research. What global political economy incentives influence fiscal balances? Do individual governments respond optimally to these international shocks? Can individual country characteristics explain varying sensitivities of national fiscal balances to international influences? Our findings highlight the relevance of such questions.
Factor analysis has been used to model covariation in many types of related variables. For example, individuals who are good at certain mental or physical tasks are very often good at other types of mental or physical tasks that are not directly related. That is, students who get an A in economics are likely to have above-average grades in other courses. Charles Spearman, a psychologist, developed factor analysis to describe the tendency of children’s performance on cognitive tasks to be positively correlated.

Researchers have recently employed dynamic latent factor models to measure global fluctuations in national real output growth and inflation rates; see, for example, Kose, Otrok, and Whiteman (2003, 2008) with respect to real output growth and Ciccarelli and Mojon (2010), Monacelli and Sala (2009), and Neely and Rapach (2011) with respect to inflation.

In the dynamic latent factor models discussed in the next section, $y_{jt}$ can also represent the demeaned national output gap, dividend-to-price ratio, unexpected inflation rate, or military spending as a share of GDP.

The comovement in net-lending-to-GDP ratios is driven almost entirely by comovement in net lending rather than by comovement in GDP. International net lending correlations are, on average, very similar if one uses the predicted value of GDP from a log linear trend as the denominator in the net lending ratio rather than GDP itself. Iceland and Japan show the most evidence of correlation through deficits and growth.

The latent world factor could also be estimated using principal components (Stock and Watson, 2002, and Bai, 2003), with inferences based on the asymptotic distribution theory in Bai (2003). Principal component estimates of the world factors are similar to the Bayesian estimates. The Bayesian approach, however, is likely to provide more accurate finite-sample results as the asymptotic theory in Bai (2003) is based on $N \to \infty$ and $T \to \infty$.

We enforce the stationarity restrictions by discarding draws of the AR parameters that do not satisfy the restrictions. We do the same to enforce the sign restriction on the factor loading for Australia. Inadmissible AR parameters and Australian loadings are rarely drawn, especially after the burn-in replications.

The OECD defines “general government accounts” as follows: “General government accounts are consolidated central, state and local government accounts, social security funds and non-market non-profit institutions controlled and mainly financed by government units” (see http://stats.oecd.org/glossary/detail.asp?ID=1095).

The OECD denotes the four measures as “central government net lending—as a percentage of GDP,” “government primary balance—as a percentage of GDP,” “cyclically adjusted government net lending—as a percentage of potential GDP,” and “cyclically adjusted government primary balance—as a percentage of potential GDP.” The OECD describes its cyclical adjustment method at http://www.oecd.org/dataoecd/0/61/36336878.pdf.

The OECD denotes these variables as “Output gap of the total economy” and “Consumer Price Index.” Full-sample dividend-to-price ratio data are unavailable for Iceland, Ireland, and Spain, and we exclude these countries when estimating the dividend-to-price ratio world factor in the section “Estimation Results for Predetermined Variables.”

The current issue of World Military Expenditures and Arms Transfers is available at http://www.state.gov/t/avc/rls/rpt/wmeat/2014/index.htm. Back issues were downloaded from https://www.icpsr.umich.edu/icpsrweb/ICPSR/series/61. Military spending data are available through 2005. Data are unavailable for Iceland, and we exclude Iceland when we estimate the military spending world factor in the section “Estimation Results for Predetermined Variables.”

The Norwegian government owns all petroleum resources on the Norwegian continental shelf. Taxes and license fees from the petroleum sector go to the Government Pension Fund of Norway, which uses them both for long-term investment and directly for government expenditures. Oil profits are taxed at very high rates, and revenues from those taxes reached $36 billion in 2011, or almost 8.6 percent of Norwegian GDP (Hsieh, 2013). See http://www.oecd.org/norway/47473811.pdf for the OECD’s definition of mainland GDP.

We use the mean of the posterior distribution as the point estimate.

The United Kingdom was also an oil exporter for most of the sample, but its oil exports were smaller in absolute value and much less important compared with the size of its economy and government budget.
Observe that the world budget surplus factor is an index, so a world surplus factor of zero in Figure 3 does not necessarily represent a balanced budget. As expected, the $\beta_i$ and $\theta_{\text{world}}$ estimates are positively correlated across countries, with correlation coefficients of 0.30, 0.53, 0.26, and 0.53 for net lending, cyclically adjusted net lending, primary surplus, and cyclically adjusted primary surplus, respectively. The world factors typically explain a substantial portion of the variability in national output gaps, price-to-divided ratios, unexpected inflation, and military spending, with averages across countries of 0.55, 0.57, 0.42, and 0.35, respectively. For brevity, we do not report the complete results for the variance decompositions; these are available upon request from the authors. Our exercise is similar in spirit to the work of Crucini, Kose, and Otrok (2011) in the context of explaining the G-7 business cycle. They first estimate a world factor in G-7 real output growth rates, which they then explain using world factors in G-7 measures of productivity, fiscal policy, monetary policy, oil prices, and terms of trade.

We also estimated fixed-effects panel regression models with national fiscal surpluses serving as regressands and national output gaps, price-to-dividend ratios, unexpected inflation, and military spending serving as regressors. For brevity, we do not report complete results but they are available upon request from the authors. The national output gap and military spending are significant determinants of national net lending and cyclically adjusted net lending, while the national output gap, the dividend-to-price ratio, and military spending are significantly related to the national primary balance and cyclically adjusted primary balance. Of course, panel estimation does not explicitly identify world factors in national budget surpluses and their determinants—the focus of our article—but it does appear to pick up aspects of the links documented in Table 2.

The complete set of common and idiosyncratic components is available upon request from the authors.

REFERENCES


Neely and Rapach


The U.S. economy is the largest in the world. In 2012, the United States accounted for 22.4 percent of the world’s gross domestic product (GDP) and 35.1 percent of the world’s total market capitalization (World Bank, 2012). The Great Recession of 2007-09 highlighted the importance of the United States to the global economy. A financial shock originating for the most part in the United States led to a worldwide downturn that had detrimental and lasting effects on both developed and emerging economies. This dynamic is summarized by this modern version of a well-known quotation: “When the U.S. sneezes, the rest of the world catches a cold.”

Given the role of the United States as a global economic leader, several recent studies investigate the spillover effects of the U.S. economy onto the economies of other nations. Arora and Vamvakidis (2004) use a fixed-effects panel regression and find that U.S. economic growth has positive effects on the rest of the world, especially developing countries. Helbling et al. (2007) use multiple methodologies to determine the effect of the U.S. economy on other countries. They conduct an event study and find that U.S. recessionary periods coincide with global downturns. They also use simple regressions and find that, controlling for potential common unobserved shocks and country-specific effects, a 1-percentage-point decline in U.S. growth leads to an average 0.16-percentage-point decline in output growth across their sample of economies.
countries. Canada, Latin American, and Caribbean countries are the most strongly influenced within their sample. Lastly, they use the more dynamic approach of structural vector autoregressions to allow for both foreign and domestic effects. They find that U.S. growth has significant effects on growth in Latin America, the Asian newly industrialized economies (Hong Kong, Korea, Singapore, and Taiwan), and some of the Association of Southeast Asian Nations (Indonesia, Malaysia, the Philippines, and Thailand). Antonakakis (2012) uses a dynamic measure of correlation to examine the synchronization of G-7 business cycles across a long time series (1870 to 2011). He finds U.S. recessions have positive effects on business cycle comovements after the 1971 breakdown of the Bretton Woods system, with an increased level of synchronization during the Great Recession.

The goal of this article is to assess the influence of U.S. output growth on the business cycles of other nations. In particular, we ask whether U.S. economic growth signals economic turning points in other countries. In our model, we cannot determine which structural innovations (shocks) drive spillovers from the United States onto the economies of other countries or whether the proximate shock leading to the turning point is global in nature. Rather, we are merely interested in the comovement between U.S. output and economic downturns in other countries. However, we do analyze the timing and duration of the effects of the U.S. economy on the business cycles of other countries. Accordingly, we could appeal to other studies regarding which driving forces occurred during a given time period.1

Despite the inability of our model to offer a complete characterization of these shocks, our study should be of relevant interest to policymakers and others interested in the dependence of foreign business cycles on the U.S. economy. Our results imply that the trajectory of U.S. output growth informs both the timing and duration of economic turning points in certain foreign economies. Proper analysis of these cross-country linkages gives policymakers, both in the United States and abroad, a better understanding of the trade-offs faced when conducting independent and coordinated actions.

Since our focus is on economic turning points, we use the regime-switching model of Hamilton (1989) with time-varying transition probabilities (TVTP) as outlined by Goldfeld and Quandt (1973), Diebold, Lee, and Weinbach (1994), and Filardo (1994). This framework allows us to identify not only the economic turning points but also the extent to which U.S. output growth influences the evolution of the underlying state—recession or expansion—of a nation’s economy. We consider regime-switching models with both two states (“recession” and “expansion”) and three states (“recession,” “low-growth expansion,” and “high-growth expansion”).

Our panel of countries includes Canada, France, Germany, Italy, Japan, Mexico, and the United Kingdom and covers the period 1960:Q2–2013:Q4. We find that U.S. output growth informs the timing and duration of recessions for Canada, Germany, the United Kingdom, and, to a lesser extent, Mexico. We find no relationship between U.S. output growth and business cycle turning points for the remaining countries (France, Italy, and Japan).

The article proceeds as follows: The next section details the regime-switching model and is followed by a section that describes the data and outlines the estimation methodology. Our results are presented in the next section, followed by our conclusions.
MODEL

Burns and Mitchell (1946) characterize the business cycle as distinct phases of expansion and recession. As defined by the National Bureau of Economic Research (NBER) Business Cycle Dating Committee, a recession is a widespread decline in economic activity typically lasting from a few months to over a year. On the other hand, expansions are characterized by positive growth in economic activity and, typically, longer durations.

Models of a country’s business cycle are typically estimated using only that country’s data. Regime shifts are characterized by sudden and persistent shifts in the growth rate of the economic indicators, usually domestic GDP. In this article, we are interested in the contagion of economic outcomes across countries. To this end, we augment the standard business cycle model to account for possible contagion by a dominant country—in this case, the United States.

The model we adopt is based on the business cycle model of Hamilton (1989), who characterizes the cycle as a two-state process with random regime changes. In his framework, the mean growth rate of a country’s output, $y_t$, depends on a latent state variable, $s_t = \{1,2\}$. The state of the economy at any time is either recession ($s_t = 1$) or expansion ($s_t = 2$). Assuming no autoregressive terms for simplicity, this model is given by

$$y_t = \begin{cases} 
\mu_1 + \epsilon_t, & \text{if } s_t = 1 \text{ (recession)} \\
\mu_2 + \epsilon_t, & \text{if } s_t = 2 \text{ (expansion)}
\end{cases},$$

where the error variance, $\epsilon_t \sim N(0,\sigma^2)$, is constant across states. Consistent with the NBER's definition of the business cycle, we restrict the average growth rate of output to be positive during expansionary periods ($\mu_2 > 0$) and negative during recessionary periods ($\mu_1 > 0$).

In principle, we could include any number of states $K$ in the model to better match certain features of business cycles. For example, Kim and Piger (2002), Kim and Murray (2002), and Billio et al. (2013) include three states in their regime-switching model of the business cycle. Additional states can reflect persistent differences in business cycle characteristics such as fast- versus slow-growth expansion regimes or deep versus shallow recessions. The generalized $K$-state model is given by

$$y_t = \begin{cases} 
\mu_1 + \epsilon_t, & \text{if } s_t = 1, \\
\mu_2 + \epsilon_t, & \text{if } s_t = 2, \\
\vdots & \vdots \\
\mu_K + \epsilon_t, & \text{if } s_t = K,
\end{cases}$$

with the identifying restriction $\mu_1 < \mu_2 < \ldots < \mu_K$. We consider both a two-state (recession and expansion) and a three-state (recession, low-growth expansion, and high-growth expansion) model for each country. We normalize the states such that $\mu_1 < 0 < \mu_2 < \mu_3$. This normalization provides econometric identification as well as an interpretation for future discussion.
Transition Probabilities

The NBER’s Business Cycle Dating Committee provides ex post historical dates for periods during which the U.S. economy is in expansion or recession. Many other countries do not have “official” business cycle turning points. The model leaves the state of the economy unobserved and, therefore, requires an assumption about the evolution process of the state variable. Ideally, a model of economic business cycles matches two features of the data: (i) Both expansions and recessions are highly persistent and (ii) expansions have longer average durations than recessions.

A standard assumption with regime-switching models is that the state variable follows a first-order Markov process with fixed transition probabilities (FTPs; e.g., as in Hamilton, 1989). The Markov property imposes that the current value of the state variable, $s_t$, is a function of its previous value, $s_{t-1}$. In the two-state model, the transition matrix governing the Markov process is represented as

$$P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix},$$

with FTPs

$$p_{ii} = \Pr[s_t = i | s_{t-1} = i],$$

where the columns of $P$ each sum to 1 (i.e., $\sum_p p_{ij} = 1$ for $i = 1,2$). Thus, if a country was in expansion during the previous period ($s_{t-1} = 2$), the probability that it remains in expansion this period ($s_t = 2$) is $p_{22}$, and the probability that the economy enters a recession this period ($s_t = 1$) is $p_{12} = 1 - p_{22}$. Similarly, given that a country was in recession during the previous period ($s_{t-1} = 1$), the probability that it remains in recession during this period ($s_t = 1$) is $p_{11}$, and the probability that the economy recovers and enters expansion this period ($s_t = 2$) is $p_{21} = 1 - p_{11}$.

Persistence is generated in the Markov process when the diagonal elements of the transition matrix are greater than the off-diagonal elements. Previous studies typically find the persistence probability of expansion, $p_{22}$, to be greater than the persistence probability of recession, $p_{11}$, coinciding with the observation that, on average, expansions are longer than recessions. For example, Hamilton (1989) found persistence probabilities for the United States of approximately 0.90 for expansions and 0.75 for recessions, implying expected durations of 10 quarters for expansions and 4 quarters for recessions, similar to those defined by the NBER.

Because we are interested in how U.S. output growth informs economic turning points of other nations, we extend Hamilton’s (1989) model to allow a foreign (U.S.) output growth rate to directly affect the evolution of the underlying economic state of other nations. We assume the Markov process is governed by TVTP, which are functions of exogenous covariates and the previous period’s state. In our case, we use the one-period lag of U.S. output growth, $y_{US,t-1}^{US}$, as the single covariate, which influences the switching process. The time-varying transition matrix in the two-state model is

$$P_t = \begin{bmatrix} p_{11,t} & p_{12,t} \\ p_{21,t} & p_{22,t} \end{bmatrix},$$

where $p_{ii,t} = \Pr[s_t = i | s_{t-1} = i; y_{US,t-1}^{US}]$. This allows for the transition probabilities to vary over time in response to changes in the U.S. output growth rate.
Here, $\alpha_j$ is the time-invariant parameter and $\beta_{ji}$ is the coefficient on lagged U.S. output growth, $y_{US}^{t-1}$. The FTP model is nested under the TVTP framework if the covariate has no effect under each state realization (i.e., $\beta_{ji} = 0$ for $i = 1,2$ and $j = 1,2$). Note that the time-invariant parameter $\alpha_{ji}$ and the coefficient $\beta_{ji}$ depend on both the previous state ($s_{t-1} = i$) and the potential current state ($s_t = j$), thereby reflecting the Markov property. Also, this parameterization allows U.S. output growth to have asymmetric effects since we assume the coefficient is state dependent (i.e., $\beta_{j1} \neq \beta_{j2}$ for $j = 1,2$ and $\beta_{i1} \neq \beta_{i2}$ for $i = 1,2$). To identify the transition parameters, we must normalize one of the state's transition parameters to be zero. For the two-state model, we use state 2: $\alpha_{2i} = 0$ and $\beta_{2i} = 0$ for $i = 1,2$.

For the general $K$-state model, the time-varying transition matrix is

$$P_t = \begin{bmatrix} p_{11,t} & p_{12,t} & \cdots & p_{1K,t} \\ p_{21,t} & p_{22,t} & \cdots & p_{2K,t} \\ \vdots & \vdots & \ddots & \vdots \\ p_{K1,t} & p_{K2,t} & \cdots & p_{KK,t} \end{bmatrix},$$

with TVTP

$$p_{ji,t} = \Pr[s_t = j|s_{t-1} = i, y_{US}^{t-1}] = \frac{\exp(\alpha_{ji} + \beta_{ji} y_{US}^{t-1})}{\sum_{k=1}^{K} \exp(\alpha_{ki} + \beta_{ki} y_{US}^{t-1})}.$$  

Determining the Effects of U.S. Output Growth

The effect of U.S. output growth on turning points of other countries appears to be summarized by the coefficient $\beta_{ji}$ in the transition equations. However, interpreting these coefficients in the logistic framework of TVTP is less straightforward than in a simple linear regression model. One way to assess the effect of U.S. output growth on the transition dynamics is by analyzing the marginal effect of a change in $y_{US}^{t-1}$ on each transition probability $p_{ji,t}$ for $j = 1,\ldots,K$. 
and $i = 1, \ldots, K$. We calculate the marginal effect of $y_{t-1}^{US}$ on $p_{ji,t}$ by taking the partial derivative of equation (2) with respect to $y_{t-1}^{US}$:

$$\frac{\partial p_{ji,t}}{\partial y_{t-1}^{US}} = p_{ji,t} (\beta_{ji} - \bar{\beta}),$$

where $\bar{\beta} = \sum_k p_{ki,t} \beta_{ki}$ is the probability-weighted mean of the coefficient across states.

In the two-state model, the marginal effect of a change in $y_{t-1}^{US}$ on the probability of recession ($s_i = 1$) simplifies to

$$\frac{\partial p_{ji,t}}{\partial y_{t-1}^{US}} = \beta_{ji} p_{ji,t} (1 - p_{ji,t}),$$

which depends on the previous period’s state. Determining the sign of this marginal effect is straightforward because it is irrespective of the value of $y_{t-1}^{US}$ and therefore time invariant. If $\beta_{1i} < \beta_{2i} = 0$, then the probability of a recession (expansion) next period falls (rises) as lagged U.S. output growth rises. We expect to find this relationship for countries whose economies tend to move with the U.S. economy. Conversely, if $\beta_{1i} > \beta_{2i} = 0$, then the probability of a recession (expansion) next period rises (falls) as lagged U.S. output growth rises. We expect to find this relationship for countries whose economies move in the opposite direction (decouple) from the U.S. economy. If $\beta_{1i} = \beta_{2i} = 0$, then the marginal effect is zero and lagged U.S. output growth does not influence the transition probabilities. Therefore, no relationship exists between U.S. output growth and economic turning points for the country under consideration.

Unlike the sign, the magnitude of the marginal effect in the two-state model is time varying because it depends on the value of $y_{t-1}^{US}$. For example, assume parameter values $\alpha_{11} = -1$ and $\beta_{11} = -1$ in a simple two-state version of our model ($K = 2$). First, consider the case where U.S. output growth is 2 standard deviations above its historical mean ($y_{t-1}^{US} = 2$). Then, the marginal effect of further changes in $y_{t-1}^{US}$ on the persistence probability of recession is $-0.05$. However, if U.S. output growth is relatively low at 2 standard deviations below its historical mean ($y_{t-1}^{US} = -2$), then the absolute magnitude of this marginal effect quadruples to $-0.20$. Thus, the current status of the U.S. economy informs not only the probability of recession in the country of interest but also the current degree of influence of U.S. output growth over this probability.

In the general $K$-state model, both the sign and magnitude of the marginal effects depend on the value of $y_{t-1}^{US}$. To fully assess the effect of U.S. output growth at different points in time, we calculate the marginal effects over a range of possible values of $y_{t-1}^{US}$.

**DATA AND ESTIMATION**

*Data*

We use the seasonally adjusted, annualized quarter-to-quarter growth rate of real GDP as our measure of economic activity growth ($y_t$) for each country. We use data from the Quarterly National Accounts database provided by the Organisation for Economic Co-operation and Development (OECD). The countries included in our sample are the G-7 counterparts of the
United States (Canada, France, Germany, Italy, Japan, and the United Kingdom) and Mexico, given its geographic proximity and economic relationship with the United States. Our time series covers 1960:Q2–2013:Q4 for Canada, Germany, Italy, Japan, and the United Kingdom; 1970:Q2–2013:Q4 for France; and 1980:Q2–2013:Q4 for Mexico. Table 1 provides summary statistics for our sample.

For the transition covariate, \( y_{US}^{t-1} \), we use the one-period lag of U.S. output growth from the OECD’s Quarterly National Accounts database for the period 1960:Q1–2013:Q3. To simplify the interpretation of the results, we standardize the time series of U.S. output growth to have zero mean and unit variance. Thus, \( y_{US}^{t-1} = 0 \) implies the United States is at its historical average growth rate—approximately 3.04 percent—over our sample period. Similarly, \( y_{US}^{t-1} = c \) means the United States is growing at \( c \) standard deviations away from its historical average growth rate. For example, \( y_{US}^{t-1} = 2 \) implies that U.S. output grew at 9.80 percent during the previous period since the standard deviation of U.S. output growth from 1960:Q1 to 2013:Q2 is approximately 3.38.

Figure 1 plots the time series of real GDP growth for a subset of our sample (Canada, Germany, and Japan). The gray bars indicate U.S. recession dates as defined by the NBER’s Business Cycle Dating Committee and are included only for reference. For each country, real GDP growth tends to fall during periods of U.S. recession, implying some connection between U.S. growth and other countries’ growth.

### Table 1

#### Sample Statistics

<table>
<thead>
<tr>
<th>Country</th>
<th>Coverage</th>
<th>Mean (( \tilde{y} ))</th>
<th>Variance (( \sigma_y^2 ))</th>
<th>Correlation with U.S. (( p_{xy} ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td>1960:Q2–2013:Q4</td>
<td>3.19</td>
<td>11.89</td>
<td>0.52</td>
</tr>
<tr>
<td>France</td>
<td>1970:Q2–2013:Q4</td>
<td>2.09</td>
<td>5.24</td>
<td>0.32</td>
</tr>
<tr>
<td>Germany</td>
<td>1960:Q2–2013:Q4</td>
<td>2.44</td>
<td>19.56</td>
<td>0.27</td>
</tr>
<tr>
<td>Italy</td>
<td>1960:Q2–2013:Q4</td>
<td>2.47</td>
<td>17.13</td>
<td>0.24</td>
</tr>
<tr>
<td>Japan</td>
<td>1960:Q2–2013:Q4</td>
<td>3.93</td>
<td>28.22</td>
<td>0.21</td>
</tr>
<tr>
<td>Mexico</td>
<td>1980:Q2–2013:Q4</td>
<td>2.39</td>
<td>28.53</td>
<td>0.26</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>1960:Q2–2013:Q4</td>
<td>2.45</td>
<td>15.43</td>
<td>0.26</td>
</tr>
</tbody>
</table>

Estimation

We estimate both the two- and three-state models using the Gibbs sampler, a Markov chain Monte Carlo (MCMC) algorithm used in a Bayesian environment. Rather than drawing from the full joint posterior distribution directly, the Gibbs sampler draws each of the four parameter blocks from their individual conditional posterior distribution given the draws for the other blocks. First, we partition the parameters and latent variables into four blocks: (i) the average growth rates, \( \mu = [\mu_1, \ldots, \mu_K] \); (ii) the error variance, \( \sigma_y^2 \); (iii) the transition probability param-
Figure 1
Real GDP Growth for Canada, Germany, and Japan

NOTE: The shaded bars indicate U.S. recessions as determined by the National Bureau of Economic Research.
SOURCE: Data from the OECD’s Quarterly National Accounts Database.
eters, \( \Gamma \); and (iv) the time series of the latent state variable, \( s = [s_1,\ldots,s_T]' \). We run the sampler for 100,000 iterations, discarding the first 50,000 to achieve convergence.

Tables 2 and 3 show the prior distributions for the parameters of the two- and three-state models, respectively. In each case, we use conjugate prior distributions. Following Kim and Nelson (1999), the steps to draw the average growth rate and error variance parameters are straightforward. The conditional posterior distribution for the vector of average growth rates, \( \mu \), is multivariate normal and the posterior for the error variance, \( \sigma^2 \), is inverse gamma.

The transition probability parameters can be rewritten as a difference random utility model (dRUM) as outlined by Frühwirth-Schnatter and Frühwirth (2010) and Kaufmann (2011). Under the dRUM, we assume each state has a continuous, latent utility value. Conditional on knowing the state at each point in time, the observed state is the one with the highest utility. The conditional posterior distribution of the transition parameter vector, \( \gamma_i \), is multivariate normal for each state \( i = 1,\ldots,K-1 \). The unobserved state variable is drawn using the filter from Hamilton (1989) with the smoothing algorithm from Kim (1994). For the general \( K \)-state model, we use the multistate extension of the filter as outlined by Kaufmann (2011).

Choosing between using two states (recession and expansion) and three states (recession, low-growth expansion, and high-growth expansion) is a model selection problem. We use the Bayesian information criterion (BIC) to choose which model is best suited for each country. BIC is calculated as

\[
BIC = -2\log \left[ L(\Theta, s, y, y^{US}) \right] + N \log(T),
\]

where \( N \) is the number of parameters in the model, \( T \) is the number of time-series observations, and \( L(\Theta, s, y, y^{US}) \) is the value of the likelihood function given model parameters \( \Theta = [\mu, \sigma^2, \alpha, \beta] \).

---

**Table 2**

Prior Distributions for the Two-State Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Prior Distribution</th>
<th>Hyperparameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu = [\mu_1, \mu_2]' )</td>
<td>( N(\mu_0, \sigma^2 I_2) )</td>
<td>( \mu_0 = [3, -3], M_0 = I_2 )</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>( \Gamma(\frac{nu}{2}, \frac{n\tau}{2}) )</td>
<td>( v_0 = 1, \tau_0 = 1 )</td>
</tr>
<tr>
<td>( \gamma = [\alpha_1, \alpha_2, \beta_1, \beta_2]' )</td>
<td>( N(g_0, G_0) )</td>
<td>( g_0 = 0, G_0 = 2I_4 )</td>
</tr>
</tbody>
</table>

**Table 3**

Prior Distributions for the Three-State Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Prior Distribution</th>
<th>Hyperparameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu = [\mu_1, \mu_2, \mu_3]' )</td>
<td>( N(\mu_0, \sigma^2 I_3) )</td>
<td>( \mu_0 = [-2, 2, 6], M_0 = I_2 )</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>( \Gamma(\frac{nu}{2}, \frac{n\tau}{2}) )</td>
<td>( v_0 = 1, \tau_0 = 1 )</td>
</tr>
<tr>
<td>( \gamma_k = [\alpha_{k1}, \alpha_{k2}, \alpha_{k3}, \beta_{k1}, \beta_{k2}, \beta_{k3}]' )</td>
<td>( N(g_0, G_0) )</td>
<td>( g_0 = 0, G_0 = 2I_6 )</td>
</tr>
</tbody>
</table>
the state vector $s$, and the data $y = [y_1, \ldots, y_T]$ and $y^{US} = [y^{US}_0, \ldots, y^{US}_{T-1}]$. The BIC accounts for the likelihood of the data while penalizing models with a large number of parameters. Raftery (1995) and Kass and Raftery (1995) show that the BIC approximates the Bayes factor of competing models; thus, it provides an adequate solution to our model selection problem. The BIC is calculated at each iteration of the Gibbs sampler, and the optimal model for each country is the one that minimizes the median BIC calculation.

**RESULTS**

Table 4 shows the model selection results for each country. The two-state model is preferred for Germany, Japan, and Mexico, while the three-state model is chosen for Canada, France, Italy, and the United Kingdom. These results suggest a more stable expansion output growth rate for the former countries, while the latter countries appear to have both low- and high-growth expansions.

Table 5 presents the estimated mean growth rate and variance parameters for each country. Germany, Japan, and Mexico each have much higher error variance than the other countries in the sample; it is possible this variance is caused by the lack of the third state in their optimal model to capture high-growth dynamics. The lack of two expansion states also explains the higher estimated mean expansionary growth rate for these countries since the model captures episodes of both high and low growth.

We discuss the remaining results in two subsections. The first outlines the estimated recession timing for each country across time. The second subsection assesses the ability of U.S. output growth to inform business cycle turning points for each country.

---

**Table 4**

Bayesian Information Criterion

<table>
<thead>
<tr>
<th>Country</th>
<th>Two-state model</th>
<th>Three-state model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td>1112.2</td>
<td><strong>1057.6</strong></td>
</tr>
<tr>
<td>France</td>
<td>769.1</td>
<td><strong>746.3</strong></td>
</tr>
<tr>
<td>Germany</td>
<td><strong>1238.0</strong></td>
<td>1259.9</td>
</tr>
<tr>
<td>Italy</td>
<td>1195.2</td>
<td><strong>1174.2</strong></td>
</tr>
<tr>
<td>Japan</td>
<td><strong>1026.3</strong></td>
<td>1048.6</td>
</tr>
<tr>
<td>Mexico</td>
<td><strong>806.8</strong></td>
<td>855.9</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>1166.1</td>
<td><strong>1161.2</strong></td>
</tr>
</tbody>
</table>

NOTE: Bold type indicates the optimal model that minimizes BIC.
Timing of Business Cycle Phases

Figure 2 presents the probability implied by our model that a country is in a state of recession at each period in our sample. In technical terms, these are the posterior probabilities of recession, $\Pr[s_t = 1|\Omega_t]$, for each country conditional on $\Omega_t$, the information at time $t$. For each $t$, $\Pr[s_t = 1|\Omega_t]$ is the percentage of Gibbs iterations for which a recession state is drawn at each period. Although all countries in our sample have experienced some similar recessions (e.g., the first oil crisis of the mid-1970s and the Great Recession of 2007-09), there are substantive differences in the timing of countries entering recessions and the durations of recessions. For example, we find that most countries entered recession after the United States had already begun the Great Recession of 2007-09. Although some countries (e.g., Canada, Mexico, and the United Kingdom) exited this recession with the United States, others (e.g., Italy and Japan) experienced lasting effects of the global downturn, leading to a “double-dip” recession.

For completeness, we plot the posterior probability of expansion in Figure 3. Countries following the two-state model (Germany, Japan, and Mexico) have a single expansion state and therefore a single posterior probability of expansion, whereas countries following the three-state model (Canada, France, Italy, and the United Kingdom) have two expansion states (low and high growth). For the latter countries, we include the posterior probabilities of the low-growth expansion state in Figure 3 and separately plot the posterior probabilities for the high-growth state in Figure 4.

Consistent with the empirical literature on business cycles, we find the expansion state(s) are highly persistent with longer average duration(s) than the recession state(s). The high-growth expansion state accounts for periods of relatively high growth prior to 1985, the beginning of the period known as the Great Moderation. For France, the high-growth expansion state also captures two notable economic periods: the movement away from dirigisme in the late 1980s and the beginning of euro integration in the late 1990s.

Table 5

Estimates for the Average Growth Rate and Variance Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Canada</th>
<th>France</th>
<th>Germany</th>
<th>Italy</th>
<th>Japan</th>
<th>Mexico</th>
<th>United Kingdom</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1$</td>
<td>-2.90</td>
<td>-3.25</td>
<td>-2.86</td>
<td>-2.57</td>
<td>-3.74</td>
<td>-4.60</td>
<td>-2.99</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>2.66</td>
<td>1.49</td>
<td>3.19</td>
<td>1.69</td>
<td>3.21</td>
<td>3.66</td>
<td>2.82</td>
</tr>
<tr>
<td>$\mu_3$</td>
<td>6.78</td>
<td>4.10</td>
<td>—</td>
<td>6.60</td>
<td>—</td>
<td>—</td>
<td>8.15</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>5.19</td>
<td>2.44</td>
<td>15.53</td>
<td>8.97</td>
<td>16.22</td>
<td>17.21</td>
<td>8.46</td>
</tr>
</tbody>
</table>

NOTE: The table shows median posterior draws for the state-dependent growth rates, $u_i$, and the variance, $\sigma^2$. 
Figure 2
Posterior Recession Probabilities

NOTE: The posterior recession probabilities for each country (y-axes) are calculated as the percentage of MCMC draws for which a recession is drawn \( s_t = 1 \). The shaded bars indicate U.S. recessions as determined by the National Bureau of Economic Research.
Figure 3
Posterior Expansion Probabilities

NOTE: The posterior expansion probabilities for each country (y-axes) are calculated as the percentage of MCMC draws for which an expansion is drawn (s = 2). For countries following the three-state model (Canada, France, Italy, and the United Kingdom), these are the posterior probabilities of the low-growth expansion state. The shaded bars indicate U.S. recessions as determined by the National Bureau of Economic Research.
Figure 4
Posterior High-Growth Expansion Probabilities

NOTE: The posterior high-growth expansion probabilities (y-axes) for countries following the three-state model (Canada, France, Italy, and the United Kingdom) are calculated as the percentage of MCMC draws for which a high-growth expansion is drawn ($s = 3$). The shaded bars indicate U.S. recessions as determined by the National Bureau of Economic Research.
Does U.S. Output Growth Drive Business Cycles?

The focus of this article is determining whether U.S. output growth informs economic turning points of other nations. In our modeling framework, this relationship is captured in the transition dynamics of the state variable. Table 6 displays the median posterior draws for the transition probability parameters for all countries in our sample. As noted in the section “Determining the Effects of U.S. Output Growth,” the coefficients $\beta_{ji}$ in the transition equations suggest how U.S. output growth influences the state dynamics of the country of interest. They are not, however, the sole determinants of the (marginal) effect of a change in lagged U.S. output growth on the transition probabilities on the business cycle of a given country. Because the marginal effects depend on both the value of lagged U.S. output growth $y_{US,t-1}$ and the previous state of the economy $s_{t-1}$, we calculate them across all possible combinations of $s_{t-1}$ and $y_{US,t-1}$. We do this for each iteration of the Gibbs sampler, thereby constructing the posterior distribution for each of the marginal effects.

Figures 5 through 11 display the marginal effect of a change in lagged U.S. output growth on each of the transition probabilities. The horizontal axis for each figure reflects different values for U.S. output growth, from –4 to +4 standard deviations from its historical average. The vertical axis plots the marginal effect of a change in U.S. output growth on the respective transition probability conditional on the value for $y_{US,t-1}$ and the previous state $s_{t-1}$. In each figure, the blue line represents the posterior median of the marginal effect, and the shaded region represents the 68 percent coverage of the posterior distribution.

Table 6

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Canada</th>
<th>France</th>
<th>Germany</th>
<th>Italy</th>
<th>Japan</th>
<th>Mexico</th>
<th>United Kingdom</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{11,t}$</td>
<td>$\alpha_{11}$</td>
<td>–0.17</td>
<td>0.76</td>
<td>–1.17</td>
<td>1.12</td>
<td>–0.49</td>
<td>0.20</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>–1.35</td>
<td>–0.81</td>
<td>–1.12</td>
<td>–0.52</td>
<td>–0.25</td>
<td>–0.91</td>
<td>–0.81</td>
</tr>
<tr>
<td>$p_{12,t}$</td>
<td>$\alpha_{12}$</td>
<td>–1.10</td>
<td>–1.46</td>
<td>–2.73</td>
<td>–0.59</td>
<td>–2.91</td>
<td>–2.78</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>–2.49</td>
<td>–1.00</td>
<td>–1.40</td>
<td>–1.36</td>
<td>–0.29</td>
<td>–0.50</td>
<td>–1.27</td>
</tr>
<tr>
<td>$p_{13,t}$</td>
<td>$\alpha_{13}$</td>
<td>–2.23</td>
<td>–3.13</td>
<td>—</td>
<td>—</td>
<td>–2.35</td>
<td>—</td>
</tr>
<tr>
<td>$\beta_{13}$</td>
<td>–1.44</td>
<td>–0.96</td>
<td>—</td>
<td>—</td>
<td>–0.35</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$p_{21,t}$</td>
<td>$\alpha_{21}$</td>
<td>–0.73</td>
<td>0.09</td>
<td>—</td>
<td>–0.18</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>–0.46</td>
<td>–0.87</td>
<td>—</td>
<td>—</td>
<td>0.39</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$p_{22,t}$</td>
<td>$\alpha_{22}$</td>
<td>2.51</td>
<td>2.94</td>
<td>—</td>
<td>—</td>
<td>3.58</td>
<td>—</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>–1.53</td>
<td>–0.45</td>
<td>—</td>
<td>—</td>
<td>–0.54</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$p_{23,t}$</td>
<td>$\alpha_{23}$</td>
<td>0.06</td>
<td>–2.25</td>
<td>—</td>
<td>—</td>
<td>–2.30</td>
<td>—</td>
</tr>
<tr>
<td>$\beta_{23}$</td>
<td>–1.03</td>
<td>–0.31</td>
<td>—</td>
<td>—</td>
<td>–0.65</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

NOTE: Median posterior draws for the parameters governing the transition probabilities, $p_{ji,t}=Pr(s_t=j|s_{t-1}=i,y_{US,t-1})$; $\alpha_{ji}$ captures the time-invariant portion of the transition probability; and $\beta_{ji}$ is the coefficient on lagged U.S. output growth. Bold values indicate that 0 lies outside the 68 percent posterior coverage.
**Figure 5**

**Marginal Effect of a Change in U.S. Output on the Transition Probabilities for Canada**

NOTE: The blue line represents the posterior median of the marginal effect of a change in U.S. output growth on the transition probability given the values for lagged U.S. output growth ($y_{US,t-1}$) and the past state ($s_{t-1}$). The shaded regions reflect the 68 percent coverage of the posterior distribution.
Figure 6
Marginal Effect of a Change in U.S. Output on the Transition Probabilities for France

NOTE: The blue line represents the posterior median of the marginal effect of a change in U.S. output growth on the transition probability given the values for lagged U.S. output growth ($y_{US,t-1}$) and the past state ($s_{t-1}$). The shaded regions reflect the 68 percent coverage of the posterior distribution.
Figure 7
Marginal Effect of a Change in U.S. Output on the Transition Probabilities for Germany

NOTE: The blue line represents the posterior median of the marginal effect of a change in U.S. output growth on the transition probability given the values for lagged U.S. output growth ($y_{US,t-1}$) and the past state ($s_{t-1}$). The shaded regions reflect the 68 percent coverage of the posterior distribution.
Figure 8
Marginal Effect of a Change in U.S. Output on the Transition Probabilities for Italy

<table>
<thead>
<tr>
<th>State Transition</th>
<th>Marginal Effect on Transition Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recession Last Period ((s_{t-1} = 1))</td>
<td>Low-Growth Expansion Past Period ((s_{t-1} = 2))</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Marginal Effect on (Pr(y_{t-1}^{US}))</td>
<td>Marginal Effect on (Pr(y_{t-1}^{IS}))</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NOTE: The blue line represents the posterior median of the marginal effect of a change in U.S. output growth on the transition probability given the values for lagged U.S. output growth ((y_{t-1}^{US})) and the past state ((s_{t-1})). The shaded regions reflect the 68 percent coverage of the posterior distribution.</td>
<td></td>
</tr>
</tbody>
</table>
Figure 9
Marginal Effect of a Change in U.S. Output on the Transition Probabilities for Japan

NOTE: The blue line represents the posterior median of the marginal effect of a change in U.S. output growth on the transition probability given the values for lagged U.S. output growth ($y_{t-1}^{US}$) and the past state ($s_{t-1}$). The shaded regions reflect the 68 percent coverage of the posterior distribution.
Figure 10

Marginal Effect of a Change in U.S. Output on the Transition Probabilities for Mexico

NOTE: The blue line represents the posterior median of the marginal effect of a change in U.S. output growth on the transition probability given the values for lagged U.S. output growth ($y_{US,t-1}$) and the past state ($s_{t-1}$). The shaded regions reflect the 68 percent coverage of the posterior distribution.
Marginal Effect of a Change in U.S. Output on the Transition Probabilities for the United Kingdom

NOTE: The blue line represents the posterior median of the marginal effect of a change in U.S. output growth on the transition probability given the values for lagged U.S. output growth ($y^{US}_{t-1}$) and the past state ($s_{t-1}$). The shaded regions reflect the 68 percent coverage of the posterior distribution.
A positive marginal effect implies that an increase in lagged U.S. output growth increases the respective transition probability \( p_{ji,t} = \Pr(s_t = j|s_{t-1} = i, y_{US,t-1}) \). Conversely, a negative marginal effect implies that an increase in lagged U.S. output growth decreases the respective transition probability. That is, for countries whose economies move with the U.S. economy, we expect to find a positive (negative) marginal effect of \( y_{US,t-1} \) on the probability of transitioning to an expansion (recession) and the duration of expansion (recession). For countries whose economies move opposite to the U.S. economy, we expect to find a negative (positive) marginal effect of \( y_{US,t-1} \) on the probability of transitioning to an expansion (recession) and the duration of expansion (recession).

For each country, we assess the ability of U.S. output growth to inform (i) the timing of entering a recession, (ii) the duration of a recession, and (iii) transitions between states of low- and high-growth expansion (for countries following the three-state model). We assess the first dynamic by examining the marginal effect of U.S. output growth on the transition probability from expansion \((s_{t-1} = 2 \text{ or } 3)\) to recession \((s_{t-1} = 1)\), so the relevant transition probabilities are \( p_{12,t} \) and \( p_{13,t} \). For recession duration, we determine whether U.S. output influences the transition probability of staying in recession this period \((s_{t-1} = 1)\) given that the economy was in recession during the previous period \((s_{t-1} = 1)\) with relevant transition probability \( p_{11,t} \). We analyze the last aspect by examining both the persistence probability of both low-expansion \((p_{22,t})\) and high-expansion \((p_{33,t})\) states in addition to the transition probabilities between the two expansion states \((p_{23,t})\) and \((p_{32,t})\).

The three countries most influenced by U.S. output growth are Canada, Germany, and the United Kingdom. For these countries, lagged U.S. output growth influences both the timing of entering a recession and the duration of a recession. The results show that the economies of each of these countries move with the U.S. economy: Higher U.S. output growth implies a lower probability of recession, and lower output growth implies a higher probability of recession \((\uparrow y_{US,t-1} \Rightarrow \downarrow p_{11,t}, \uparrow p_{22,t} \text{ for all } i)\). Figure 7 presents the marginal effects for Germany, which follows the simpler two-state model. For Germany, the marginal effect of U.S. output growth is largest (in absolute terms) at low levels of \( y_{US,t-1} \), or when the United States is likely in a state of recession. Therefore, when the U.S. economy is in dire circumstances (as signaled by low output growth), Germany is more susceptible to any further movements in U.S. output relative to more “normal” economic times.

In addition to informing the timing and duration of recessions, U.S. output growth also influences the transition dynamics of low- and high-growth expansion for Canada (see Figure 5). When U.S. growth is relatively low (i.e., below its historical mean), increases in U.S. output growth imply a higher persistence of low-growth expansion \((\uparrow p_{22,t})\). However, when U.S. growth is relatively high (i.e., above its historical mean), increases in U.S. output growth (i) decrease the duration of low-growth expansion and (ii) increase both the probability of transitioning to high-growth expansion \((\uparrow p_{32,t})\) and the persistence probability of high-growth expansion \((\uparrow p_{33,t})\). This result reflects the strong economic relationship between Canada and the United States since it informs not only the timing of recessions but also the timing of varying degrees of expansion.
For Mexico, lagged U.S. output growth informs the duration of a recession but not the timing of entering a recession. When U.S. output growth falls, the persistence probability of recession in Mexico rises ($\uparrow p_{11}$), implying a longer expected duration of recession. The lack of influence of U.S. output growth on the timing of Mexico entering a recession could be due to the fact that Mexico experienced idiosyncratic recessions unrelated to the United States (e.g., the 1994 Mexican peso crisis), which tended to be shorter than coincident recessions with the United States (e.g., the recession of the early 1980s and the Great Recession of 2007-09).

The results for France, Italy, and Japan suggest that lagged U.S. output growth does not influence the timing or duration of recessions for these countries. For France and Italy, increases in U.S. output growth increase the persistence probability of high-growth expansion ($\uparrow p_{33}$) but only at low levels of U.S. output growth.

Recent studies on business cycle synchronization offer two possible explanations for our results: stage of development and common language. Regarding the first explanation, Kose, Otrook, and Prasad (2012) find that emerging market economies and advanced economies have decoupled during the globalization period, but the economies of countries within each respective group have converged. This finding is consistent with our result that the United States is more informative for the business cycles of advanced countries such as Canada, Germany, and the United Kingdom and less so for the developing country in our sample, Mexico.

Another plausible explanation is that countries with a common language tend to have similar business cycles. We find that U.S. output growth informs the business cycles for each of the countries in our sample whose de facto or official language is English.

**CONCLUSION**

In this article, we assessed whether the U.S. economy drives business cycle turning points of other nations. We extended the nonlinear business cycle model of Hamilton (1989) to allow U.S. output growth to influence the probability of a country moving between states of expansion and recession. We found that the United States does inform the timing and duration of recessions for Canada, Germany, the United Kingdom, and, to a lesser extent, Mexico. Additionally, we found no informative relationship between U.S. output growth and the business cycles of France, Italy, and Japan.

It is important to keep in mind that our results suggest only that the U.S. economy does not appear to lead the economies of France, Italy, and Japan. If the business cycles in these countries react intraquarterly to fluctuations in U.S. output, the leading relationship of the United States would show up as a false negative in the estimation. Further, if a common world shock affects the United States before other countries, the result might be a false positive. However, our analysis provides a framework for approaching the question of Granger causality across business cycles.
NOTES

1 For analysis of the specific mechanisms (trade openness, financial market linkages, and so on) by which the United States transmits shocks to the rest of the world, see Calvo, Leiderman, and Reinhart (1993); Kose and Yi (2001); Uribe and Yue (2006); Mackowiak (2007); Edwards (2010); Bayoumi and Bui (2010); and Kazi, Wagan, and Akbar (2013).

2 In this case, we assume that the foreign output growth rate is exogenous and unaffected by the domestic regime.

3 Note that we cannot infer causality of the business cycle in the structural sense, but rather we assess if U.S. output acts as an informative indicator of other countries’ turning points. Therefore, for the countries for which our model indicates that U.S. output growth is not a significant indicator, this assessment does not imply a lack of structural mechanisms that propagate shocks between the two nations.

4 We consider values for $y_{US,t}$ between $-4$ standard deviations and $+4$ standard deviations from its historical mean. This corresponds to a range of $-10.5$ to $16.6$, which includes the historical minimum ($-8.7$) and maximum ($15.3$) values of U.S. output growth.

5 See Artis, Chouliarakis, and Harischandra (2011); Francis, Owyang, and Savascin (2012); and Ductor and Leiva-Leon (2014).

REFERENCES


How Does Informal Employment Affect the Design of Unemployment Insurance and Employment Protection?

Emilio Espino and Juan M. Sánchez

The authors use a simple model to study the optimal design of unemployment insurance and employment protection. Workers are risk averse and face the possibility of unemployment. Firms are risk neutral and face random shocks to productivity. Workers can participate in a shadow economy, or informal sector. The model yields several lessons. First, countries should encourage formal employment to address the issue of informal employment. In extreme cases, such encouragement translates into high severance payments and negative payroll taxes. Along these same lines, unemployment payments cannot be too large. In fact, when the risk of informality is extreme, the authors find that unemployment benefits should be negative, which is (in effect) a positive tax on the lack of formal employment. (JEL D82, H55, I38, J65)


ow does the risk of informality (hereafter, informality) affect the optimal design of unemployment insurance and employment protection? To study this question, we extend the simple model of unemployment insurance developed by Blanchard and Tirole (2008) to add informality. In particular, we consider an extended model in which individuals can both work in the informal sector (the shadow economy), where employment is not observed by the government, and simultaneously claim unemployment insurance payments.

It is important to account for the possibility of informality because in many economies a large share of total economic activity takes place in the informal sector. According to the work of Schneider and Enste (2000) and Schneider, Buehn, and Montenegro (2010), the size of the informal economy is nonnegligible for developed countries and very large for developing ones (Table 1). For example, the U.S. shadow economy during the 2000s is about 9 percent of gross domestic product (GDP). The size of the informal sector is much larger for less-developed countries. For instance, during the 2000s informality as a share of GDP was about 40 percent in Brazil.
Studying informality and unemployment insurance is relevant because “unreported wages” are the most common form of informality (Schneider and Enste, 2000). If employment in the informal sector is unreported, individuals can receive labor earnings and collect unemployment benefits at the same time. In fact, according to Fuller, Ravikumar, and Zhang (2013), unemployment insurance overpayments in the United States amounted to $3.3 billion in 2011. About 65 percent of such overpayment cases were related to individuals collecting unemployment insurance despite being employed.

The unemployment insurance model presented here is extremely simple. In particular, the model is static, since all decisions are made in a single period. Workers are risk averse, so they need insurance, but firms are risk neutral. Unemployment exists because productivity is random and firms are better off firing low-productivity workers. Informality is modeled as in Álvarez-Parra and Sánchez (2009). Individuals have the option of working in an informal sector at a low wage. However, unlike Álvarez-Parra and Sánchez (2009), here we assume that the optimal policy must prevent informality. Thus, we assume the costs of informality (not modeled) are high such that in equilibrium zero informality is optimal; we discuss this point in more detail later.

We use the model to study the optimal design of unemployment insurance and employment protection under the risk of informality. We focus on the implications of the risk of

Table 1
Size of the Shadow Economy in the 2000s: Selected Economies

<table>
<thead>
<tr>
<th>Country</th>
<th>Percent of GDP</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>8.6</td>
</tr>
<tr>
<td>Japan</td>
<td>10.8</td>
</tr>
<tr>
<td>New Zealand</td>
<td>12.4</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>12.5</td>
</tr>
<tr>
<td>Australia</td>
<td>13.8</td>
</tr>
<tr>
<td>France</td>
<td>14.9</td>
</tr>
<tr>
<td>Canada</td>
<td>15.6</td>
</tr>
<tr>
<td>Germany</td>
<td>15.8</td>
</tr>
<tr>
<td>Sweden</td>
<td>18.8</td>
</tr>
<tr>
<td>Portugal</td>
<td>22.6</td>
</tr>
<tr>
<td>Spain</td>
<td>22.8</td>
</tr>
<tr>
<td>Italy</td>
<td>27.0</td>
</tr>
<tr>
<td>Ecuador</td>
<td>33.4</td>
</tr>
<tr>
<td>Brazil</td>
<td>39.8</td>
</tr>
<tr>
<td>Zambia</td>
<td>49.9</td>
</tr>
<tr>
<td>Bolivia</td>
<td>66.4</td>
</tr>
</tbody>
</table>

SOURCE: Schneider, Buehn, and Montenegro (2010).
informality on the tax structure that implements the optimally designed unemployment insurance and employment protection. Despite its simplicity, the framework provides some sharp insights. First, in the economy with the risk of informality, low-productivity workers who would be unemployed in the formal economy without this risk are encouraged to work in the formal sector. Along these same lines, when the risk of informality is extreme—in the sense that the wage paid in the informal sector is high—the government may want to establish high severance payments. Similarly, in such situations, the government must set negative payroll taxes to make employment in the formal sector attractive. Another lesson from our framework is that unemployment payments cannot be too generous in economies with labor informality because unemployed workers could potentially claim unemployment benefits in the formal sector and work in the informal sector. In fact, if the risk of informality is extreme, unemployment benefits must be negative, which can be interpreted as a tax on informality.

The literature studying the optimal design of unemployment insurance in dynamic models is large and began with the seminal work of Hopenhayn and Nicolini (1997). The problem is modeled as a repeated principal-agent problem involving a risk-averse agent (the worker) and a risk-neutral principal (the government) that cannot monitor the agent’s search effort. These authors emphasize that unemployment payments should depend on the entire employment history of individuals. Their main finding is that unemployment payments should decrease during the unemployment spell and post-reemployment wage taxes should increase with the duration of unemployment.

Hopenhayn and Nicolini’s (1997) study, however, assumes that all employment takes place in the formal sector (employment is observable). In a related paper, Álvarez-Parra and Sánchez (2009) incorporate an informal sector into Hopenhayn and Nicolini’s (1997) framework. The key finding of Álvarez-Parra and Sánchez is that unemployment must decline abruptly to zero after several months. This design provides incentives for workers to concentrate on their job search during the first month of unemployment (when payments are being received). If individuals do not find a job during this period, they will spend most of their time working in the informal sector.

Our analysis is organized as follows. The following section presents the benchmark economy and characterizes the optimal design of unemployment insurance and employment protection and its implementation. In a separate section, we introduce the risk of informality into the model and study its impact on the optimal design and implementation. We then present the results of solving the model numerically. This section is the key to understanding the role of key parameters. The final section presents conclusions and policy implications.

**THE BENCHMARK ECONOMY**

The benchmark economy is used to illustrate the outcome with no informal sector or private information. The economy is composed of a continuum of mass 1 of ex ante identical workers. Workers are risk averse, with utility function $U$. Individuals who do not have a job can produce $b$ units of the consumption good at home.
On the other side, there is a continuum of mass 1 of ex ante identical entrepreneurs. Entrepreneurs are risk neutral, and each entrepreneur can start and run a firm. There is a fixed cost $I$ of creating a firm, which is the same for all entrepreneurs. If a firm is created, a worker is hired, and the productivity of the match is then revealed. Productivity is given by $y$ from the cumulative distribution function $G(y)$ with density $g(y) > 0$ on $[0,1]$. It is useful to define $Y(x) = \int_0^x ydG(y)$ as the output produced in the formal sector. Note that $Y''(x) < 0$, and we assume that $Y''(x) < 0$.

The firm, but not the worker (or, for that matter, third parties such as an insurance company or the state) observes $y$. Then the firm must take a decision: It can either retain the worker and produce or it can lay off the worker, who then becomes unemployed. Realizations of productivity matches are i.i.d across firms, and there is no aggregate risk.

Finally, there is a government that can levy taxes and provide unemployment insurance. This implies the following sequence of events:

- firms are created;
- workers and firms are matched;
- shocks are observed;
- production and employment decisions are taken; and
- taxes are collected and transfers are distributed.

**Optimal Allocation under Full Information**

The full-information allocation can be characterized as defining a threshold level of productivity below which workers are laid off. Call this threshold $y$. Let $w$ be the payment to workers who remain employed and $s$ be the payment to workers who are laid off and become unemployed.

The optimal allocation maximizes the expected utility of the representative agent subject to the economy’s resource constraint; that is,

$$\max_{\{w,s,y\}} G(y)U(b+s) + (1-G(y))U(w),$$

subject to

$$\int_y yG'(y)dy = G(y)s + (1-G(y))w + I.$$

The optimal design, $(w^*, s^*, y^*)$, dictates that consumption by employed and unemployed workers is the same; that is, $b + s^* = w^*$. This implies that the government provides full insurance to the representative risk-averse worker. The optimal design also dictates that the threshold of productivity, $\bar{y}$, is such that all workers with a productivity match higher than what they can produce at home, $y \geq b$, will work. All other workers, on the other hand, are laid off and become unemployed. This translates into $\bar{y}^* = b$.

We can use the resource constraint to write

$$w^* = Y(b) + bG(b) + I,$$

and so it defines the consumption (or net wage) of the employed worker, $w^*$. 
**Implementation of the Full-Information Optimal Allocation**

The implementation can be analyzed in three stages. At stage I, the government announces the payroll tax paid by the firm for each employed worker, $\tau$; the severance tax paid by the firm for each laid-off worker, $f$; and the unemployment benefits provided directly by the government, $s$. At stage II, the firm and the worker are matched and sign a contract that defines the noncontingent wage if the worker is not fired, $w$, and the productivity cutoff level, $\bar{y}$. At stage III, the firm (and not the worker) observes the realization of the productivity match, $y$, and decides to either retain the worker and pay wage ($w + \tau$) or fire the worker and pay layoff costs, $f$.

As usual, it is easier to solve the problem by working backwards. Given ($w^*, s^*, \bar{y}^*$), the government tries to design taxes ($\tau, f$) such that the optimal allocation studied earlier can be decentralized. Note that stages II and III define the optimal behavior of the representative firm.

**Stage III: Productivity Cutoff Level.** For the marginal worker with productivity $\bar{y}^*$, a firm must be indifferent between retaining him and firing him; that is,

$$\bar{y}^* - (w^* + \tau) = -f.$$

**Stage II: Free Entry Condition.** Ex ante, the representative firm is indifferent between starting the firm at cost $I$ and not starting the firm. As the entrepreneurs operating the firms are risk neutral, the corresponding ex ante zero-profit condition reads

$$I = \int_{\bar{y}}^{\infty} yG'(y)dy - G(\bar{y})f - (1-G(\bar{y}))(w^* + \tau).$$

**Stage I: The Government’s Budget Constraint.** The government sets taxes such that its budget constraint is satisfied; that is,

$$(1-G(\bar{y}))(\tau) = G(\bar{y})(-f + s^*).$$

Note that the government’s budget constraint coupled with the feasibility condition implies that the free entry condition (stage II) is satisfied. Hence, the conditions in stages I and III define the optimal ($\tau^*, f^*$) given ($w^*, s^*, \bar{y}^*$).

Importantly, as optimality dictates that (i) workers with productivity above the level of home production will produce $\bar{y}^* = b$ and (ii) the consumption of workers and unemployed individuals is the same, achieving full insurance—that is, $w^* = b + s^*$. Moreover, the conditions in stages I and III imply that $f^* = s^*$ and $\tau^* = 0$. That is, the contribution rate—the ratio of layoff taxes to unemployment benefits ($f^*/s^*$)—is 1. Payroll taxes are not used to make the firm fully internalize the cost of firing a worker.\(^4\)

**THE ECONOMY WITH RISK OF INFORMALITY**

Consider an extended version of the original setting that includes an informal labor market. That is, once a worker is matched with a firm,\(^5\) the worker can choose to move to work in the informal sector before the productivity draw has been revealed.
The employment status of an informal worker is unobservable and, in particular, cannot be distinguished a priori from unemployment. This means there is no technology available to detect employment in the informal sector and, consequently, individuals employed in the informal sector can pretend they are unemployed and collect unemployment benefits. Finally, if the worker is fired, he cannot move to the informal sector immediately. The wage in the informal sector is denoted by $w_I$. Importantly, we assume that $w_I > b$.

Note that in our framework it is not immediately clear that eliminating informality is necessarily optimal. Actually, informality is included in the optimal allocation studied by Álvarez-Parra and Sánchez (2009). Here, we assume it is desirable not to have workers in the informal sector for reasons beyond the scope of this article (crime, terrorism, and so on). In particular, the assumption is that informality generates a social cost that it is always larger than any benefit it may generate. This assumption is formalized with a new constraint. Given $(w, s)$, this ex ante incentive compatibility constraint is

$$\max_{(w, s)} G(y) U(w) + (1-G(y)) U(b+s) \geq U(w_I+s),$$

which be rewritten as

$$(1-G(y)) (U(w) - U(b+s)) \geq U(w_I+s) - U(b+s).$$

**Optimal Allocation with Risk of Informality**

Now we proceed to characterize the constrained optimal allocation in this setting. Notice that under these assumptions, incentives are provided so that no worker decides to work in the informal sector; thus, the optimal design must solve

$$\max_{(w, s, \pi)} G(y) U(w) + (1-G(y)) U(b+s),$$

subject to

$$\int_0^1 y G'(y) dy = G(y) s + (1-G(y)) w + I,$$

$$(1-G(y)) (U(w) - U(b+s)) \geq U(w_I+s) - U(b+s).$$

Let $\lambda$ and $\chi$ denote the Lagrange multipliers corresponding to the feasibility and the incentive compatibility constraints, respectively.

From the first-order conditions (FOCs) we obtain that

$$U'(b+s) = U'(w^*) + \frac{\chi}{1+\chi} U'(w_I+s).$$

Therefore, as $\lambda^* > 0$, this implies there is no full insurance because $U$ is concave; that is, $w^* > b + s$. The risk of labor market informality limits the provision of risk sharing as the government must limit the amount of unemployment benefits to reduce the value of informality.
On the other hand, observe that the FOCs also imply that

\[
\bar{y}' = b - \left[ ((b + s') - w') - \left( \frac{1}{U'(w')} \right) \left( U(b + s') - U(w') \right) \right].
\]

Therefore, since \( U \) is strictly concave and \( w' > b + s' \), it follows that the term in brackets is positive and so \( \bar{y}' < b \); that is, there is more employment in the allocation with risk of informality than in the allocation without risk of informality.

The intuition of this result can be grasped as follows. Under the risk of informality, the government wants not only to decrease the value of an informal job but also to increase the value of a formal job. Therefore, reducing \( \bar{y}' \) increases the likelihood that a worker will keep his formal job and be paid a salary \( w' > b + s' \). This is the main effect of the risk of informality on the optimal design of unemployment insurance and employment protection: Less-productive formal sector matches are preserved to provide incentives and thus make the informal sector less attractive.

**Implementation of the Informality Risk Optimal Allocation**

Given \( (w^*, s^*, \bar{y}^*) \), the government tries to design taxes \((\tau, f)\) such that the (constrained) optimal allocation studied earlier can be decentralized. Again, we solve the problem backwards. Recall that stages II and III define the optimal behavior of the representative firm.

**Stage III: Productivity Cutoff Level.** For the marginal worker with productivity \( \bar{y}' \), a firm must be indifferent between retaining him and firing him and so the government must set \((\tau, f)\) such that

\[
\bar{y}' = (w' + \tau) = -f.
\]

**Stage II: Free Entry Condition.** Ex ante, the representative firm must be indifferent between starting the firm at cost \( I \) and not starting the firm in an equilibrium where no worker moves to the informal sector because of the incentive compatibility constraint. So,

\[
I = \int_{\bar{y}'}^{\infty} yg'(y)dy - G(\bar{y}')f - (1 - G(\bar{y}'))(w' + \tau).
\]

**Stage I: The Government’s Budget Constraint.** Taxes must satisfy the government’s budget constraint such that

\[
(1 - G(\bar{y}'))\tau = G(\bar{y}'(-(f + s')).
\]

As before, the government’s budget constraint coupled with the feasibility condition implies that the free entry condition is satisfied and so equations in stages I and III determine the optimal taxes \((\tau^*, f^*)\). The combination of taxes that implements the optimally designed system is \( \tau^* = G(\bar{y}')(s + \bar{y}' - w') \) and \( f^* = s' + \tau' - \bar{y}' \).

The risk of informality implies that formal jobs should not be taxed but rather should be subsidized. To see this, note that if \( \tau^* \geq 0 \), then the government’s budget constraint implies that \( f^* \leq s' \) and so \( \tau^* - f^* \geq s' \). The condition derived above implies that
and therefore $b + s' \geq w'$. But this contradicts the fact that full insurance is not attained as long as $\lambda^* > 0$. Consequently, this implies that $\tau^* < 0$. Importantly, it follows from the government’s budget constraint that $f^* \geq s^*$. Thus, to provide incentives for firms to create more formal jobs, the government makes the contribution ratio larger than 1. That is, to finance subsidies, firms must pay more to fire a worker than what that fired worker will receive from unemployment.

**NUMERICAL EXERCISES**

In this section, we perform numerical exercises that add to the understanding of the implications of informality for the design of optimal policies. Several assumptions must be made about functional forms and parameter values as follows: The utility function is logarithmic, and the distribution of the productivity, $y$, is $N(1, \sigma^2)$ truncated between 0 and 1.

We solve for two economies. The first is the benchmark economy, referred to as *No informality*, which is basically the model developed by Blanchard and Tirole (2008). The second is the benchmark economy extended to allow for an informal sector as described in the previous section; we refer to this case as *Risk of informality*.

All figures display the optimal policy for different values of the informal wage-to-home production ratio, $w_i/b$. Recall that working in the informal sector requires workers to relinquish the opportunity to work in the formal sector. Thus, if $w_i/b < 1$, there is actually no risk of informality because its return is too low. Thus, we consider $w_i/b \in [1, 2]$.

The *Risk of informality* case has significant effects on the optimal design of unemployment insurance and employment protection. In particular, it makes it desirable to design the tax structure such that it provides incentives for firms to hire (and not fire) less-productive workers. This tax structure makes the informal sector less attractive at the cost of not firing less-productive workers who would be unemployed in the absence of the risk of entering the informal market.

First, we show how the proportion of people unemployed changes as the returns to informality increase (Figure 1). Recall that in the *No informality* case, the employment threshold, $\bar{y}^*$, is equal to $b$, as shown by the red line in the figure. For the *Risk of informality* case, one key result is that the risk of informality implies that the employment threshold must be lower and, as a consequence, unemployment must be lower. In particular, if the value of staying at home, $b$, is half of the informal sector wage, $w_i$, the threshold of productivity for employment implies that unemployment must be about 24 percent with no informality and about 10 percent with the informality risk. Thus, one lesson from our framework is that formal employment must be heavily subsidized in countries with high informality. This requirement is a result of the constraint in the design of the optimal policy that prevents workers from joining the informal sector. In the *Risk of informality* model, low-productivity workers are encouraged to work in the formal sector, whereas in the *No informality* model, they would be unemployed.
But how is a higher level of employment in the formal sector achieved? Figure 2, which exhibits the behavior of payroll taxes, shows the subsidy to employment. Taxes are always in the benchmark economy with no informality. The risk of informality requires subsidies to provide incentives to firms so they destroy relatively fewer matches with low-productivity workers. For high levels of $w_I/b$, making employment in the formal sector attractive requires the government to subsidize formal jobs so that effective matches are more likely, thereby increasing the utility of these formal jobs relative to the informal ones. However, recall that the larger $w_I/b$, the lower the unemployment benefits as informality needs to be made less attractive and, consequently, fewer and fewer subsidies are needed as incentives because unemployment payments are used more intensively to provide incentives.

Figure 3 shows the value of unemployment payments. Again, for values of $b = w_I$, the two cases coincide. In that case, the unemployment payment is such that consumption by unemployed and employed workers is the same. In contrast, for values of $w_I/b > 1$, informality implies that unemployment payments must decrease as the return to informality increases. This happens because unemployment payments are received with probability 1 by workers who are tempted to join the informal sector, while they are received with probability less than 1—actually $G(\bar{y})$—by those who decide not to join the informal sector. Thus, $s$ must decrease to guarantee that workers do not participate in the informal sector. This is the second lesson from our framework: To make the informal sector less attractive, unemployment payments cannot be as large in economies with labor informality.
**Figure 2**  
Payroll Taxes

![Graph showing Payroll Taxes with Payroll Tax, $\tau$, on the y-axis and Informal Sector Wage, $w/b$, on the x-axis. The graph includes two lines: one for No informality and one for Risk of informality.](image)

**Figure 3**  
Unemployment Insurance Payments

![Graph showing Unemployment Insurance Payments with Payment to Unemployed Workers, $s$, on the y-axis and Informal Sector Wage, $w/b$, on the x-axis. The graph includes two lines: one for No informality and one for Risk of informality.](image)
Figure 4 illustrates that severance payments are increasing with respect to the relative returns to informality $w_I/b$. Recall that these payments are used to implement the desired level of employment. Thus, since unemployment must decrease with $w_I/b$ (see Figure 1), it must be the case that severance payments increase. This is another lesson from our framework: When there is the risk of informality, severance payments must be used to prevent unemployment, but the payments are not transferred to unemployed individuals; rather, they are used to finance the negative payroll tax.

In addition, note that for very high values of $w_I/b$, the value of $s$ can indeed be negative. This is a key lesson from our framework: Unemployment benefits must be negative in situations of extreme informality. This requirement can be interpreted as an indirect tax on informality. Allowing negative unemployment payments is in contrast to the assumption in Álvarez-Parra and Sánchez (2009). In their case, unemployment payments are bounded below by zero. However, we can reconcile our implications with theirs. Since they studied a dynamic setting, the comparison must be the expected discounted sum of unemployment payments in their case and unemployment benefits in our static setting. They find that the duration of payments is shortened by the presence of informality. Thus, the expected discounted sum of benefits is reduced in their framework, as it is here in ours.

Figure 5 displays total consumption by the unemployed. For values of $w_I/b > 1$, the consumption by unemployed workers is lower with informality. Actually, the higher the relative return to informality, the lower is consumption by the unemployed. This is a direct conse-
Figure 5
Consumption by Unemployed Workers

Consumption by Unemployed Workers, $c_u$

- No informality
- Risk of informality

Informal Sector Wage, $w/b$

Figure 6
Consumption by Employed Workers

Consumption by Employed Workers, $c_e$

- No informality
- Risk of informality

Informal Sector Wage, $w/b$
sequence of the behavior of unemployment payments. Recall that such payments decrease to discourage workers from participating in the informal sector.

The consumption by workers, which coincides with their wage \( w \), is shown in Figure 6. Two forces move \( w \) for values of \( w_I/b > 1 \). First, as \( w_I/b \) increases, lower-productivity workers become employed and naturally the wage decreases—this is the only force that could move the wage in the No informality model. Second, since \( w \) is received only by workers who do not participate in the informal sector, providing incentives to keep workers out of the informal sector is helpful. This explains why consumption by employed workers is higher than in the No informality model. This is a further lesson from our framework: Working in the formal sector must be subsidized in economies with informality.

Finally, Figure 7 shows the cost of informality. This cost is computed in terms of the percent of consumption that workers would be willing to give up ex ante to eliminate the possibility of informality. Thus, this cost captures how costly it is to prevent such individuals from entering the informal sector. This cost is positive because informality is unobservable; therefore, the government must distort the optimal allocation to deal with the risk of informality. In addition, the cost is increasing in the relative returns to informality because as \( w_I/b \) increases, it is increasingly difficult to prevent the existence of informality.
CONCLUSION

This article studies optimal unemployment insurance and employment protection in the presence of the risk of informality. The analysis is performed in a simple static model along the lines of Blanchard and Tirole (2008) that is modified to incorporate the risk of informality.

We find that employment in the formal sector must be encouraged when the risk of informality exists. In particular, if the wage that individuals can obtain in the informal sector is quite high (extreme risk of informality), the government should set high severance payments and negative payroll taxes. For similar reasons, unemployment benefits cannot be as large in economies with the risk of informality as in those with no such risk. In fact, if the risk of informality is extreme, informality is taxed by setting negative unemployment benefits.

NOTES

1 This number is likely to be downward biased because it is based on random audits that most likely include only unemployment insurance recipients with jobs in the formal sector.

2 We assume that the cost is small enough, such that the return of creating a firm relative to what agents can produce at home implies that having a firm in the formal sector is optimal.

3 Note here that the “planner” in this economy is directly choosing (i) who works and who does not and (ii) the consumption by every agent, as usual.

4 Here, the term “internalization” is used in the sense of Pigou (1920).

5 This can be interpreted broadly, either because the match between the worker and the firm is new or because the worker was previously working for that firm, but his productivity follows an i.i.d. process with draws every period.

REFERENCES


