

FEDERAL RESERVE BANK OF ST. LOUIS

# REVIEW

MAY/JUNE 2002

VOLUME 84, NUMBER 3

**Conducting Monetary Policy Without Government Debt:  
The Fed's Early Years**

*David C. Wheelock*

**Unemployment Insurance Claims and Economic Activity**

*William T. Gavin and Kevin L. Kliesen*

**Did "Right-to-Work" Work for Idaho?**

*Emin M. Dinlersoz and Rubén Hernández-Murillo*

**Predicting Exchange Rate Volatility:  
Genetic Programming Versus GARCH and RiskMetrics™**

*Christopher J. Neely and Paul A. Weller*



# REVIEW

*Director of Research*

**Robert H. Rasche**

*Deputy Director of Research*

**Cletus C. Coughlin**

*Review Editor*

**William T. Gavin**

---

*Banking*

**R. Alton Gilbert**

**Frank A. Schmid**

**David C. Wheelock**

*International*

**Christopher J. Neely**

**Michael R. Pakko**

**Patricia S. Pollard**

*Macroeconomics*

**Richard G. Anderson**

**James B. Bullard**

**Michael J. Dueker**

**Hui Guo**

**Kevin L. Kliesen**

**Michael T. Owyang**

**Jeremy M. Piger**

**Daniel L. Thornton**

*Regional*

**Howard J. Wall**

---

*Managing Editor*

**George E. Fortier**

*Assistant Editor*

**Lydia H. Johnson**

*Graphic Designer*

**Donna M. Stiller**

*Review* is published six times per year by the Research Division of the Federal Reserve Bank of St. Louis. Single-copy subscriptions are available free of charge. Send requests to: Federal Reserve Bank of St. Louis, Public Affairs Department, P.O. Box 442, St. Louis, MO 63166-0442, or call (314) 444-8808 or 8809.

The views expressed 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.

© 2002, The Federal Reserve Bank of St. Louis. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if this copyright notice is included. Please send a copy of any reprinted, published, or displayed materials to George Fortier, Research Division, Federal Reserve Bank of St. Louis, P.O. Box 442, St. Louis, MO 63166-0442; george.e.fortier@stls.frb.org. Please note: Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis. Please contact the Research Division at the above address to request permission.

# Contents

## Volume 84, Number 3

### 1 **Conducting Monetary Policy Without Government Debt: The Fed's Early Years**

*David C. Wheelock*

The Federal Reserve implements its monetary policy by using open market operations in U.S. government securities to target the federal funds rate. A substantial decline in the stock of U.S. Treasury debt could interfere with the conduct of monetary policy, possibly forcing the Fed to rely more heavily on discount window lending or to conduct open market transactions in other types of securities. Either would cause the implementation of monetary policy to resemble the methods used by the Fed before World War II. This paper describes two things: (i) how the Fed implemented monetary policy before the war and (ii) the conflicts that arose within the Fed over the allocation of private-sector credit when discount window loans and Fed purchases of private securities were a substantial component of Federal Reserve credit. Those conflicts help explain the Fed's failure to respond vigorously to the Great Depression. The experience suggests that a renewed reliance on the discount window or on open market operations in securities other than those issued by the U.S. Treasury could hamper the conduct of monetary policy if it leads to increased pressure on the Fed to affect the allocation of credit.

### 15 **Unemployment Insurance Claims and Economic Activity**

*William T. Gavin and Kevin L. Kliesen*

Economic forecasters pay especially close attention to labor market indicators during periods of economic uncertainty. Labor market data are thought to provide early evidence about changes in the course of the economy. This article examines whether monthly changes in labor market indicators are useful for predicting real GDP. It then examines whether weekly

changes in initial and continuing unemployment insurance claims are useful for helping to predict changes in important labor market indicators. Incoming monthly data on nonfarm payroll jobs and the index of aggregate weekly hours help predict changes in real GDP growth, but data on the civilian unemployment rate do not. The authors also find that unemployment insurance claims help to predict changes in monthly labor variables. As others have found, these predictions work best in periods of recession. However, this article shows that there was also some predictive ability during the 1990s expansion.

### 29 **Did "Right-to-Work" Work for Idaho?**

*Emin M. Dinlersoz and Rubén Hernández-Murillo*

The Idaho legislature passed their right-to-work law in 1986. Idaho provides an interesting case study for the effects of the law because it adopted the law during a period in which unionization had already declined substantially in the United States. This paper provides evidence on the industrial organization consequences of right-to-work laws by jointly examining the trends in Idaho's unionization rate and industrial performance before and after 1986, using neighboring states as controls. We find that unionization declined substantially starting a few years before the passage of the law, and this decline persisted thereafter. Furthermore, the post-law growth rates in manufacturing employment and the number of establishments far exceeded their pre-law counterparts. As a result, Idaho gradually became more like an "average" right-to-work law state in terms of its unionization rate and the employment share of manufacturing, diverging from the patterns exhibited by states without a right-to-work law.

### 43 **Predicting Exchange Rate Volatility: Genetic Programming Versus GARCH and RiskMetrics™**

*Christopher J. Neely and Paul A. Weller*

This article investigates the use of genetic programming to forecast out-of-sample daily volatility in the foreign exchange market. Fore-

casting performance is evaluated relative to GARCH(1,1) and RiskMetrics™ models for two currencies, the Deutsche mark and the Japanese yen. Although the GARCH and RiskMetrics™ models appear to have an inconsistent marginal edge over the genetic program using the mean-squared-error (MSE) and  $R^2$  criteria, the genetic program consistently produces lower mean absolute forecast errors (MAE) at all horizons and for both currencies.

# Conducting Monetary Policy Without Government Debt: The Fed's Early Years

David C. Wheelock

**T**he Federal Reserve System is the largest single owner of U.S. Treasury securities in the world, excluding federal government accounts and trust funds such as the Social Security and Medicare trust funds. As of January 2, 2002, the Federal Reserve System Open Market Account held \$554.8 billion of U.S. government securities, or about 16 percent of the stock of Treasury securities held outside of government accounts and trust funds.<sup>1</sup> The Fed has acquired this portfolio through its monetary policy operations. Currently, the Fed implements its monetary policy by setting a target for the federal funds rate and using open market operations in U.S. government and agency securities to achieve that target. Purchases of securities supply reserves to the banking system, and thus tend to put downward pressure on the funds rate, whereas sales of securities remove reserves and put upward pressure on the funds rate.<sup>2</sup>

Federal Reserve holdings of government securities are the principal source of the nation's currency and depository institution reserve balances, and hence the U.S. monetary base. In principle, the Fed could add to the stock of bank reserves and currency by purchasing any asset, but U.S. Treasury securities offer at least two advantages over alternative assets for conducting monetary policy. First, because the Treasury market is extremely large and highly liquid, the Fed is able to conduct large transactions that give it substantial control over depository institution reserve balances and, hence, the federal funds rate, without having a disruptive impact on market prices.<sup>3</sup> Second, by using open market transactions in Treasury debt to implement monetary policy, the Fed avoids directly affecting the allocation of private capital, which Federal Reserve Chairman

Alan Greenspan (2001) and other System officials have cited as an important consideration in the conduct of monetary policy.<sup>4</sup>

Because of the close relationship between Federal Reserve holdings of U.S. Treasury securities and the monetary base, as well as the advantages of using open market operations in Treasury securities to implement monetary policy, a substantial decline in the outstanding stock of Treasury debt would pose a major challenge to policymakers. Indeed, the stock of Treasury securities available to private holders, including the Fed, declined from 1997 to 2001, prompting Federal Reserve officials to consider how the Fed might conduct monetary policy in a world without a deep, liquid Treasury securities market.

The substantial size and liquidity of the U.S. Treasury securities market emerged during World War II. The stock of outstanding Treasury debt ballooned during the war and remained large during the ensuing 50 years because of nearly continuous annual federal budget deficits. Thus, for evidence on how monetary policy might be conducted without substantial reliance on Treasury debt, it is necessary to look either to the experiences of other countries or to the Fed's own history before World War II. Accordingly, this article describes the implementation of Federal Reserve monetary policy before World War II and highlights that era's experiences that offer lessons for the conduct of policy in a possible future world without Treasury debt.

The record of Federal Reserve policy before World War II is not good, and some scholars contend that the poor performance of monetary policy was caused by the System's desire also to affect the

---

<sup>1</sup> Board of Governors of the Federal Reserve System Release H.4.1 (January 2, 2002).

<sup>2</sup> Fed transactions take the form of outright purchases and sales, as well as repurchase and matched sale-purchase transactions, of Treasury and agency securities. The Fed enters the market nearly every business day to offset influences on reserve markets beyond the Fed's immediate control, such as changes in the amount of currency in circulation or in the size of U.S. Treasury balances at Federal Reserve Banks, that otherwise would cause the federal funds rate to deviate from the Fed's target. See Board of Governors of the Federal Reserve System (1994) for additional information about the implementation of monetary policy.

<sup>3</sup> In 2001, the average *daily* volume of outright transactions in U.S. government securities, as reported by primary dealers, was \$298 billion. By contrast, in 2001, the Federal Reserve purchased an average of \$5.7 billion of Treasury securities per *month*. As noted, however, the Federal Reserve also engages in repurchase and matched purchase-sale agreements. See Dupont and Sack (1999) for an overview of the U.S. Treasury securities market.

<sup>4</sup> See also Broadus and Goodfriend (2001).

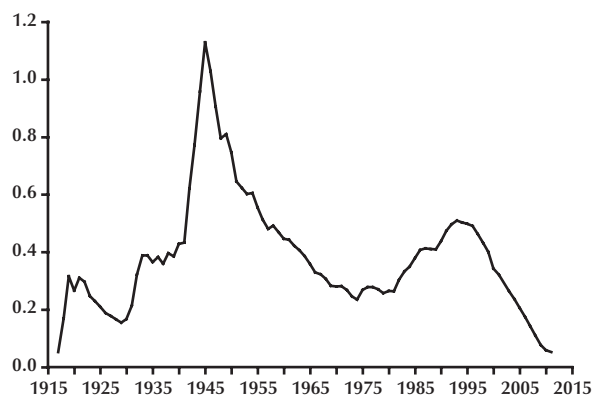
---

David C. Wheelock is an assistant vice president and economist at the Federal Reserve Bank of St. Louis. Heidi L. Beyer provided research assistance.

© 2002, The Federal Reserve Bank of St. Louis.

**Figure 1****U.S. Government Debt / GDP**

Annual data, 1917-2011\*



\*CBO forecasts for 2001-11.

SOURCE: Board of Governors of the Federal Reserve System, Bureau of Economic Analysis, and Robert Gordon, *Macroeconomics* (2000).

private allocation of credit. This article describes the Fed's desire that the credit it supplied to markets not encourage financial speculation or other forms of "unproductive" activity and how that desire caused policymakers to tighten credit overzealously in response to a perceived misallocation in the late 1920s and remain too tight as the economy collapsed into the Great Depression. The views that led to this outcome do not influence Federal Reserve policy today. However, as Goodfriend (1994) and Schwartz (1992) contend, the Fed is at times pressured to conduct a targeted credit policy. Moreover, if the Fed were forced to conduct monetary policy using private debt instruments, which could occur if the stock of U.S. Treasury debt were to decline substantially, such pressures might increase. The Fed's prewar experience described here provides one example of how the conduct of an effective monetary policy can be compromised by pressures to affect the allocation of private-sector credit.<sup>5</sup>

The next section discusses recent changes in the size of the U.S. Treasury debt and possible implications of a substantial decline in the outstanding stock of debt for the implementation of monetary policy. Following sections describe how the Fed's founders intended the System to conduct policy, the development of Federal Reserve monetary policy during the 1920s, and the conflicts created by the Fed's desire to prevent Federal Reserve credit from

financing speculative activity. Those conflicts help explain the Fed's failure to respond aggressively to the Great Depression and illustrate how a policy focused on the usage of Federal Reserve credit can interfere with the implementation of an effective stabilization policy.

## THE RISE AND (POSSIBLE) FALL OF THE STOCK OF TREASURY DEBT

As of December 31, 2001, the outstanding debt of the U.S. federal government totaled \$5943.4 billion, of which \$2549.0 billion was held by U.S. government agencies and trust funds such as the Social Security and Medicare trust funds. The remaining \$3394.4 billion of outstanding debt was held by the "public," consisting of private individuals, financial institutions and other firms, state and local governments, foreign concerns, and the Federal Reserve.<sup>6</sup> The stock of government debt held by the public reached a year-end peak in 1997 at \$3846.7 billion. Since then, the surplus revenues of government trust funds invested in Treasury securities have exceeded the amount by which total Treasury debt has increased.<sup>7</sup>

Often, the stock of government debt is measured relative to national output and, presumably, the implications of a given amount of debt for monetary policy depend on the size of the economy. As Figure 1 illustrates, the ratio of U.S. government debt held by the public to gross domestic product (GDP) soared during World War II; it then declined steadily to the mid-1970s before rising again to a peak in 1993. Through September 2001 (the end of the fiscal year), the stock of Treasury debt then grew at a slower rate than did U.S. GDP. Projections by the Congressional Budget Office and other forecasters indicate that the debt-to-GDP ratio will continue to decline for at least the next decade. Past experience indicates that the size of the federal debt

<sup>5</sup> Whereas the implementation of monetary policy using private credit instruments can lead to potential conflicts between the conduct of stabilization policy and the allocation of credit, analogous conflicts can arise if the central bank conducts open market transactions in any asset. For example, in the nineteenth and early twentieth centuries, U.S. silver mining interests often pressed the federal government to purchase and coin silver. Conceivably, such efforts to raise the relative price of silver could conflict with monetary policy objectives. See Friedman and Schwartz (1963, pp. 483-91) for a discussion of a silver purchase program implemented in the 1930s.

<sup>6</sup> See < [www.publicdebt.treas.gov/opd/opdpdodt.htm](http://www.publicdebt.treas.gov/opd/opdpdodt.htm) > .

<sup>7</sup> On September 30, 2001, the close of fiscal year 2001, the stock of debt held by the public totaled \$3339.3 billion.



is difficult to forecast at horizons of more than a year or two (see Kliesen and Thornton, 2001). Nevertheless, if true, recent projections indicate that by mid-decade the debt-to-GDP ratio could fall to a level not observed since the 1920s, and by about 2010 fall to a level not observed since before World War I.<sup>8</sup>

A substantial decline in the volume of outstanding Treasury securities would have repercussions for both the Fed and the financial system. Treasury securities, especially Treasury bills, serve as liquid, risk-free investments and collateral for banks and other financial market participants. For the Fed, either a substantial increase in discount window borrowing, which typically is secured by loans and other private claims, or greater use of securities other than those issued by the U.S. Treasury in the conduct of open market operations would expose the System to more credit risk than it faces today. Aside from possibly affecting monetary policy, such exposure could complicate other duties the Fed performs. For example, if the Fed were to become a major creditor of the banks that it supervises, then any aggressive actions it might take as a bank supervisor to deal with problem banks could increase the probability of losses by the Fed as a bank creditor.

The Fed has relied mainly on open market operations in Treasury debt to implement monetary policy since World War II. Recently, Federal Reserve Chairman Alan Greenspan (2001) summarized why Treasury securities are a convenient asset for the Fed to conduct its operations: “First, the liquidity of the market allows the Federal Reserve to make substantial changes in reserves in a short period of time, if necessary. Second, the size of the market has meant that the effects of the Federal Reserve’s purchases on the prices of Treasury securities have been minimal. Third, Treasury securities are free of credit risk...[and] we believe that the effects of Federal Reserve operations on the allocation of private capital are likely to be minimized when Federal Reserve intermediation involves primarily the substitution in the public’s portfolio of one type of instrument that is free of credit risk—currency—for another—Treasury securities.”

Greenspan went on to identify how the Fed might respond to a substantial decline in the stock of Treasury debt: “One possibility is to expand the use of the discount window by auctioning such credit to financially sound depository institutions... Another possibility is to add new assets to those the Fed is currently allowed by law to buy for its portfolio.” Greenspan cited Ginnie Mae securities

and certain types of municipal or foreign government obligations as examples of securities that the Fed might use for open market operations. Either increased reliance on the discount window (in which depository institutions borrow reserves mainly against collateral other than U.S. Treasury securities) or an expansion of the financial assets the Fed purchases in the open market would cause the implementation of monetary policy to resemble more closely the methods used by the Fed before World War II.

## WHAT THE FED’S FOUNDERS INTENDED

In discussing the advantages of conducting open market operations in U.S. Treasury debt, Chairman Greenspan (2001) argued that “it is important that government holdings of assets not distort the private allocation of capital” and that “if the Treasury debt is paid down...then the Federal Reserve will have to find alternative assets that still provide substantial liquidity and minimize distortions to the private allocation of capital.” The notion that actions to implement stabilization policy should not distort the private allocation of capital is not controversial today. When the Federal Reserve System was established in 1914, however, its founders very much did intend the System to favor certain uses of private sector capital over others. Moreover, the founders had no expectation that the Fed would conduct stabilization policy as we know it today. In short, the Fed’s founders envisioned that the Fed would conduct *credit* policy, but not *monetary* policy, in that Federal Reserve operations were expected to influence the private allocation of credit but not regulate the growth rate of the monetary base or the level of interest rates to achieve macroeconomic stability.<sup>9</sup>

### The Discount Window

The Fed’s founders expected that Federal Reserve Banks, like the central banks of Europe, would serve mainly as lending institutions for their member commercial banks. Before the Fed’s establishment, the U.S. commercial banking system suffered numer-

<sup>8</sup> Projected annual ratios of debt to GDP for 2001–11 plotted in Figure 1 are from a Congressional Budget Office (2001) forecast made in August 2001, which was the latest available as of year-end 2001.

<sup>9</sup> This distinction is discussed in Goodfriend (1994) and Broadus and Goodfriend (2001).

ous banking panics and, at times, high failure rates (see Dwyer and Gilbert, 1989). Proponents of the Federal Reserve System argued that the Fed would make the banking system more stable by providing commercial banks with a ready source of reserves to meet fluctuations in the demands for credit and currency. To perform that role, Federal Reserve Banks were given a lending facility—their discount windows—through which they would rediscount eligible financial assets for member commercial banks in exchange for currency or reserve deposit balances.<sup>10</sup> Member banks were required to maintain minimum reserve balances with their Reserve Bank, and the Reserve System provided currency (Federal Reserve notes) and payments services for member institutions.<sup>11</sup>

The Federal Reserve Act provided that “any Reserve Bank may discount notes, drafts, and bills of exchange arising out of actual commercial transactions; that is, notes, drafts, and bills of exchange issued or drawn for agricultural, industrial, or commercial purposes.” In addition, to be eligible for rediscount, agricultural paper could have a maturity of no more than six months, whereas nonagricultural paper had to mature in 90 days or less.

The limits that the Fed’s founders placed on the type of securities eligible for rediscount with Reserve Banks reflected conventional banking principles of the time, the so-called Real Bills Doctrine. By limiting discount loans to short-term commercial and agricultural loans, the Fed’s founders expected that the System would supply a sufficient volume of credit to accommodate growth and fluctuations in real economic activity without causing inflation or speculation. The Federal Reserve Act explicitly prohibited the rediscount of “notes, drafts, or bills covering merely investments or issued or drawn for the purpose of carrying or trading stocks, bonds, or other investment securities, except bonds and notes of the Government of the United States.” By ruling such securities ineligible, the Fed’s founders sought to prevent Federal Reserve credit from being used to finance transactions or investments that had no obvious, direct connection to the production, distribution, or sale of specific products or commodities.<sup>12</sup>

Apparently, the authors of the Federal Reserve Act believed it more important to specify precisely the type of securities that would be eligible for rediscount than to specify criteria for setting the discount rate. The Federal Reserve Act stated only that Reserve Bank discount rates should be determined “with a view of accommodating commerce and business.”

Reserve Banks were required to maintain a gold reserve against their liabilities, however, which imposed implicit bounds on their lending rates.

### **Open Market Operations in Bankers Acceptances**

The coalition of interests supporting the founding of the Federal Reserve System included both small community banks and large money center banks.<sup>13</sup> The large money center banks were particularly interested in promoting the dollar as an international currency and thereby increasing the share of international transactions financed by U.S. banks (Broz, 1997). Accordingly, the Federal Reserve Act permitted U.S. commercial banks to issue bankers acceptances to finance foreign trade. The Act further sought to encourage the development of a U.S. acceptance market by permitting Federal Reserve Banks to acquire acceptances by rediscount or open market purchase. The Reserve Banks established interest rates at which they purchased all eligible acceptances offered to them, and such purchases were a major source of Federal Reserve credit during the System’s first two decades.

### **Open Market Operations in U.S. Government Securities**

In addition to open market purchases of bankers acceptances, the Federal Reserve Act authorized Reserve Banks “to buy and sell, at home or abroad, bonds and notes of the United States, and bills, notes, revenue bonds, and warrants . . . issued . . . by any State [sic], county, district, political subdivision, or municipality in the continental United States.” This provision was intended to provide the Reserve Banks with a source of revenue in the event that income from

<sup>10</sup> At the time, most commercial and agricultural loans were made on a discount basis. Hence, when such loans were offered to Reserve Banks in exchange for currency or reserve balances, the Reserve Banks rediscounted the paper at the current discount rate.

<sup>11</sup> The Fedwire system, which the Fed established in 1918 to effect inter-bank payments electronically, was among the innovations enhancing the liquidity of the payments system. See Gilbert (1998) for an analysis of how the founding of the Federal Reserve affected the efficiency of the U.S. payments system.

<sup>12</sup> Burgess (1936) and West (1977) discuss the objectives of the Fed’s founders in detail.

<sup>13</sup> This is not to say that all banks favored the creation of the Fed. Numerous small, state-chartered banks elected not to join the System, and many opposed the Fed’s check collection practices (see Gilbert, 1998). For further analysis of the Fed’s “membership problem,” see White (1983).



rediscounts and the provision of services was insufficient to meet Bank expenses (Chandler, 1958, p. 76). The Fed's founders did not contemplate that open market operations would be used to influence the level of market interest rates or the growth rate of the money stock in an effort to stabilize the price level or economic activity. It was not long, however, before Federal Reserve policymakers discovered that open market operations could affect the level of interest rates and, potentially, influence economic activity.

## THE SOURCES OF FEDERAL RESERVE CREDIT

Federal Reserve credit constitutes the Fed's contribution to the stock of bank reserves and currency in circulation, the sum of which is referred to as the monetary base or high-powered money. Other sources of the monetary base include Treasury currency outstanding (e.g., coins) and, historically, the monetary gold stock. The principal sources of Federal Reserve credit are discount window loans and Fed purchases of U.S. government securities. Before World War II, Fed purchases of bankers acceptances also contributed meaningfully to Fed credit.<sup>14</sup>

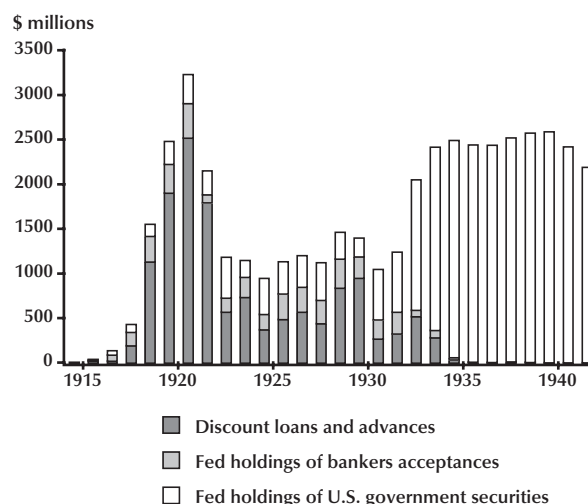
Figure 2 illustrates the relative volumes of discount window loans, Federal Reserve holdings of acceptances, and Fed holdings of U.S. government securities during 1914-41. Total Federal Reserve credit grew sharply during World War I, when the Fed committed itself to helping finance the war effort. The Federal Reserve Act was amended in 1916 to permit member banks to borrow directly from the Fed using U.S. government securities and other eligible assets as collateral. Discount loan volume soared when the Fed established preferential discount rates for advances secured by U.S. government securities that guaranteed banks a profit on their holdings of such securities. In June 1917, discount window loans outstanding (rediscounts and advances) totaled \$197 million, of which 13 percent were advances against member bank holdings of U.S. government securities. In June 1919, by contrast, discount loans totaled \$1818 million; of these, 87 percent represented rediscounts of, or advances against, U.S. government securities (Board of Governors, 1943, p. 340).

The Federal Reserve also purchased considerable U.S. government securities in the open market during the war. Fed holdings of Treasury securities increased from \$66 million in June 1917 to \$236

**Figure 2**

### Principal Sources of Federal Reserve Credit

Annual data, 1914-41



SOURCE: Board of Governors of the Federal Reserve System.

million in June 1919. The percentage of total Fed credit outstanding accounted for by Fed purchases of U.S. government securities remained at approximately 10 percent over the two years. Meanwhile, the Fed also acquired bankers acceptances in the open market. Throughout the war, more Federal Reserve credit was extended by the purchase of acceptances than by open market purchases of U.S. government securities.

The Fed retained preferential discount rates on loans secured by U.S. government securities after World War I. By late 1919, however, declining reserve ratios at several Reserve Banks prompted the Banks to increase their discount rates and to discontinue preferential rates on loans secured by U.S. government securities.<sup>15</sup> Discount window loan volume dropped sharply in 1921-22. From a peak of \$2808 million in October 1920, discount loans outstanding fell to less than \$400 million in August 1922 (Board of Governors, 1943, p. 374).

The Reserve Banks sought to offset the loss of revenue associated with the decline in discount loans by purchasing government securities in the

<sup>14</sup> Other sources of Fed credit include check float and Federal Reserve purchases of foreign currency.

<sup>15</sup> Reserve Banks were required by law to maintain reserves of gold and eligible securities against their outstanding liabilities.

open market (Chandler, 1958, p. 209). Open market operations in government securities remained an important source of Federal Reserve credit throughout the 1920s and even more so during the early 1930s, when discount window loans and Federal Reserve purchases of bankers acceptances dwindled. By 1932, open market purchases of U.S. government securities had become the dominant source of Federal Reserve credit, and for the remainder of the decade the Fed made almost no discount loans and purchased almost no bankers acceptances.

## MONETARY VERSUS CREDIT POLICY

The changes in the composition of Federal Reserve credit during the 1920s reflect well the evolution from the self-regulating *credit* policy envisioned by the Fed's founders toward a modern *monetary* policy. By the mid-1920s, the Fed had begun to use open market operations in U.S. government securities to influence money market conditions, with the twin goals of promoting domestic economic stability and the international gold standard. Credit policy came back to the fore in 1928-29, however, when the Fed sought to check stock market speculation without unduly restricting the flow of credit to "legitimate" borrowers. The consequent stance of monetary policy proved extremely restrictive, the stock market crashed, and the U.S. economy collapsed.<sup>16</sup> The Fed did not ease monetary policy aggressively in response to the collapse, however, in part because some Fed officials feared that loose monetary policy would reignite financial speculation. This section discusses the origins of Federal Reserve monetary policy during the 1920s and the conflict between monetary and credit policy that emerged during the late 1920s. The following section explores how this conflict influenced the setting of monetary policy during the Great Depression.

### The Birth of Monetary Policy

Although the provisions of the Federal Reserve Act permitting the Reserve Banks to acquire government securities were little more than an afterthought, Reserve Bank officials, especially at the Federal Reserve Bank of New York, observed that their purchases affected market interest rates and credit conditions. After World War I, the Reserve Banks formed a committee of Reserve Bank governors, headed by Federal Reserve Bank of New York Governor Benjamin Strong, to establish policies for the conduct of open market operations and to

coordinate open market purchases for all Reserve Banks.<sup>17</sup> Strong, according to Chandler (1958), was among the first Reserve System officials to comprehend the impact of open market operations on money market and credit conditions, as well as to favor the use of open market operations to achieve general macroeconomic goals.

Under Strong's leadership, the Fed began to use open market operations in U.S. government securities during the 1920s to implement an active *monetary* policy. This policy clashed with the *credit* policy objectives of members of the Federal Reserve Board and some Reserve Banks. This conflict came to a head over how to control stock market speculation in 1928-29 and how to respond to the economic depression that followed the stock market crash in 1929.

Strong directed two major monetary policy operations during the 1920s, involving substantial open market purchases in 1924 and 1927. The motivation for these operations has been debated. Chandler (1958), Friedman and Schwartz (1963), Meltzer (1997), and Wicker (1966) all argue that Strong was motivated by a desire to ease money market conditions, but they disagree about Strong's ultimate objective. Friedman and Schwartz (1963) contend that in both years Strong was motivated primarily by a desire to promote domestic recovery from a recession. Wicker (1966), however, argues that Strong's primary motivation was to redirect the international flow of gold away from the United States toward the United Kingdom, in an effort to help Britain first restore, then preserve, gold convertibility of the pound. Chandler (1958) and Meltzer (1997) contend that both objectives were important, and Wheelock (1991) reports econometric evidence consistent with their conclusion.

Whatever Strong's motivation, his use of open market operations caused considerable controversy within the Federal Reserve System. Strong's initiative irritated members of the Federal Reserve Board, who believed that the committee of governors had overstepped its authority. Several members of the Board also opposed open market purchases, espe-

<sup>16</sup> Schwartz (1981) and Hamilton (1987) argue that tight monetary policy in 1928-29 was an important cause of the Great Depression.

<sup>17</sup> From 1914 to 1935, the chief executive officer of each Federal Reserve Bank held the title "governor," as did the chair of the Federal Reserve Board. The Banking Act of 1935 changed the title of Reserve Bank chief executives to "president" and assigned the title "governor" to each member of the Federal Reserve Board, which was renamed the Board of Governors of the Federal Reserve System.

cially in 1927, on economic grounds. Most members of the Board, and officials of some Reserve Banks, believed that Federal Reserve credit should be extended only at the initiative of member commercial banks through the rediscounting of commercial and agricultural loans. Otherwise, those officials argued, the Fed risked contributing to speculative activities that could prove harmful to the economy.

Strong, on the other hand, contended that the Fed could help lift the economy during a period of weakness by using open market purchases to ease monetary conditions. At a meeting of Reserve Bank governors in 1926, Strong argued: "Should we go into a business recession while the member banks were continuing to borrow [from Reserve Bank discount windows]...we should consider taking steps to relieve some of the pressure which this borrowing induces by purchasing Government securities and thus enabling member banks to reduce their indebtedness" (quoted in Chandler, 1958, pp. 239-40). In Strong's view, by enabling banks to repay their discount window borrowings, open market purchases would ease money market conditions and promote economic recovery.

### ***The Stock Market and the Return of Credit Policy***

Although the disagreements about open market purchases in 1924 and 1927 were sharp, the most heated debates within the System during the 1920s focused on how the Fed should respond, if at all, to the rising stock market. The rapid increase in stock prices and growth of loans from banks and brokers to finance stock purchases in 1928 and 1929 concerned System officials who sought to ensure that reserves supplied by the Fed were not being used to finance speculation. Most members of the Federal Reserve Board favored a "direct action" policy in which member banks with outstanding loans to finance stock purchases would be prohibited from borrowing at the discount window. Board members thought that by enforcing this restriction, discount rates would not have to rise and thereby penalize borrowers with "legitimate" credit demands.

Officials of Federal Reserve Banks, however, generally believed it neither practical nor desirable for the Fed to affect the private allocation of credit. In a 1925 memorandum, Benjamin Strong asked rhetorically how the Fed should respond to calls for action against real estate and stock market speculation or, for example, to "too much enthusiasm in automobile production":

Where does our responsibility lie? Must we accept parenthood for every economic development in the country? That is a hard thing for us to do. We would have a large family of children. Every time any one of them misbehaved, we might have to spank them all. There is no selective process in credit operations. If we undertake "direct action" in one case, we would be saddled with the responsibility for direct action in all cases. Have we infallible good judgment as well as sufficient knowledge to play the role of parent?...Of one thing I am sure...and that is that we have no direct responsibility to deal with isolated situations, and must rely for the development of our policy upon estimates of the whole situation. (quoted in Chandler, 1958, p. 428)

In Strong's view, the Fed should be concerned with the stock market, or any other particular market, only to the extent that it bears on the behavior of the economy as a whole.

To the extent that a rising stock market meant that monetary policy should become tighter, Strong favored raising the discount rate and conducting open market sales, rather than placing special restrictions on banks that made stock market loans. Such restrictions, he argued, would not limit the flow of credit to the stock market: "the money will go into the stock exchange anyway" (quoted in Chandler, 1958, p. 430). Even if the Fed lent only to banks that made no stock market loans, Strong claimed, reserves supplied through the discount window (or via open market purchases) could still end up enabling banks in the aggregate to increase stock market loans: "If we create an addition to the volume of credit by our open-market operations or by our discounts, the banks which get it [i.e., the credit] pass it along through all the channels through which credit circulates in our banking system—and we cannot control what happens to it. Some of it will go in one direction and some of it will go in another, and the nature and the use of our funds is perfectly impossible to control" (quoted in Chandler, 1958, pp. 431-32).

### ***Direct Action***

At its meeting on January 11, 1928, the Federal Reserve's open market committee decided to implement a more restrictive monetary policy, defined as "somewhat firmer money conditions," so as to



“check unduly rapid further increases in the volume of credit” (quoted in Chandler, 1971, p. 38). The Reserve Banks also began to increase their discount rates and, for the most part, these initial restrictive actions were supported by the Federal Reserve Board. One Board member, however, dissented from all moves to tighten policy. In explaining his vote against a discount rate increase in 1928, Edward Cunningham stated that “I feel that increases in the discount rate for the purpose of restricting stock market activities should only be resorted to when other means within the power of the Board have failed to accomplish the objective. I am not in favor of penalizing agriculture and business because of the indirect use of credit for investments in brokers loans” (quoted in Chandler, 1971, p. 43).

Despite further discount rate increases and open market sales in 1928, Fed officials were frustrated by their apparent inability to control the flow of credit to the stock market. Chandler (1971, pp. 52-53) summarizes the quandary the Fed found itself in at the beginning of 1929:

By late January 1929 the Federal Reserve’s policy of restriction had been in effect about a year. Monetary and credit conditions had changed markedly during the period. Member bank borrowings at the Federal Reserve had nearly doubled, rising to nearly \$900 million, equal to 37 percent of total bank reserves... The total volume of bank credit was barely above its level of a year earlier. Interest rates had risen sharply... Call-loan rates averaged above 7 percent in December 1928 and frequently reached considerably higher levels. However, the Federal Reserve had not achieved its objective of curbing stock speculation. Share prices rose 38 percent in the year... Brokers’ loans reached the unprecedented level of \$6.4 billion; this reflected an increase of 45 percent for the year... Domestic business activity was still at high and rising levels, but even here there were warning signs in the form of decreasing availability of mortgage money, a downturn in construction, and increasing difficulties in floating long-term bond issues.

In these circumstances, disagreements within the System over how to respond to the stock market became more heated.

Federal Reserve Board officials believed that the Reserve Banks had not properly administered

their discount windows and were permitting member banks to borrow reserves to support “speculative” lending, meaning primarily loans to stock brokers and dealers and to customers for the purpose of purchasing securities. Although Reserve Banks rediscounted only loans and securities that were eligible as defined by the Federal Reserve Act, Board officials argued that commercial banks should be forced to liquidate their speculative loans before being permitted to rediscount (or borrow against) eligible paper. On February 2, 1929, the Federal Reserve Board sent a letter to each of the 12 Reserve Banks in which the Board stated that

The Federal Reserve Act does not... contemplate the use of the resources of the Federal reserve banks for the creation or extension of speculative credit. A member bank is not within its reasonable claims for rediscount facilities at its Federal reserve bank when it borrows either for the purpose of making speculative loans or for the purpose of maintaining speculative loans.

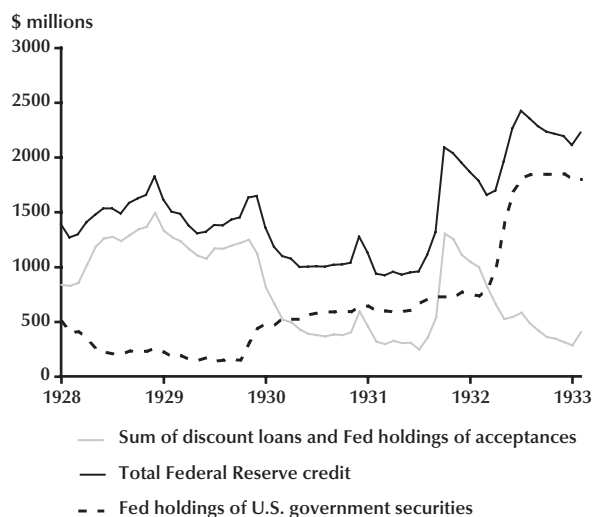
The letter went on to request that each Reserve Bank report to the Board as to “a) how they keep themselves fully informed of the use made of borrowings by their member banks, b) what methods they employ to protect their institution against the improper use of its credit facilities by member banks, and c) how effective these methods have been” (quoted in Chandler, 1971, pp. 56-57).

Although the Board’s instructions to the Reserve Banks were vague—for example, the terms “speculative credit” and “speculative loans” were not defined—the Reserve Banks made some effort to comply with the Board’s request that they administer their discount windows more tightly. At the same time, the Reserve Banks pressed for increases in their discount rates, but were denied by the Federal Reserve Board. Consequently, Reserve Bank officials grew increasingly frustrated with the Board’s “direct action” policy of tightly restricting access to the discount window. As George Norris, Governor of the Federal Reserve Bank of Philadelphia, complained to one Board member:

This whole process of “direct action” is wearing, friction-producing, and futile. We are following it honestly and energetically, but it is manifest, beyond...doubt, that it will never get us anywhere. It is like punching at a mass of dough. You make a dent where

**Figure 3****Principal Sources of Federal Reserve Credit**

Monthly data, January 1928–February 1933



SOURCE: Board of Governors of the Federal Reserve System.

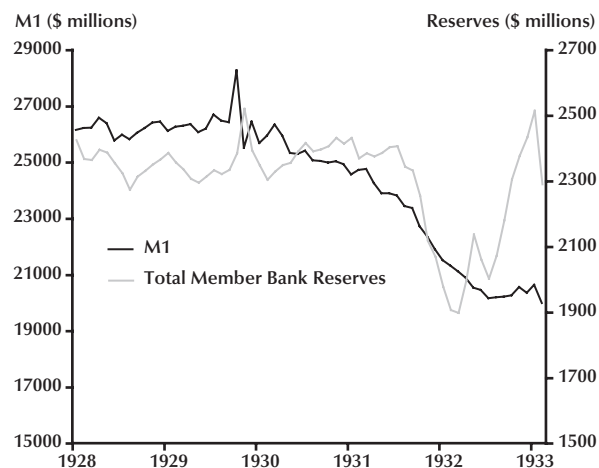
you hit, but the mass swells up at another point. As long as we maintain a discount rate which is absurdly low, and out of proportion to all other rates, the present conditions will continue. Our 5 per cent rate is equivalent to hanging a sign out over the door "Come in," and then we have to stand in the doorway and shout "Keep out." It puts us in an absurd and impossible position. (Quoted in Chandler, 1971, p. 66)

## THE CRASH AND GREAT DEPRESSION

The Federal Reserve Board eased its policy of "direct action" in June 1929, when economic activity had begun to slow and Fed officials were concerned that credit had become too tight. In August, however, the discount rate of the Federal Reserve Bank of New York was increased in an effort to discourage borrowing, and market interest rates remained high until the stock market crashed in October. Rates then fell sharply. The crash prompted the Federal Reserve Bank of New York to make large open market purchases while also lending heavily through its discount window. The Federal Reserve, however, did not respond aggressively to the sharp decline in economic activity that followed the stock market crash or to the banking panics that occurred over the

**Figure 4****M1 Money Stock and Total Bank Reserves**

Monthly data, January 1928–February 1933



SOURCE: Board of Governors of the Federal Reserve System and Friedman and Schwartz (1963) for M1.

next three years.<sup>18</sup> One reason for the Fed's inaction is that Federal Reserve officials remained mired in debate over whether the System should attempt to channel credit to "appropriate" uses or pursue an active stabilization policy. This section reviews that debate, focusing on the arguments of Fed officials who opposed the use of expansionary monetary policy to revive the economy.

### The Fed's Response

The Federal Reserve Bank of New York purchased some \$160 million of government securities for its own account immediately following the stock market crash, and the System purchased another \$150 million of securities before the end of 1929. During 1930 and most of 1931, however, Fed purchases of government securities were insufficient to offset net declines in discount window loans and Fed purchases of bankers acceptances. Hence, total Federal Reserve credit outstanding fell (see Figure 3). As Friedman and Schwartz (1963) show, the money stock began to fall in this phase of the Depression (see Figure 4).

Monetary contraction accelerated in the fourth quarter of 1931, when speculation that the United

<sup>18</sup> Macroeconomic conditions during the Depression are summarized in Wheelock (1992).



States would abandon the gold standard triggered bank runs and a gold outflow. Banks borrowed heavily from the discount window to replace lost reserves and Federal Reserve credit increased sharply. The Fed did not make substantial open market purchases of government securities, however, claiming that it lacked sufficient gold reserves.<sup>19</sup>

Easier monetary policy did come in 1932 when, under pressure from Congress, the Fed purchased some \$1 billion of U.S. government securities between March and August.<sup>20</sup> The total increase in Federal Reserve credit was less than \$1 billion because of declines in discount window loans and Fed holdings of bankers acceptances, but member bank reserves increased and the money stock stopped falling. The banking crisis resumed in early 1933, however, triggering a series of state bank suspensions and prompting President Franklin Roosevelt to declare a national bank holiday and suspend the gold standard when he took office in March. The Federal Reserve, meanwhile, stayed in the shadows as the new president took charge of macroeconomic policy.

### ***How Should the Fed Respond to a Decline in Economic Activity?***

Why the Federal Reserve failed to respond more aggressively to the Great Depression has been the subject of considerable research.<sup>21</sup> Certainly, there were officials in the System who advocated a more vigorous response to the Depression. Officials of the Federal Reserve Bank of New York, for example, argued that recovery required low interest rates, a strong bond market, and a sufficient supply of reserves to free member banks from having to obtain discount window loans.<sup>22</sup> In July 1930, New York Fed Governor George Harrison wrote to his counterparts at other Reserve Banks urging that the Fed “do everything possible and within its power to facilitate a recovery of business.” He went on to advocate open market purchases: “In previous business depressions, recovery has never taken place until there has been a strong bond market,” and, moreover, “we cannot foresee any appreciable harm” from making open market purchases (quoted in Friedman and Schwartz, 1963, p. 370).

Outside of New York, however, many Fed officials were convinced that Federal Reserve credit should *contract* with declines in economic activity and loan demand. Those officials claimed that in the absence of demand for loans by business and agri-

cultural borrowers, reserves created by an expansion of Federal Reserve credit would be used to finance speculation. Many argued that open market purchases during recessions in 1924 and 1927 had been a mistake and had contributed to the financial speculation that they saw as responsible for the subsequent economic depression. For example, Adolph Miller, a member of the Federal Reserve Board who had voted against open market purchases in 1927, testified in 1931 that the operation “was the greatest and boldest operation ever undertaken by the Federal Reserve System, and, in my judgment, resulted in one of the most costly errors committed by it or any banking system in the last 75 years... That was a time of business recession. Business could not use and was not asking for increased money at that time” (U.S. Senate, 1931, p. 134).

Miller’s view was not unique among System officials. In response to a written question from the Senate Banking Committee in 1931 about open market purchases in 1924 and 1927, officials of the Federal Reserve Bank of Richmond wrote that “we think United States securities should not have been purchased in these periods, and the aim should have been to decrease rather than augment the total supply of Federal Reserve credit” (U.S. Senate, 1931, p. 100). Officials of the Federal Reserve Bank of Philadelphia responded similarly, arguing that Federal Reserve credit should never be extended except at the initiative of member banks. Other Reserve Banks replied that the open market purchases of 1924 and 1927 had been justified, but were too large.

<sup>19</sup> Burgess (1936, pp. 285-86) argues that the Fed was constrained by a lack of gold reserves, but Friedman and Schwartz (1963, pp. 399-406) dismiss the Fed’s excuse. See also Chandler (1971, pp. 182-91).

<sup>20</sup> The Glass-Steagall Act of 1932 permitted the Fed to back Federal Reserve notes with U.S. government securities, which greatly eased the Fed’s gold reserve requirement.

<sup>21</sup> See Friedman and Schwartz (1963), Wicker (1966), Brunner and Meltzer (1968), Wheelock (1991, 1992), and Meltzer (1994).

<sup>22</sup> During the Depression, Fed officials interpreted historically low levels of borrowed reserves and interest rates as indicating that monetary conditions were exceptionally easy. No one in the System, and almost no one outside the Fed, recognized that a falling price level and widespread banking panics, let alone a decline in the money stock, meant that monetary policy was in fact exceptionally tight. Friedman and Schwartz (1963) argue that the Fed would have responded much more aggressively to the Depression had Benjamin Strong not died in 1928. During the 1920s, however, Strong advocated basing the volume of open market operations on the levels of borrowed reserves and market interest rates; the Fed’s anemic response to the Depression does not seem inconsistent with that framework (see Wheelock, 1990, 1991, 1992).

Those officials who were critical of open market purchases during the 1920s tended to argue that Federal Reserve credit should be extended only at the initiative of member banks, through the discount window or by sales of bankers acceptances to the Fed. Open market purchases of government securities, by contrast, constituted “artificial” easing, which Federal Reserve Bank of St. Louis Governor William McChesney Martin Sr. argued was “unwise” and possibly “hazardous” (quoted in Chandler, 1971, p. 142). The Federal Advisory Council argued similarly in 1930 that “the present situation will be best served if the natural flow of credit is unhampered by open-market operations” (quoted in Friedman and Schwartz, 1963, p. 373). Such operations, claimed Chairman Richard Austin of the Federal Reserve Bank of Philadelphia, “lays us open to the apparent undesirable charge that the action is not justified by the demand for credit but for some other purpose, it may be for boosting business, making a market for securities, or some other equally criticizable cause that certainly will come back to plague us” (quoted in Chandler, 1971, p. 136).

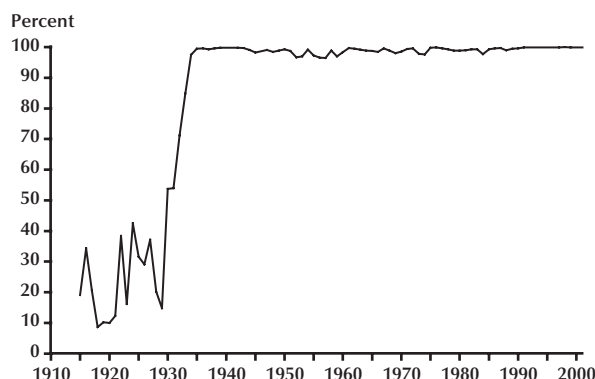
Several Fed officials argued that monetary policy could do little to bring about recovery from the Depression. Officials of the Federal Reserve Bank of Philadelphia, for example, concluded that “the correction must come about through reduced production, reduced inventories, the gradual reduction of consumer credit, the liquidation of security loans, and the accumulation of savings through the exercise of thrift” (quoted in Chandler, 1971, p. 137). And, in response to a proposal for open market purchases, Governor James McDougal of the Federal Reserve Bank of Chicago replied that the market already had an “abundance” of funds. Further, he argued that the Fed should “maintain a position of strength, in readiness to meet future demands... rather than to put reserve funds into the market when not needed” (quoted in Friedman and Schwartz, 1963, p. 371).

In the view of McDougal and several other Fed officials, open market purchases would have little positive impact on economic activity and could in fact interfere with economic recovery by delaying the liquidation of loans and speculative investments that, in their view, was necessary for recovery to begin. Moreover, in the absence of an obvious demand for Federal Reserve credit, as evidenced by discount window borrowing or sales of bankers acceptances to Reserve Banks, McDougal and others believed that reserves created by open market purchases could result in a dangerous misallocation

**Figure 5**

### Government Security Holdings as a Percent of Total Adjusted Federal Reserve Credit\*

Annual data, 1915-2001



\*Total adjusted Federal Reserve credit = Sum of discount loans and advances to depository institutions and Federal Reserve holdings of U.S. government and other securities.

SOURCE: Board of Governors of the Federal Reserve System.

of credit. They were not able to prevent open market purchases altogether, but their resistance undoubtedly slowed the Fed's response to the Great Depression.

## LESSONS

Ironically, the Fed's unwillingness to purchase a large volume of government securities early in the Depression ultimately may have contributed to such purchases becoming the dominant source of Federal Reserve credit. As the Depression continued and banking panics worsened, commercial banks became increasingly unable and unwilling to come to the discount window. Some banks lacked eligible collateral for discount window loans, while others feared that borrowing would trigger deposit withdrawals by giving the appearance of weakness.<sup>23</sup> By 1932, discount window borrowing and Federal Reserve purchases of bankers acceptances had fallen to minimal levels, where they stayed throughout the remainder of the decade. As Figure 5 illustrates, Fed holdings of U.S. government securities had become by far the most important source of Federal Reserve credit. Since 1934, the size of the Fed's

<sup>23</sup> Chandler (1971, pp. 225-33) concludes that borrowing was reduced to some extent by a lack of eligible collateral and a heightened reluctance to borrow. Wheelock (1990) finds that discount window borrowing declined more during 1930-33 than could be explained simply by the decline in economic activity.

portfolio of U.S. government securities has always comprised over 90 percent of total adjusted Federal Reserve credit outstanding (the sum of discount loans and Fed holdings of U.S. government and other securities).

Monetary policy lay dormant from the mid-1930s to 1951. Neither the size of the Fed's government security portfolio nor total Fed credit outstanding changed substantially between 1934 and 1941.<sup>24</sup> During World War II and for several years subsequently, the Fed's open market operations were directed entirely at maintaining low and stable yields on U.S. Treasury securities, while discount window borrowing and Fed acceptance purchases remained minimal. An agreement between the Federal Reserve and the Treasury in March 1951 (the "Accord") freed the Fed from rigid support of U.S. Treasury security prices, enabling the System to pursue broader policy objectives (Hetzel and Leach, 2001a). Under William McChesney Martin Jr., who became Chairman of the Federal Reserve's Board of Governors following the Accord, the Fed initiated an active monetary policy designed to limit inflation and the amplitude of business cycles (Hetzel and Leach, 2001b). To achieve those goals, the Fed has relied primarily on open market operations in U.S. government securities to manipulate the volume of bank reserves and influence market interest rates. Although at times the quantity of discount window borrowing has been an operational target for open market policy, discount loans have been a far less important source of Federal Reserve credit since 1951 than they were before 1934, as have Fed purchases of bankers acceptances (see Figure 5).<sup>25</sup>

Whereas the Great Depression was a defining moment in the conduct of U.S. monetary policy (Calomiris and Wheelock, 1998), it perhaps had even more impact on the regulation of the financial system and the government's role in credit allocation. A host of federal loan corporations and other agencies to allocate credit, such as the Reconstruction Finance Corporation, were founded or expanded. The widely held view that stock market speculation and commercial bank involvement in the underwriting, sale, and financing of security purchases had caused the Depression led to fundamental reforms of securities markets and the banking system, including the Glass-Steagall Act of 1933, which prohibited the commingling of commercial and investment banking.

The Federal Reserve also was given expanded powers to influence the allocation of credit. The

Federal Reserve Act was amended in 1933 to authorize the Fed to set minimum margin requirements for stock market loans, while giving the Federal Reserve Board clear authority to deny discount window loans to banks that made speculative loans. At the same time, the definition of acceptable collateral for discount window loans was broadened and the Fed was authorized under certain circumstances to make loans to nonmember banks, groups of banks, and even to individuals, partnerships, and corporations (Hackley, 1973).

Since the Accord, the Federal Reserve has effectively insulated its monetary policy from credit allocation. Discount window lending has been a small fraction of total Federal Reserve credit, and the Fed largely discontinued the purchase of bankers acceptances in the 1950s, though authorization to purchase acceptances was not eliminated until 1998. Moreover, discount window loans and other Federal Reserve transactions, such as foreign exchange market intervention and warehousing, are prevented from affecting the monetary base by means of offsetting open market operations.

Goodfriend (1994) contends that pressure to allocate credit could still be detrimental to monetary policy. At times, Congress and the Administration have called upon the Fed to lend to distressed firms and governments, such as Penn Central Corporation in 1970 and New York City in 1975 (see Schwartz, 1992). Although the Fed has usually resisted such calls, Goodfriend (1994) argues that pressure put on the Fed to conduct targeted credit policy threatens the Fed's independence, which he views as crucial to the conduct of effective monetary policy. Arguably, if the Fed were to rely more heavily on discount window lending or to conduct open market operations in assets other than U.S. Treasury securities, the System could face intensified pressure to alter the composition of its asset portfolio. The experience of the Fed during the Great Depression suggests that a desire to affect the allocation of credit, even one

<sup>24</sup> The Fed's few open market purchases maintained a constant-size portfolio as holdings matured. Gold inflows from abroad poured reserves into the U.S. banking system, however, and banks amassed high levels of reserves in excess of legal requirements. This "golden avalanche" produced rapid growth of the money stock (Friedman and Schwartz, 1963). The Fed did raise bank reserve requirements in 1936 and 1937, fearing the inflationary potential of excess reserves, and lowered them under pressure from the Administration in 1938 when the economy slipped into recession. See Calomiris and Wheelock (1998) for discussion.

<sup>25</sup> See Meulendyke (1989) for an overview of monetary policy since the Accord.



that originates from within the Federal Reserve System, could undermine its monetary policy.

When direct lending to commercial banks was an important source of Federal Reserve credit, Federal Reserve officials became concerned that banks were not employing the reserves they acquired through the discount window appropriately. Elaborate collateral requirements were imposed by the Federal Reserve Act, and at various times the Fed reiterated that borrowing is a privilege, not a right. Still, Fed officials became dissatisfied with the use of credit and sought to impose tight controls to limit borrowing. The result was an exceptionally tight monetary policy that carried over into the Great Depression, when many Fed officials feared that aggressive monetary easing would only reignite financial speculation.

The Federal Reserve is unlikely to repeat the egregious error of contracting Federal Reserve credit and the monetary base during a serious economic downturn. However, if direct lending to financial institutions or open market operations in assets other than U.S. Treasury securities become important in the implementation of monetary policy, the Fed's early history warns that new pressures to conduct a *credit* policy could arise that might hamper the conduct of *monetary* policy.

## REFERENCES

- Board of Governors of the Federal Reserve System. *Banking and Monetary Statistics, 1914-1941*. Washington, DC, 1943.
- \_\_\_\_\_. *The Federal Reserve System: Purposes and Functions*. Washington, DC, 1994.
- Broaddus, J. Alfred Jr. and Goodfriend, Marvin. "What Assets Should the Federal Reserve Buy?" Federal Reserve Bank of Richmond *2000 Annual Report*, 2001.
- Broz, J. Lawrence. *The International Origins of the Federal Reserve System*. Ithaca, NY: Cornell University Press, 1997.
- Brunner, Karl and Meltzer, Allan H. "What Did We Learn from the Monetary Experience of the United States in the Great Depression?" *Canadian Journal of Economics*, May 1968, pp. 334-48.
- Burgess, W. Randolph. *The Reserve Banks and the Money Market*. New York: Harper and Brothers, 1936.
- Calomiris, Charles W. and Wheelock, David C. "Was the Great Depression a Watershed for American Monetary Policy?" in Michael D. Bordo, Claudia Goldin, and Eugene N. White, eds., *The Defining Moment: The Great Depression and the American Economy in the Twentieth Century*. Chicago: University of Chicago Press, 1998, pp. 23-66.
- Chandler, Lester V. *Benjamin Strong: Central Banker*. Washington, DC: The Brookings Institution, 1958.
- \_\_\_\_\_. *American Monetary Policy, 1928-1941*. New York: Harper and Row, 1971.
- Congressional Budget Office. *The Budget and Economic Outlook*. August 2001. <www.cbo.gov> (as posted July 11, 2001).
- Dupont, Dominique and Sack, Brian. "The Treasury Securities Market: Overview and Recent Developments." Board of Governors of the Federal Reserve System *Federal Reserve Bulletin*, December 1999, pp. 785-806.
- Dwyer, Gerald P. Jr. and Gilbert, R. Alton. "Bank Runs and Private Remedies." Federal Reserve Bank of St. Louis *Review*, May/June 1989, 71(3), pp. 43-61.
- Friedman, Milton and Schwartz, Anna J. *A Monetary History of the United States, 1867-1960*. Princeton, NJ: Princeton University Press, 1963.
- Gilbert, R. Alton. "Did the Fed's Founding Improve the Efficiency of the U.S. Payments System?" Federal Reserve Bank of St. Louis *Review*, May/June 1998, 80(3), pp. 121-42.
- Goodfriend, Marvin. "Why We Need an 'Accord' for Federal Reserve Credit Policy: A Note." *Journal of Money, Credit, and Banking*, August 1994, 26(3), pp. 572-80.
- Greenspan, Alan. "The Paydown of Federal Debt." Remarks before the Bond Market Association, White Sulphur Springs, West Virginia, 27 April 2001.
- Hackley, Howard H. *Lending Functions of the Federal Reserve Banks: A History*. Washington, DC: Board of Governors of the Federal Reserve System, 1973.
- Hetzel, Robert L. and Leach, Ralph F. "The Treasury-Fed Accord: A New Narrative Account." Federal Reserve Bank of Richmond *Economic Quarterly*, Winter 2001a, 87, pp. 33-55.
- \_\_\_\_\_. and \_\_\_\_\_. "After the Accord: Reminiscences on the Birth of the Modern Fed." Federal

- Reserve Bank of Richmond *Economic Quarterly*, Winter 2001b, 87, pp. 57-64.
- Hamilton, James D. "Monetary Factors in the Great Depression." *Journal of Monetary Economics*, March 1987, 19(2), pp. 145-69.
- Kliesen, Kevin L. and Thornton, Daniel L. "The Expected Federal Budget Surplus: How Much Confidence Should the Public and Policymakers Place in the Projections?" Federal Reserve Bank of St. Louis *Review*, March/April 2001, 83(2), pp. 11-24.
- Meltzer, Allan H. "Why Did Monetary Policy Fail in the 1930s?" Working paper, 1994.
- \_\_\_\_\_. "New Procedures, New Problems." Working paper, 1997.
- Meulendyke, Ann-Marie. *U.S. Monetary Policy and Financial Markets*. New York: Federal Reserve Bank of New York, 1989.
- Schwartz, Anna J. "Understanding 1929-1933," in Karl Brunner, ed., *The Great Depression Revisited*. Boston: Martinus Nijhoff, 1981, pp. 5-48.
- \_\_\_\_\_. "The Misuse of the Fed's Discount Window." Federal Reserve Bank of St. Louis *Review*, September/October 1992, 74(5), pp. 58-69.
- U.S. Senate, Committee on Banking and Currency, 71st Congress, 3rd Session. *Operation of the National and Federal Reserve Banking Systems*. Washington, DC: Government Printing Office, 1931.
- West, Robert C. *Banking Reform and the Federal Reserve, 1863-1923*. Ithaca, NY: Cornell University Press, 1977.
- Wheelock, David C. "Member Bank Borrowing and the Fed's Contractionary Monetary Policy During the Great Depression." *Journal of Money, Credit, and Banking*, November 1990, 22, pp. 409-26.
- \_\_\_\_\_. *The Strategy and Consistency of Federal Reserve Monetary Policy, 1924-1933*. Cambridge: Cambridge University Press, 1991.
- \_\_\_\_\_. "Monetary Policy in the Great Depression: What the Fed Did and Why." Federal Reserve Bank of St. Louis *Review*, March/April 1992, 74(2), pp. 3-28.
- White, Eugene N. *The Regulation and Reform of the American Banking System, 1900-1929*. Princeton: Princeton University Press, 1983.
- Wicker, Elmus R. *Federal Reserve Monetary Policy, 1917-1933*. New York: Random House, 1966.



# Unemployment Insurance Claims and Economic Activity

William T. Gavin and Kevin L. Kliesen

**A**lthough the Federal Open Market Committee (FOMC) monitors a large number of economic series when deciding whether to alter the current stance of its policy, it is generally accepted that policymakers, as well as financial markets, pay especially close attention to labor market indicators during periods of economic uncertainty. The reason, in short, is that changes in labor market activity are thought to be useful predictors for changes in real gross domestic product (GDP), the broadest measure of economic activity.

The main indicators of activity in the labor market include the civilian unemployment rate, nonfarm payroll employment, and average weekly hours, all of which are reported monthly in the *Employment Situation* from the Bureau of Labor Statistics (BLS). Indeed, the release of the monthly employment report seemingly rivals the post-FOMC meeting press release as the single most anticipated economic event in the financial markets. Given its significance, therefore, it is probably not too surprising that economists and market participants try to anticipate changes in this and other labor market indicators.

When it comes to forecasting monthly changes in the unemployment rate or the number of new nonfarm jobs created or destroyed, it appears that many economists and market participants pay particularly close attention to the report on initial unemployment insurance claims. This report, which is published by the Employment Training Administration (ETA), an agency within the U.S. Department of Labor, attempts to measure, on a weekly basis, labor flows from the ranks of the employed to the ranks of the unemployed (initial claims). The report also measures the total number of people currently unemployed who are eligible to receive unemployment insurance benefits (continuing claims).

We begin with a brief review of the important monthly labor market data and their usefulness to economists, policymakers, and financial market participants. We then examine whether these labor market indicators are useful for predicting concurrent growth rates of real GDP. Finally, the paper will examine whether there is significant information to be gleaned from weekly changes in initial and continuing unemployment claims for predicting these monthly labor market indicators.

## LABOR MARKET DATA

There are three major sources of data for the labor market: the household survey, the establishment survey, and the reports of state agencies that collect information about employment for the unemployment insurance program. The former two comprise the information that is found in the monthly employment report, while the latter is the source of information for the weekly unemployment insurance claims data.

### The Household Survey

The household survey collects information from a small but representative sample of households. Currently, about 60,000 households are surveyed either in person or by telephone each month by the Bureau of Census. This survey, although it comprises less than 0.06 percent of the roughly 107 million households in the United States, is meant to be a representative sample of the U.S. civilian noninstitutional population, from which trends in labor market activity can be inferred. From that survey, known as the Current Population Survey (CPS), the BLS culls information on the demographics of the job market, such as race, age, sex, educational level, and detailed information about those who are unemployed, such as the duration of their unemployment.

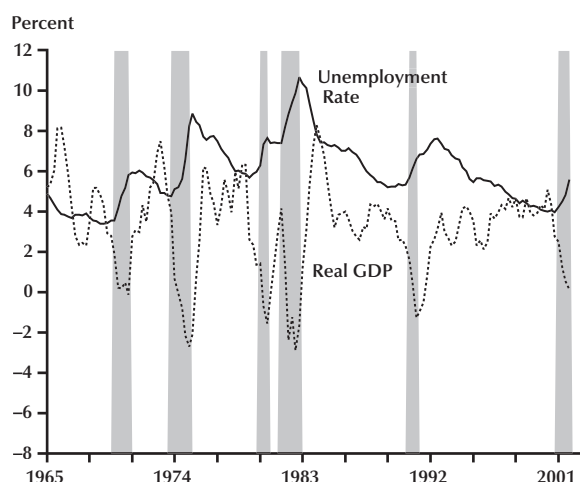
The most important information from the CPS is the unemployment rate, which is plotted in Figure 1. Here, the monthly unemployment rates are averaged to get a quarterly rate. Since there is thought to be a significant cyclical relationship between changes in the unemployment rate and changes in aggregate output, also included is the four-quarter growth rate of GDP.<sup>1</sup> Visual evidence

---

William T. Gavin is a vice president and economist and Kevin L. Kliesen is an economist at the Federal Reserve Bank of St. Louis. The authors thank Cynthia Ambler of the Department of Labor for providing information on the unemployment insurance program. Rachel J. Mandal and Thomas A. Pollmann provided research assistance.

---

<sup>1</sup> The relationship between real GDP growth and the unemployment rate is sometimes characterized by Okun's law. Named after the late economist Arthur Okun, the "law" says that for every percentage point that real GDP growth is above (below) its potential growth, the unemployment rate will fall (rise) by one-half a percentage point. See Mankiw (1998).

**Figure 1****Unemployment Rate and Real GDP**

NOTE: The unemployment rate is a quarterly average of monthly rates. Real GDP is shown as a four-quarter growth rate. Bars indicate periods of recession.

suggests that, during recessions, the unemployment rate usually rises as real GDP declines. At other times, though, the relationship does not hold very well, suggesting that trends in the unemployment rate are not a reliable indicator of GDP growth. Indeed, as shown in Table 1, the correlation between the four-quarter growth of real GDP and the contemporaneous value of the unemployment rate is negative, but with a relatively low value ( $-0.27$ ). One reason why the two series might not be more closely correlated is that the unemployment rate lags the cycle.<sup>2</sup>

Another reason why changes in the underlying trend of the unemployment rate appear unrelated to the business cycle is the influence of microeconomic factors. These include changes in the benefits associated with being unemployed, changes in the demographics of the labor force, and cultural changes in family structure and work habits. Regarding the latter two factors, the large increase in the unemployment rate in the late 1960s and 1970s was associated with a growing number of young workers and women entering the labor force. And since the unemployment rates for young workers and women were higher than the average, this change in the composition of the labor force was associated with a rising trend in the unemployment rate. In the 1990s, as the baby boomers aged (fewer young people entering the labor force) and the labor force

participation rate of women approached the rate of men, the unemployment rate has gradually declined.

### **The Establishment Survey**

The establishment survey, also known as the Current Employment Statistics (CES) program, includes labor input information from about 350,000 nonagricultural establishments that employ about 39 million people. (Establishments are not the same as firms, but rather, they are distant parts of a firm in different locations. For example, the Federal Reserve Bank of St. Louis is a firm with establishments in St. Louis, Little Rock, Louisville, and Memphis.) The time series we use on payroll jobs and hours worked come from the CES. The data on employment growth from the CES are considered to be more accurate than the data from the CPS because the establishment survey has much greater coverage. Although the establishments surveyed are not representative, they nonetheless are the largest establishments and account for about 30 percent of the workforce (compared with 0.06 percent for the household survey).<sup>3</sup>

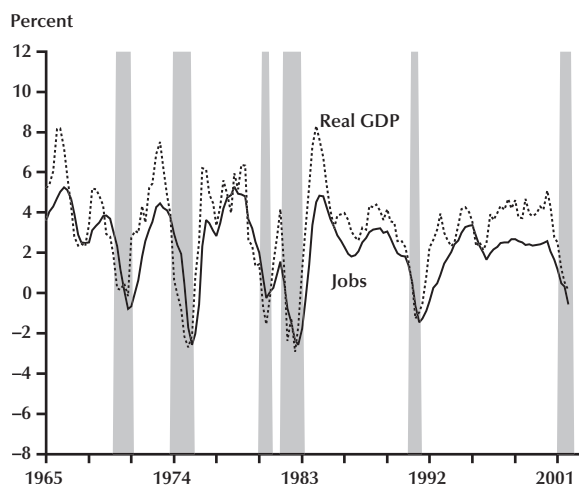
Figure 2 shows the four-quarter growth rate in jobs as well as the four-quarter growth rate in GDP. As can be seen visually in Figure 2, and statistically in Table 1, there is a much closer correlation between jobs growth and GDP growth than there is between the unemployment rate and GDP growth. The correlation between the four-quarter growth of real GDP and nonfarm payroll jobs is high, 0.79. Here the cycles appear to coincide. However, there are sustained periods of productivity growth during which the economy grows faster than the work force. Most obvious in the chart are the decade of the 1960s and the five years following 1995. It appears that these periods of high productivity growth tend to occur during expansions.

The other major series that comes from the establishment survey is the index of hours worked.<sup>4</sup>

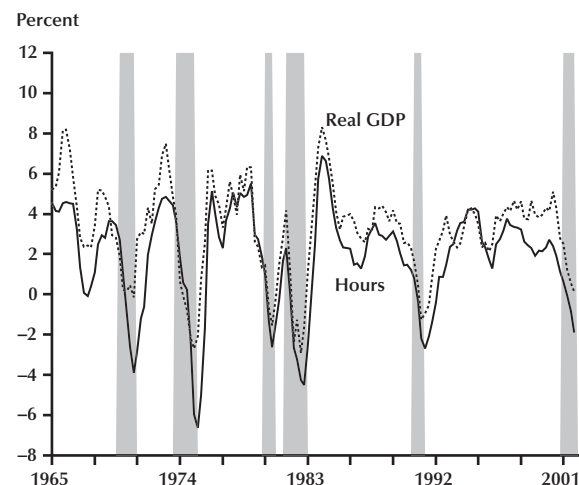
<sup>2</sup> Further evidence of this assertion is that the average duration of unemployment is included in the Conference Board's list of lagging indicators. Its weight places it seventh out of seven in terms of its contribution to the index.

<sup>3</sup> The results from this large sample are adjusted for the bias that exists between the composition of the approximately 350,000 large establishments surveyed and the composition of the roughly five million smaller establishments that are not included in the survey. This bias adjustment process, as it is known, is being replaced with a completely different methodology. See Getz (2000).

<sup>4</sup> The index of aggregate hours worked is the product of average weekly hours and employment of production or nonsupervisory workers. See BLS *Handbook of Methods*.

**Figure 2****Nonfarm Payroll Jobs and Real GDP**

NOTE: Both payroll jobs and real GDP are shown as four-quarter growth rates. Bars indicate periods of recession.

**Figure 3****Nonfarm Hours Worked and Real GDP**

NOTE: Both hours worked and real GDP are shown as four-quarter growth rates. Bars indicate periods of recession.

**Table 1****Cross Correlations Between Real GDP Growth and Growth of Labor Market Variables**

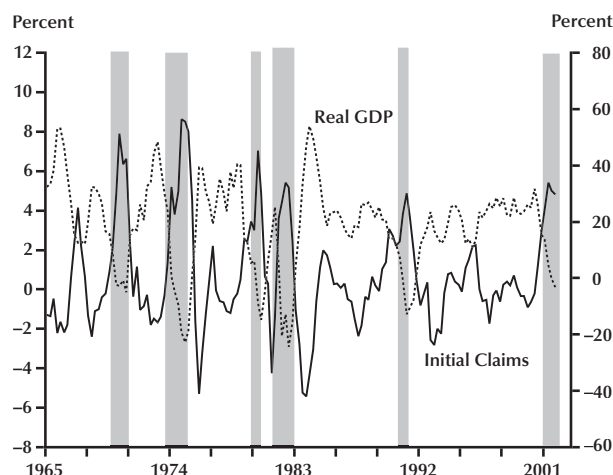
|                   | Continuing claims | Real GDP | Hours | Initial claims | Jobs  | Unemployment rate |
|-------------------|-------------------|----------|-------|----------------|-------|-------------------|
| Continuing claims | 1                 | -0.82    | -0.86 | 0.90           | -0.70 | 0.14              |
| Real GDP          |                   | 1        | 0.87  | -0.79          | 0.79  | -0.27             |
| Hours             |                   |          | 1     | -0.69          | 0.94  | -0.37             |
| Initial claims    |                   |          |       | 1              | -0.50 | 0.05              |
| Jobs              |                   |          |       |                | 1     | -0.48             |
| Unemployment rate |                   |          |       |                |       | 1                 |

NOTE: Correlations of four-quarter growth rates except for the unemployment rate, which is in levels.

Figure 3 shows that growth in hours worked also moves closely with output growth over the business cycle. Indeed, among the labor variables cited earlier, the cross-correlations reported in Table 1 show that the growth of hours worked has the highest correlation with the growth of real GDP (0.87 over the sample period in Figure 3). Like jobs growth, the movement in the growth of aggregate hours is procyclical and appears to coincide with output growth.

One of the reasons why the monthly labor report from the establishment survey is considered so important is because it provides early information

about GDP growth. To understand why this is so, note that a given month's report is released on the first Friday of the next month. For example, on the first Friday of each January, the Department of Labor will release information about the labor market in the previous December. The market will already have received labor market data for October and November. Labor data for the fourth quarter will be available in the first week of January, but the Department of Commerce will not release the advance estimate of fourth-quarter GDP growth until the last week of January.

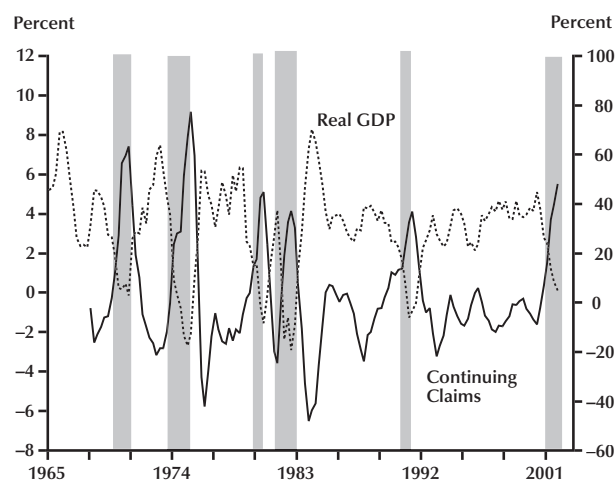
**Figure 4****Initial Claims and Real GDP**

NOTE: Both initial claims and real GDP are shown as four-quarter growth rates. Bars indicate periods of recession.

The initial release of payroll jobs and hours worked is based on the establishment survey. According to the BLS, the most recent two months of estimates from the establishment data are considered preliminary because not all of the surveys have been returned and processed. Conceivably, then, the BLS may report up to three different estimates (current plus two subsequent revisions) of nonfarm job gains or losses for any month. But even these are still only preliminary, since the data on jobs and hours will be revised the following year with the annual benchmark revisions. The purpose of the benchmark revisions is to tie together the *sample-based* estimates that underpin the monthly establishment data with the actual “universe” counts of jobs, wages, and earnings that are reported to employment security agencies of the 50 states, the District of Columbia, Puerto Rico, and the Virgin Islands. Thus, the third source of information about the labor market is that reported to the Department of Labor by the state agencies that administer the federal-state unemployment insurance program.

### **Covered Employment and Wages Program**

This program, also known as the ES-202 program, is a joint venture between the BLS and the state employment security agencies. The purpose of the

**Figure 5****Continuing Claims and Real GDP**

NOTE: Both continuing claims and real GDP are shown as four-quarter growth rates. Bars indicate periods of recession.

program is to provide a comprehensive accounting of nonagricultural employment and wage data by industry at the national, state, and local levels. Thus, coverage under the ES-202 program is nearly universal. In 1994, more than 96 percent of all wage and salary civilian jobs were covered by the ES-202 program, while covered employees accounted for nearly 93 percent of the wage and salary component of national income. Those excluded from the program coverage include agricultural workers, the military, and segments of state and local government employees.

This statewide information is aggregated to the national level by the ETA. Each week, the ETA releases statistics for the number of individuals filing new or continuing claims under the unemployment insurance (UI) program. The UI program is a joint arrangement between the federal government and individual state governments. Its purpose is to provide temporary unemployment benefits to eligible recipients. Though there are some common characteristics, each state operates under its own laws and, accordingly, sets program eligibility requirements. See the appendix for more detail on the program and its eligibility requirements.

Figure 4 shows the widely reported series, initial claims for unemployment insurance. Because growth in initial claims is much more variable than real GDP, their growth rates are shown on the right-hand



scale in Figure 4. Initial claims are clearly countercyclical. Table 1 shows a high negative correlation between the four-quarter growth rates of the two series,  $-0.79$ . Despite this high correlation, the National Bureau of Economic Research does not place much weight on initial claims when it comes to determining business cycle peaks and troughs. In a question-and-answer section in the on-line issue of *The NBER's Recession Dating Procedure* dated October 8, 2001, the following was posted:

- Q: How do the movements of unemployment claims inform the Bureau's thinking?  
 A: A bulge in jobless claims would appear to forecast declining employment, but we don't use forecasts and the claims numbers have a lot of noise.

The weekly initial claims report also includes a series on those individuals that continue to draw unemployment compensation, otherwise known as continuing claims. Figure 5 shows the four-quarter growth of continuing claims and real GDP. Continuing claims are also much more variable than GDP, and their growth rates are shown in Figure 5 on the right-hand scale. Like initial claims, continuing claims are also countercyclical. Visually, it is difficult to distinguish the co-movement between initial and continuing claims, although growth of the latter appear to vary less. Either way, their correlations with GDP growth are virtually identical, as seen in Table 1.

## THE LABOR MARKET AND GDP

From both a theoretical and empirical standpoint, the labor market is an important element in the economy. Output production requires combinations of labor, land, capital, and other factors. About two-thirds of the payments for factors go to the labor component. From the point of view of data collection, perhaps our best measure of economic activity is a measure of the number of people working.

As we saw in Figures 1 through 5, labor market indicators move in tandem with output over the business cycle. There is a considerable literature showing that monthly data in general, and labor data in particular, can be used to predict current quarter GDP. Miller and Chin (1996) survey this literature and report their own research showing that monthly information about hours worked helps predict GDP growth. Recently, Koenig, Dolmas, and Piger (2001) reported that monthly employment

growth is a significant predictor of current-quarter GDP growth. Some private sector economists have even developed a "real time" model of aggregate economic activity that uses both initial and continuing claims to predict *monthly* changes in real GDP.<sup>5</sup> Accordingly, monthly data should be able to predict quarterly GDP because the Bureau of Economic Analysis (BEA) uses monthly labor market data as an input to formulas that are used to estimate GDP components.

We evaluate the predictive content of monthly labor market data by adding these variables one at a time to a univariate autoregressive model of real GDP growth. We construct quarterly time series of incoming monthly labor market data. We forecast the current-quarter real GDP growth rate using incoming monthly labor market data from the same quarter. The general form of the forecasting model is

$$(1) \quad y_t = c + \beta_j LM_{j,t}^k + \sum_{i=1}^4 \delta_i y_{t-i} + \varepsilon_t,$$

where  $y_t = \ln(GDP_t/GDP_{t-1}) \times 400$ , the average annualized growth rate of GDP in the current quarter, and  $LM_j^k$  is one of five labor market variables measured at the end of each of the three months in the quarter. The five labor market variables indexed by  $k$  include the unemployment rate and the annualized growth rates of payroll jobs, aggregate hours worked, initial claims for unemployment insurance, and continuing claims for unemployment insurance. The labor market variable is indexed by  $j$  to indicate which month of the current quarter is being used in the forecast. For example, at the end of the first month,  $LM_{1,t}^1$  is the newly reported unemployment rate; at the end of the second month,  $LM_{2,t}^1$  is the average of unemployment rates for the first two months of the quarter; and, at the end of the third month,  $LM_{3,t}^1$  is the average of the three months of quarter  $t$ . Remember that the labor report for the third month of a quarter arrives three to four weeks before the first GDP report for that quarter. We consider each labor market variable separately. We also include four lags of GDP growth.

We use current vintage data in this forecasting experiment.<sup>6</sup> In this experiment, we begin by estimating a model using data from 1967:Q2 through

<sup>5</sup> See Hatzius (2001).

<sup>6</sup> Koenig, Dolmas, and Piger (2001) have shown that it is possible to get a better forecast by using real-time vintage data. They show analytically that, if the revisions to data are not predictable, then the real-time vintage data will yield a forecast model with a smaller out-of-sample forecast error than one would get using current vintage data.



1991:Q3.<sup>7</sup> This model is then used to forecast the fourth quarter of 1991. We then update the forecasting model with 1991:Q4 data and use the newly estimated model to forecast the second quarter. That is, we update the model recursively and tabulate the forecasts through 2001:Q3. We then calculate the root-mean-squared error (RMSE) for that model forecast.

We examine six models. The first is simply the autoregressive (AR) component of the model excluding the labor market variable. The next five correspond to the labor market variables. The civilian unemployment rate is measured in level form, whereas the remaining four variables are measured as annualized growth rates. Note that in the case of these four variables, we calculate the first month's growth rate by taking the log ratio of the variable in the first month to the average of the three months in the previous quarter. In the second month we take the log of the ratio of the average of the first two months in the current quarter to the average of the three months in the previous quarter. In the third month we just look at the ratio of the three-month averages. In all cases involving variables in the GDP forecasting equation, we annualize growth rates.

To assess the statistical significance of the accuracy of the alternative model forecasts, we use two tests developed for nested forecasting models. In each case, we compare a model with lags of GDP growth and a labor market variable with a model that includes only lags of GDP growth. First, we use an out-of-sample F test of the null hypothesis that the model with the labor market variable has no predictive content for real GDP growth once the autoregressive model is taken into account. This test, developed by McCracken (1999), is given by

$$\text{OOS-F} = P \left[ \frac{MSE_{AR} - MSE_{LM^k}}{MSE_{LM^k}} \right],$$

where OOS-F is the out-of-sample F test,  $P$  is the number of forecasts made,  $MSE_{AR}$  is the mean-squared error for the AR model forecasts, and  $MSE_{LM^k}$  is the mean-squared error for the models that include labor market variables. McCracken derives the limiting distribution of this test statistic under the null hypothesis and reports percentiles of the OOS-F statistic. He derives tables under alternative methods of updating the forecasting models. We use a recursive scheme. The critical values of the test statistic depend on which scheme is used and two other factors: (i) the number of labor market variables

included (in each case we have one in each model) and (ii) the ratio ( $P/R$ ) of the number of forecasts ( $P$ ) to the number of observations used to estimate the model that was used to make the first forecast ( $R$ ). Percentiles of the distribution are listed in the tables. We are comparing nested models so we use a one-sided test. When the MSE of the forecasts from the unrestricted model is larger than the MSE from the restricted model, this test statistic is negative.

The second test we use is an out-of-sample test for encompassing. (Encompassing is simply that, if one forecast incorporates all of the relevant information, then adding information from the other forecast will not help predict the actual value.) We use an encompassing test of the null hypothesis that the AR model encompasses the model augmented with the labor market variable. This test, developed by Clark and McCracken (2000), is given by

$$\text{ENC-CM} = P \left[ \frac{MSE_{AR} - MCPE}{MSE_{LM^k}} \right],$$

where ENC-CM is the encompassing test proposed by Clark and McCracken (2000) and MCPE is the mean cross product of the forecast errors from the restricted (AR) and unrestricted (LM) models.<sup>8</sup> Clark and McCracken derive the limiting distribution of this test statistic under the null hypothesis and report percentiles of the ENC-CM statistic. As in the case of the OOS-F statistic, the limiting distribution depends on the method used to update the forecasting models, the number of parameters restricted to zero, and the ratio,  $P/R$ . The percentiles of the distribution are shown in Table 2. Again, we are comparing nested models, so we use a one-sided test. The statistic will be negative only if the average cross product is positive and larger than the mean-squared error of the forecasts from the AR model.

The results of our evaluation are shown in Table 2. The first column reports results for the AR model (which excludes contemporaneous labor market data). This is the benchmark model and is nested in all the others. The RMSE of the forecast from the AR model for the period from 1990:Q1 to 2001:Q3 is 2.17 percent, with an adjusted  $R^2$  of 6 percent for the last model estimated—that is, estimated over the period from 1967:Q2 to 2001:Q2.

<sup>7</sup> Except in the case where the labor market variable is continuing claims. These data begin in January 1968.

<sup>8</sup> The MCPE is calculated by the following formula:

$$\frac{1}{T} \sum_{t=1}^T \hat{e}_{it} \hat{e}_{jt}.$$

**Table 2****Evaluation of GDP Forecasts (1990:Q1 through 2001:Q3)**

|   | AR (4)<br>model | Unemployment<br>rate | Payrolls<br>jobs | Hours<br>worked | Initial<br>claims | Continuing<br>claims |
|---|-----------------|----------------------|------------------|-----------------|-------------------|----------------------|
| <b>RMSEs (out-of-sample forecasts)</b>                      |                 |                      |                  |                 |                   |                      |
| First month   | 2.17            | 2.17                 | 1.81             | 1.95            | 2.18              | 2.23                 |
| Second  |                 | 2.17                 | 1.81             | 1.76            | 2.03              | 2.07                 |
| Third   |                 | 2.17                 | 1.78             | 1.66            | 1.93              | 1.93                 |
| Adjusted R <sup>2</sup> *                                   | 0.06            | 0.06                 | 0.57             | 0.63            | 0.41              | 0.47                 |
| <b>McCracken out-of-sample F test<sup>†</sup></b>           |                 |                      |                  |                 |                   |                      |
| First month   |                 | -0.31                | <b>20.32</b>     | <b>10.75</b>    | -0.40             | -2.64                |
| Second  |                 | -0.28                | <b>20.31</b>     | <b>24.34</b>    | <b>6.39</b>       | <b>4.48</b>          |
| Third   |                 | -0.28                | <b>22.29</b>     | <b>32.65</b>    | <b>12.48</b>      | <b>12.25</b>         |
| <b>Clark-McCracken nested encompassing test<sup>‡</sup></b> |                 |                      |                  |                 |                   |                      |
| First month   |                 | 0.11                 | <b>26.75</b>     | <b>24.42</b>    | <b>6.38</b>       | <b>6.44</b>          |
| Second  |                 | 0.00                 | <b>31.49</b>     | <b>35.05</b>    | <b>13.98</b>      | <b>12.17</b>         |
| Third   |                 | -0.07                | <b>34.60</b>     | <b>39.86</b>    | <b>20.65</b>      | <b>17.87</b>         |

NOTE: \*Adjusted R<sup>2</sup> is for the full 1967:Q1 to 2001:Q2 sample using the 3-month models.

<sup>†</sup>The null hypothesis is that the AR model is more accurate than the model with the labor market variable. Here  $P/R=0.52$ . For  $P/R=0.4$ , the 99th, 95th, and 90th percentiles for the OOS-F tests are 2.768, 1.298, and 0.814, respectively; and for  $P/R=0.6$ , the 99th, 95th, and 90th percentiles for the OOS-F tests are 3.719, 1.554, and 0.796, respectively.

<sup>‡</sup>The null hypothesis is that the AR model encompasses the model with the labor market variable. Here  $P/R=0.52$ . For  $P/R=0.4$ , the 99th, 95th, and 90th percentiles for the Clark-McCracken encompassing tests are 2.098, 1.079, and 0.685, respectively; and for  $P/R=0.6$ , the 99th, 95th, and 90th percentiles for the encompassing tests are 2.662, 1.312, and 0.791, respectively.

Bold values indicate that the null hypothesis is rejected at the 99th percentile.

The next column shows the results using the unemployment rate. As was suggested by Figure 1, changes in the unemployment rate since 1990 do not appear to help predict current-quarter real GDP growth. The explanatory power of unemployment was no better than with GDP alone, and the out-of-sample forecasts were slightly worse, although the difference is small. The OOS-F statistics in the middle section of Table 2 are all negative, thus we cannot reject the hypothesis that the AR model is more accurate than the model that includes the unemployment rate. The same is true for the ENC-CM statistics. They are all below the 90th percentile value.

The next two columns show the models using growth in jobs and an index of hours worked. Here, there appears to be predictive information in the growth of payroll employment in all three months. Note, however, that the addition of the second month does not result in a lower RMSE for the model that includes payroll jobs. For the aggregate hours model, adding information from the second and third

months lowers the RMSEs. These models also display a much higher in-sample explanatory power than does the model that includes the unemployment rate. For the models that include payroll jobs and hours worked, the OOS-F tests always reject the hypothesis that the AR model is more accurate. We can also reject the hypothesis that the AR model encompasses these models.

Finally, the two series using the unemployment insurance data lead to lower RMSEs only in the cases with two and three months of claims data included. Here the difference is large enough so that we can reject the null hypothesis that the AR model is more accurate than the models that include two or three months of claims (both initial and continuing). The RMSEs for the models that use initial or continuing claims from the first month only are generally higher than the RMSE from the AR model. The adjusted R<sup>2</sup> values for the models with three months of initial and continuing claims are 0.41 and 0.47, respectively. Looking at the encompassing tests in the bottom

section of Table 2, we see that we can reject the hypothesis that the AR model encompasses the model augmented with initial claims, even when the RMSEs are larger than the benchmark case.

In summary, we find that—consistent with previous empirical research—labor market data does help to predict GDP growth. In the next section, we examine the ability of weekly data on initial and continuing claims to predict the monthly time series on unemployment, payroll jobs, and the index of hours worked.

### PREDICTING MONTHLY LABOR MARKET DATA USING UNEMPLOYMENT INSURANCE DATA

In the previous section we saw that initial and continuing claims for unemployment insurance are not very useful for predicting real GDP growth during the concurrent quarter. However, data on monthly employment and hours worked did help to predict GDP growth. Therefore, it would be useful to be able to predict employment and hours worked using the weekly claims data. Furthermore, many economists and financial analysts use weekly claims data to predict monthly changes in the unemployment rate. The payoff from this exercise is potentially quite large, since unexpected changes in the unemployment rate can be a significant market mover; moreover, these changes can sometimes induce immediate changes in monetary policy.<sup>9</sup> A typical example of the analysis that posits a causality between unemployment insurance claims and the unemployment rate may be found in the following *Monetary Policy Report to the Congress*:

Employment continued to decline in December and January but much less than in the preceding two months. Manufacturing and its related industries lost jobs at a slower pace, and employment leveled off in other private industries. The unemployment rate moved up to 5.8 percent in December but then ticked down to 5.6 percent in January. The recent reversal of the October and November spikes in new claims for unemployment insurance and in the level of insured unemployment also point to some improvement in labor market conditions early this year. (Board of Governors of the Federal Reserve System, February 2002, p. 20)

A recent study by Montgomery et al. (1998) uses monthly initial claims data to forecast the quarterly unemployment rate. The study finds some support for the predictive content of monthly initial claims.<sup>10</sup> The contribution of initial claims was concentrated in periods when unemployment was rising. McConnell (1998) reports a similar finding in which she uses initial claims data to forecast payroll jobs growth. In her study, initial claims helped to predict payroll jobs, but only during periods of recession. In this study we are looking at the ability of the weekly data to predict the monthly series: not only the unemployment rate, but the jobs and hours worked data as well.

We use a model analogous to equation (1) to evaluate the ability of the unemployment insurance claims data to predict the monthly labor statistics:

$$(2) \quad LM_t^k = c + \beta_j Weekly_{j,t}^a + \sum_{i=1}^{12} \delta_i LM_{t-i}^k + \varepsilon_t,$$

where the dependent variable is one of three monthly labor market series: the unemployment rate, growth in payroll jobs, and growth in the index of hours worked. Here, the growth rates are monthly. There are two alternative weekly series used on the right-hand side of equation (2): initial claims and continuing claims. The data on initial claims are released on Thursdays and apply to the previous week that ended five days earlier. The data on continuing claims released at the same time apply to the week that ended 12 days earlier.

We create five monthly series from each of these latter two data series, initial and continuing claims. The first weekly series is the data reported on the first Thursday following the first Friday of the month, the normal release date for the *Employment Situation*. We take the logarithm of the ratio of this weekly release to the average for the previous month. The second weekly series is the logarithm of the ratio of the average data reported on the first and second Thursdays (following the first Friday of the month) to the previous month's average, and so forth. We do not create a fifth series because there is not always a fifth Thursday. Instead, we create a series that we call the last week, which includes the

<sup>9</sup> See Jordan (1992).

<sup>10</sup> They started with seasonally adjusted data, but as usual in these time-series models, they had to include the seasonal terms to get rid of the residual correlation. We briefly examined ARIMA and Bayesian VAR methods, but, overall, none generate more accurate forecasts than the univariate regressions reported herein.

**Table 3****Regression Output for Period from February 1968 to November 2001**

|                          | Initial claims |             |       | Continuing claims |             |       |
|--------------------------|----------------|-------------|-------|-------------------|-------------|-------|
|                          | $\beta$        | t Statistic | SEE   | $\beta$           | t Statistic | SEE   |
| <b>Unemployment rate</b> |                |             |       |                   |             |       |
| AR only                  | —              | —           | 0.161 | —                 | —           | 0.161 |
| First week               | 0.003          | 2.11        | 0.161 | 0.021             | 4.27        | 0.158 |
| Second                   | 0.007          | 3.81        | 0.159 | 0.038             | 8.29        | 0.149 |
| Third                    | 0.008          | 4.92        | 0.157 | 0.038             | 9.27        | 0.146 |
| Fourth                   | 0.009          | 5.28        | 0.156 | 0.036             | 10.01       | 0.144 |
| Last                     | 0.009          | 5.36        | 0.156 | 0.036             | 10.33       | 0.143 |
| <b>Payroll jobs</b>      |                |             |       |                   |             |       |
| AR only                  | —              | —           | 0.171 | —                 | —           | 0.171 |
| First week               | -0.006         | -4.02       | 0.168 | -0.030            | -5.93       | 0.164 |
| Second                   | -0.012         | -6.82       | 0.162 | -0.043            | -8.95       | 0.156 |
| Third                    | -0.014         | -7.87       | 0.159 | -0.045            | -10.95      | 0.150 |
| Fourth                   | -0.014         | -8.20       | 0.158 | -0.044            | -12.34      | 0.145 |
| Last                     | -0.014         | -8.38       | 0.157 | -0.043            | -12.88      | 0.143 |
| <b>Hours worked</b>      |                |             |       |                   |             |       |
| AR only                  | —              | —           | 0.477 | —                 | —           | 0.477 |
| First week               | -0.012         | -2.63       | 0.474 | -0.061            | -4.13       | 0.468 |
| Second                   | -0.024         | -4.89       | 0.464 | -0.087            | -6.20       | 0.456 |
| Third                    | -0.031         | -6.17       | 0.456 | -0.097            | -7.86       | 0.444 |
| Fourth                   | -0.034         | -7.15       | 0.449 | -0.103            | -9.74       | 0.429 |
| Last                     | -0.034         | -7.32       | 0.448 | -0.102            | -10.09      | 0.426 |

NOTE: The values in the table are estimates of  $\beta$ , its t statistic, and the standard error for equation (2).

average of the data released in the first four weeks when there is no fifth week available.

The estimation results for the weekly initial claims models using the full data set are shown in Table 3. The estimation period includes the months from February 1968 through November 2001. We estimated OLS models for the three labor market variables. Each model included a constant and 12 lags of the dependent variable as well as our weekly series that use information about unemployment insurance claims. There are three sections in Table 3. The top section shows the results for the unemployment rate. The first row reports the standard error of the equation (SEE) for the autoregressive (AR) model which excludes the claims data. The estimate of the coefficient on the weekly initial claims data is reported in the first column of results with the t

statistic for that coefficient reported in the second column. The third column reports the SEE for the equation. The last three columns report the analogous results for continuing claims. The middle section includes the results for payroll jobs, and the bottom section reports the results for hours worked. Overall, the in-sample fit improved with the accumulation of information throughout the month. Uniformly, the data on continuing claims do a better job of predicting the labor market variables than do the initial claims data. This condition is true in spite of the extra-week delay in reporting information about continuing claims.

### **An Out-of-Sample Forecasting Exercise**

To evaluate the predictive content of the claims data, we conduct an out-of-sample forecasting

Table 4

### Evaluation of Monthly Forecasts of Labor Market Indicators (Current Month Forecasts from January 1990 to November 2001)

|  | Unemployment rate |                   | Payroll jobs   |                   | Hours worked   |                   |
|--|-------------------|-------------------|----------------|-------------------|----------------|-------------------|
|  | Initial claims    | Continuing claims | Initial claims | Continuing claims | Initial claims | Continuing claims |
| <b>RMSE (% at monthly rates)</b>                 |                   |                   |                |                   |                |                   |
| AR model   |                   | 0.135             |                | 0.106             |                | 0.371             |
| First week                                       | 0.134             | 0.137             | 0.103          | 0.105             | 0.369          | 0.369             |
| Second   | 0.133             | 0.136             | 0.105          | 0.108             | 0.374          | 0.371             |
| Third  | 0.132             | 0.131             | 0.103          | 0.102             | 0.371          | 0.362             |
| Fourth   | 0.133             | 0.130             | 0.102          | 0.097             | 0.371          | 0.357             |
| Last   | 0.133             | 0.130             | 0.102          | 0.097             | 0.372          | 0.357             |
| <b>McCracken out-of-sample F test*</b>           |                   |                   |                |                   |                |                   |
| First week                                       | 1.64              | -4.17             | <b>8.68</b>    | 1.30              | 1.19           | 1.87              |
| Second   | <b>4.37</b>       | -2.41             | 1.04           | -5.52             | -1.90          | -0.29             |
| Third  | <b>5.02</b>       | <b>6.97</b>       | <b>6.43</b>    | <b>9.25</b>       | -0.07          | <b>6.98</b>       |
| Fourth   | 3.52              | <b>10.09</b>      | <b>11.85</b>   | <b>25.32</b>      | 0.22           | <b>11.37</b>      |
| Last   | 2.89              | <b>9.18</b>       | <b>10.11</b>   | <b>25.65</b>      | -0.37          | <b>11.29</b>      |
| <b>Clark-McCracken nested encompassing test†</b> |                   |                   |                |                   |                |                   |
| First week                                       | 1.87              | 2.52              | <b>9.03</b>    | <b>16.30</b>      | 1.85           | <b>6.20</b>       |
| Second   | <b>5.02</b>       | <b>11.45</b>      | <b>17.62</b>   | <b>22.47</b>      | <b>4.56</b>    | <b>9.28</b>       |
| Third  | <b>7.39</b>       | <b>18.60</b>      | <b>23.64</b>   | <b>36.99</b>      | <b>7.80</b>    | <b>15.84</b>      |
| Fourth   | <b>8.09</b>       | <b>20.62</b>      | <b>31.00</b>   | <b>52.32</b>      | <b>11.61</b>   | <b>22.57</b>      |
| Last   | <b>8.05</b>       | <b>20.50</b>      | <b>31.74</b>   | <b>53.58</b>      | <b>12.35</b>   | <b>22.88</b>      |

NOTE: \*The null hypothesis is that the AR model is more accurate than the model with the labor market variable. Here  $P/R=0.58$ . For  $P/R=0.6$ , the 99th, 95th, and 90th percentiles for the OOS-F tests are 3.719, 1.554, and 0.796, respectively.

†The null hypothesis is that the AR model encompasses the model with the labor market variable. Here  $P/R=0.58$ . For  $P/R=0.6$ , the 99th, 95th, and 90th percentiles for the Clark-McCracken encompassing tests are 2.662, 1.312, and 0.791, respectively.

Bold values indicate that the null hypothesis is rejected at the 99th percentile.

experiment. Again, we are using current vintage data to construct these out-of-sample forecasts. We begin by estimating the model over the period from February 1968 through December 1989. As before, we update the model each month before we make the next forecast, recursively computing the forecasts through November 2001. The RMSEs of the forecasts are reported in the top section of Table 4. The first row of results include the RMSEs from the forecasts made by the autoregressive models. Here the sample period includes the months from January 1990 through November 2001.

The out-of-sample forecasting results are not entirely consistent with the in-sample fit where continuing claims always outperformed initial

claims. Here, initial claims appears to do a better job forecasting the unemployment rate and payroll job growth early in the month and continuing claims does better late in the month. In the second and third sections of Table 4, we report the out-of-sample tests for equality of MSEs and encompassing, respectively. Again, we compare the forecasts from the full model with forecasts from the AR model that is nested within each of the full models. Therefore, we use the tests for nested models that were described above. The forecasting method was recursive, there is one restriction on the AR model, and the  $P/R$  ratio for this experiment is 0.58, so we use the percentiles for  $P/R = 0.6$  where the 99th percentile for the OOS-F test statistic is 3.719.



Table 5

**Evaluation of Monthly Forecasts of Labor Market Indicators (Current Month Forecasts from April 1991 to February 2001—Expansion Months Only)**

|  | Unemployment rate |                   | Payroll jobs   |                   | Hours worked   |                   |
|--|-------------------|-------------------|----------------|-------------------|----------------|-------------------|
|  | Initial claims    | Continuing claims | Initial claims | Continuing claims | Initial claims | Continuing claims |
| <b>RMSE (% at monthly rates)</b>                 |                   |                   |                |                   |                |                   |
| AR (12) model                                    | 0.128             |                   | 0.095          |                   | 0.374          |                   |
| First week                                       | 0.127             | 0.132             | 0.092          | 0.097             | 0.371          | 0.373             |
| Second   | 0.126             | 0.133             | 0.097          | 0.103             | 0.375          | 0.379             |
| Third  | 0.126             | 0.128             | 0.097          | 0.100             | 0.374          | 0.375             |
| Fourth   | 0.127             | 0.128             | 0.097          | 0.095             | 0.373          | 0.369             |
| Last   | 0.128             | 0.129             | 0.097          | 0.094             | 0.373          | 0.369             |
| <b>McCracken out-of-sample F test*</b>           |                   |                   |                |                   |                |                   |
| First week                                       | 1.13              | -6.95             | <b>6.82</b>    | -6.05             | 1.93           | 0.70              |
| Second   | <b>3.04</b>       | -8.64             | -5.82          | -19.20            | -0.63          | -3.49             |
| Third  | <b>3.24</b>       | -1.29             | -4.88          | -11.43            | 0.06           | -0.89             |
| Fourth   | 0.56              | -0.93             | -4.64          | -0.50             | 0.13           | <b>3.32</b>       |
| Last   | 0.19              | -1.84             | -6.05          | 0.76              | 0.13           | <b>3.18</b>       |
| <b>Clark-McCracken nested encompassing test†</b> |                   |                   |                |                   |                |                   |
| First week                                       | 2.08              | 1.25              | -8.31          | -5.94             | 1.71           | <b>3.93</b>       |
| Second   | <b>5.86</b>       | <b>8.69</b>       | 0.24           | -0.55             | <b>3.54</b>    | <b>6.11</b>       |
| Third  | <b>7.37</b>       | <b>13.60</b>      | <b>4.20</b>    | <b>8.14</b>       | <b>6.04</b>    | <b>10.05</b>      |
| Fourth   | <b>7.94</b>       | <b>14.37</b>      | <b>9.12</b>    | <b>17.05</b>      | <b>9.28</b>    | <b>15.05</b>      |
| Last   | <b>7.92</b>       | <b>14.26</b>      | <b>9.25</b>    | <b>18.96</b>      | <b>9.80</b>    | <b>15.12</b>      |

NOTE: \*The null hypothesis is that the AR model is more accurate than the model with the labor market variable. Here  $P/R=0.36$ . For  $P/R=0.4$ , the 99th, 95th, and 90th percentiles for the OOS-F tests are 2.768, 1.298, and 0.814, respectively.

†The null hypothesis is that the AR model encompasses the model with the labor market variable. Here  $P/R=0.36$ . For  $P/R=0.4$ , the 99th, 95th, and 90th percentiles for the Clark-McCracken encompassing tests are 2.098, 1.079, and 0.685, respectively.

Bold values indicate that the null hypothesis is rejected at the 99th percentile.

Using a 1 percent critical region, the F tests reject the null hypothesis that the AR forecast of the unemployment rate is better than the initial claims forecast for the second and third weeks, but not for the fourth and last weeks. This hypothesis is rejected for the continuing claims data in the models where at least three weeks of data are available. We can reject the null hypothesis for payroll jobs as well, when we can use initial claims data for all but the model with the first two weeks of data. As we found with the unemployment rate, the null hypothesis is rejected in the cases using at least three weeks of continuing claims data. We cannot reject the null hypothesis in the case of hours worked for initial claims, but we can for cases including three or more

weeks of continuing claims data. The encompassing tests are reported in the bottom panel of Table 4. In all but a few cases involving models with just the first-week data, we can reject the null hypothesis that the AR model encompasses the models including the claims variables at the 99th percentile.

Our forecasting period included the recession that began in July 1990 and ended in March 1991, as well as the first nine months of the current recession. Both Montgomery et al. (1998) and McConnell (1998) conclude that initial claims data can forecast labor market variables, but only in times of recession and rising unemployment. Therefore we calculated the forecasting performance of these models during the 10 years of expansion from April 1991 through

February 2001. These results are reported in Table 5. Looking at expansion months only, we find much less information in the claims data. However, there is still some evidence that initial claims data help to predict the unemployment rate and continuing claims data help to predict growth in hours worked. Again, even though the AR model often had a lower RMSE, we could always reject the hypothesis that the AR model encompassed the model that included the claims data when we used at least three weeks of data.

## CONCLUSION

Empirical evidence and economic theory suggest that changes in labor market conditions will have significant effects on aggregate output. Evidence presented in this paper further suggests that incoming monthly data on nonagricultural payroll jobs and the index of aggregate weekly hours help predict changes in real GDP growth. Changes in the civilian unemployment rate are less significant. This finding suggests that predicting monthly changes in jobs or hours growth would be helpful in predicting real GDP growth. Many economists and financial market analysts strive to do this by tracking initial claims for state unemployment insurance benefits, which are released weekly. This article has shown that there is some statistically significant marginal information in the unemployment insurance claims data, even during periods of expansion. However, information about continuing claims appears to be at least as important as the information about initial claims that usually appears in the headlines.

## REFERENCES

- Board of Governors of the Federal Reserve System. *Monetary Policy Report to the Congress*. February 2002.
- Clark, Todd E. and McCracken, Michael W. "Tests of Equal Forecast Accuracy and Encompassing for Nested Models." Working Paper RWP 99-11, Federal Reserve Bank of Kansas City, November 2000.
- Getz, Patricia M. "Implementing the New Sample Design for the Current Employment Statistics Survey." *Business Economics*, October 2000, 35(4), pp. 47-50.
- Hatzius, Jan. "Jobless Claims Imply Contraction, but Only at a Glacial Pace." *Goldman Sachs Economics Goldman U.S. Daily*, 15 June 2001.
- Jordan, Jerry L. "What Monetary Policy Can and Cannot Do." Federal Reserve Bank of Cleveland *Economic Commentary*, 15 May 1992.
- Koenig, Evan F.; Dolmas, Sheila and Piger, Jeremy M. "The Use and Abuse of 'Real-Time' Data in Economic Forecasting." Working Paper 2001-015A, Federal Reserve Bank of Dallas, 2001.
- Mankiw, N. Gregory. *Principles of Economics*. Orlando, FL: Harcourt Brace, 1998.
- McConnell, Margaret M. "Rethinking the Value of Initial Claims as a Forecasting Tool." Federal Reserve Bank of New York *Current Issues in Economics and Finance*, November 1998, 4(11), pp. 1-6.
- McCracken, Michael W. "Asymptotics for Out of Sample Tests of Causality." Unpublished manuscript, Louisiana State University, November 1999.
- Miller, Preston J. and Chin, Daniel M. "Using Monthly Data to Improve Quarterly Model Forecasts." Federal Reserve Bank of Minneapolis *Quarterly Review*, Spring 1996, 20(2), pp. 16-33.
- Montgomery, Alan L.; Zarnowitz, Victor; Tsay, Ruey S. and Tiao, George C. "Forecasting the U.S. Unemployment Rate." *Journal of the American Statistical Association*, June 1998, 93(442), pp. 478-93.

## Appendix

### METHODOLOGY OF THE UNEMPLOYMENT INSURANCE CLAIMS DATA

#### *Data Series and Sources*

Each week, state government employment offices report the number of individuals filing claims for unemployment insurance benefits. The state offices then report the figures to the Office of Workforce Security in the ETA.<sup>11</sup> They are then published in the *Unemployment Insurance Weekly Claims Report*, which is issued by the ETA. Also published in this report are continuing claims for state unemployment insurance benefits (insured unemployment), which is another closely monitored indicator.

#### *Eligibility Requirements*

Individuals who file for unemployment insurance benefits are not automatically eligible for benefits. To qualify for benefits, the worker must first demonstrate that they have a work history, otherwise known as an “attachment to the labor force.” In most states, this requirement is met by having earned a minimum amount of money in a job that is covered by the law. In some states, a person is eligible if they have merely worked a minimum amount of time in covered employment. Covered employment excludes self employment, small farms, and small domestic operations. Once the person is deemed monetarily eligible, the reason for the claim is then examined. Although a common reason stems from an unintended loss of employment, some states disburse benefits to individuals who are following a spouse to a new job. If an unfavorable ruling results, the claimant may appeal the decision.

#### *Waiting Period Requirements*

In general, individuals do not receive benefit checks for two to three weeks after they are classi-

fied as eligible. Moreover, there is an additional lag in those states that have a one-week waiting period. This means that they cannot claim benefits for that week. Most states require that claimants file for benefits every two weeks. For every week a person claims benefits, they are required to be available and actively seeking work, and, among other things, they cannot refuse a suitable job.

#### *Type of Claims*

The initial claims series that is reported weekly comprises two types of claims: *new* and *additional*. A new claim is defined as the first initial claim filed in person, by mail, telephone, or other means to request a determination of entitlement to and eligibility for compensation. This results in an agency-generated document to determine monetary eligibility. An additional claim is a subsequent initial claim filed (i) during an existing benefit year due to new unemployment and (ii) when a break of one week or more has occurred in the claim series due to intervening employment. Thus, these claims are reported only when there has been intervening employment since the last claim was filed. Claims that follow breaks due to illness, disqualification, unavailability, or failure to report for any reason other than job attachment are not reported. Thus, if a person has multiple occurrences of unemployment during their benefit year, the first one is counted as a new initial and the others are counted as additional initials. Both numbers are incorporated into the published weekly counts and thus represent new emerging unemployment for that week.

---

<sup>11</sup> The claims data are not derived from the ES-202 program, which is the source of employment and wage data by industry at the national, state, and county levels. Thus, they are not drawn from the sample of data that is used to construct the establishment data in the monthly *Employment Situation* report, nor are they used to calculate the unemployment rate, which comes from the household survey.





# Did “Right-to-Work” Work for Idaho?

Emin M. Dinlersoz and Rubén Hernández-Murillo

## RIGHT-TO-WORK LAWS AND ECONOMIC ACTIVITY

The right-to-work (RTW) law ensures that workers are not forced to join unions or pay union dues as a condition of employment.<sup>1</sup> Despite many years of research, the impact of these laws on a state’s economic performance is still a controversial issue. Using a diverse set of data and methods, a sizeable body of literature has concentrated on understanding whether the passage of RTW laws matters.<sup>2</sup> RTW laws continue to be an important issue on states’ agendas and a source of fierce campaigning by pro- and anti-union groups. For instance, in September 2001, Oklahoma adopted the RTW law after a lengthy period of campaigns for and against it.

States with RTW laws usually offer additional policies as part of a pro-business profile designed to attract new firms and boost industrial development. This is the view taken by Holmes (1998), who uses the RTW law as a proxy for the state’s business-friendly climate. He studies the effects of pro-business policies on economic activity by examining the performance of manufacturing industries across state borders where one state has a RTW law and the other does not. His analysis identifies a large, positive impact of an overall favorable business climate, but the effects cannot be traced to any particular state legislation, such as a RTW law.

Many states passed RTW laws in the mid-1940s to early 1950s. Since then, except for the 2001 adoption by Oklahoma, only two other states adopted them: Louisiana in 1976 and Idaho in 1986. Indiana adopted the law in 1957, but repealed it in 1965. It is natural to think that economic conditions today

are quite different from those that prevailed during the earlier period when many states passed the law en masse. An important question then is whether the late adopters of this law have experienced any real benefits.

## Idaho’s Case

In this paper, we reassess the economic impact of the RTW law by focusing on Idaho’s experience.<sup>3</sup> Idaho adopted their RTW law in 1986, at a time when the decline in unionization in the U.S. had substantially run its course.<sup>4</sup> Was the passage of the law merely a gesture that simply reflected a trend of decline in unionization, or did it have a significant influence in making Idaho a more attractive location for business in the years following the adoption? Our goal is to provide some evidence on how Idaho’s unionization rate and industrial performance evolved over time, both before and after the passage of the RTW law, thereby contributing to the literature on the effect of business-friendly policies on states’ industrial performance.

One important aspect of Idaho’s experience is that the passage of the law itself was a long and controversial process that took nearly two years. The critical events related to the legislation process are summarized in Abraham and Voos (2000). The original bill was introduced to Idaho’s House in January 1985, and the law was eventually passed in November 1986, after a lengthy political and bureaucratic process involving several confrontations between pro-law and anti-law groups, as well as a veto and several delays. The law finally took effect in 1987.

A detailed investigation of other business policies adopted in Idaho around 1987 reveals that there were no other major changes in Idaho’s business

---

Emin M. Dinlersoz is an economist at the University of Houston. Rubén Hernández-Murillo is an economist at ITAM, Mexico. The authors thank Gordon Dahl, Roger Sherman, Lori Taylor, and seminar participants at the October 2001 Federal Reserve System Conference on Regional Analysis in San Antonio, Texas, for comments and suggestions. Barry Hirsch and David Macpherson provided useful suggestions to compute the estimates of unionization rates. This article was written when the authors were conducting research at the Federal Reserve Bank of St. Louis.

© 2002, The Federal Reserve Bank of St. Louis.

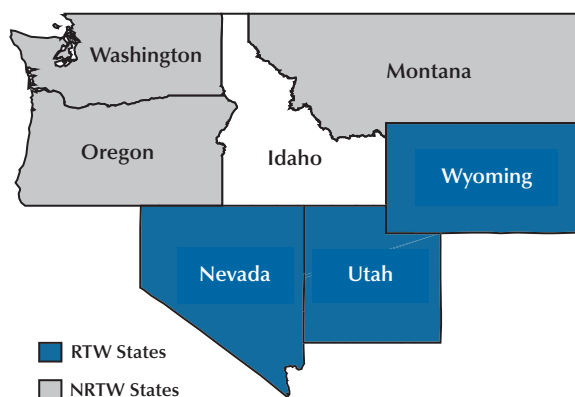
---

<sup>1</sup> Section 14(b) of the Taft-Hartley Act, passed in 1947 by Congress, reaffirms states’ rights to pass RTW laws. These laws may or may not apply to federal workers, depending on the specifics.

<sup>2</sup> See Moore and Newman (1985) and Moore (1998) for a comprehensive review of this literature.

<sup>3</sup> Louisiana is also a candidate for such a study. However, the unavailability of long time series data before Louisiana’s adoption year (1976) prevents the investigation of this case in detail.

<sup>4</sup> Goldfield (1987) reports that between 1954 and 1978 the union membership rate in the United States declined from 34.7 percent to 23.6 percent. See Goldfield (1987) for a comprehensive analysis of the declining unionization in the United States. According to Hirsch, Macpherson, and Vroman (2001), the union membership rate declined from 29.3 percent in 1964 to 24.1 percent in 1977, and then to 13.6 percent in 2000.

**Figure 1**

climate regarding incentives for new investments or firm relocation.<sup>5</sup>

Idaho offers an interesting case study not only because it is a late adopter, but also because three of its six neighboring states have had the RTW law for a long time and three have traditionally been non-right-to-work (NRTW) law states.<sup>6</sup> Figure 1 shows Idaho and its neighbors, which provide potential controls against which to judge Idaho's performance. Clearly, these states are imperfect controls. However, among all other states, Idaho's neighbors seem to be a natural choice for comparison for the reason, if nothing else, that we can control for common region-specific factors that do not vary over time. Responses to nationwide economic fluctuations vary substantially across regions. Focusing on a particular region minimizes this problem. In analyzing the evolution of unionization rates, we also consider the experience of states with an industry mix that was similar to that of Idaho to account for differences arising from the composition of industrial activity.

Our empirical analysis has two main parts. First, we look at the evolution of the unionization rate before and after the law. We find that there was a large decline in unionization between 1981 and 1984, the year before the bill was introduced to the legislature. The unionization rate then rebounded somewhat until 1987, the year the law officially took effect, but continued to decline persistently thereafter. Idaho's unionization rate gradually became very similar to the average unionization rate of other RTW states with a similar industrial mix. When we compare Idaho's unionization rate also to that of its geographic neighbors, we find that, particularly in the manufacturing sector, Idaho's unionization rate exhibits a significantly faster decline.

Second, we investigate the manufacturing sector's performance pre- and post-law. We observe that in the post-law period, Idaho experienced a significant and persistent annual growth in manufacturing employment and in the number of establishments, as opposed to virtually zero growth in both of these variables in the pre-law period. The difference between the pre-law and post-law growth rates in Idaho was significantly larger compared with other states in the region. In addition, we find that the fraction of total manufacturing employment in large manufacturing establishments increased significantly in Idaho after the law was passed. The average size of large manufacturing establishments also grew substantially in the post-law period.<sup>7</sup> Our observations are consistent with the hypothesis that Idaho became more attractive for large plants because of declining unionization.

Overall, our findings indicate that the increase in Idaho's industrial growth rate is strongly related to the decline in unionization. While we are tempted to associate the patterns observed with the passage of the law itself, the timing of the decline in the unionization rate prevents such a definitive conclusion. The large decline in unionization started about four years prior to the almost two-year-long bureaucratic process that eventually led to the passage of the law. This prompts us to consider the hypothesis that the passage of the law might actually have been a consequence of the decline in unionization and growing anti-unionism in Idaho, rather than a cause. Consequently, while the declining unionization appears to be responsible for the strong post-law growth trends in Idaho, we cannot fully ascribe the initiation of the trends to the law itself. The *passage* of the law, however, seems to have strengthened and reinforced the trends.

## Literature Review

One expects that a first-order effect of the passage of a RTW law would be a reduction in the

<sup>5</sup> We examined, in particular, the *Directory of Incentives for Business Investment and Development in the United States*, published by the National Association of State Development Agencies.

<sup>6</sup> The RTW neighbors, Nevada, Utah, and Wyoming, adopted the law in 1951, 1955, and 1963, respectively. The time period between these years and our first observation year (1975) is long enough to give us some comfort that the potential effects of the RTW law must have already been realized to a large extent in these states.

<sup>7</sup> In general, larger establishments are more likely to be unionized and, therefore, have more incentives to avoid unions. See Long (1993), Galarneau (1996), and Lowe (1998) for evidence on this in Canada.

union membership rate. There are several reasons why this might be the case. As Ellwood and Fine (1987) point out, the most obvious reason is that the passage of the law makes unions less attractive to workers because unions no longer have the ability to enforce payments and fines. These effects depress new union organizing and also deter the replacement of decertified unions. If a state's labor force is growing, then less union organizing means also a reduction in the union membership rate.

Most earlier studies, surveyed by Moore and Newman (1985) and Moore (1998), found a weak relationship between the passage of RTW laws and the level of the union membership rate. However, this does not mean that unionization activity was not influenced by RTW laws. Using 1951-77 data for 50 states on new union organizing activity (a measure of new membership flow into unions, rather than the level of unionization), Ellwood and Fine (1987) presented convincing evidence that the passage of RTW laws led to a decline in new union organizing of about 46 percent for the first five years after the legislation and 30 percent during the next five. This reduction in organizing disappears after a decade. The level of union membership, as a result, declines in most states by about 5 to 10 percent after the 10 years, which may not have been detected by the econometric methods used in the previous studies. Further tests reveal that these findings are robust to time-invariant differences across states. Idaho's experience provides a natural setting to further assess the evolution of the union membership rates before and after the passage of the law. Since we are looking at the same state both before and after, time-invariant state-specific factors should be irrelevant for the pattern of evolution in the union membership rates.

As we mentioned before, an important concern is whether declining union strength is a catalyst for the passage of RTW laws, as opposed to being a result of it. If the passing of RTW laws is a consequence rather than the cause, then the reduction in union organizing should be visible during the immediate years before the passage of the law. Ellwood and Fine (1987) investigated this possibility by analyzing the evolution of new union organizing for seven states prior to the adoption of the law; they detected no reduction in union organizing during that period and concluded that the decline in union organizing is likely to have been caused by the passage of RTW laws.

According to the anecdotal evidence in Kendrick

(1996), one possible source of the events that led to the eventual passage of the law was the "Bunker Hill" incident. In 1984, employees of the Bunker Hill mining company in Idaho voted for voluntary pay cuts and other concessions to keep the company from going out of business. The union headquarters in Pittsburgh overruled this vote, resulting in a loss of 1500 jobs. The Bunker Hill incident might have initiated a change in attitude toward unions in Idaho. If this is the case, then a growing anti-unionism in the state might be the reason for the eventual passage of the law.

The rest of our paper is organized as follows. We present evidence in the next section on the evolution of unionization before and after the RTW law was enacted, followed by evidence on the growth in manufacturing.

## PATTERNS OF UNIONIZATION IN IDAHO

### *Unionization Across Industries*

We used data from the Census Bureau's Current Population Survey (CPS) to estimate unionization rates. We describe the characteristics of the data and methodology in the appendix. The employment and establishments data for our analysis of manufacturing comes from the Census Bureau's County Business Patterns data set and is also described in the appendix.

We start our analysis by examining the evolution of the unionization rate in Idaho. We compare the trends in Idaho's unionization rate with the average trend in both RTW states and NRTW states that had an industrial mix similar to that of Idaho in the years prior to the passage of the law, 1977-86. For this we construct a measure of dispersion using the employment shares in broadly defined industries.<sup>8</sup> We identi-

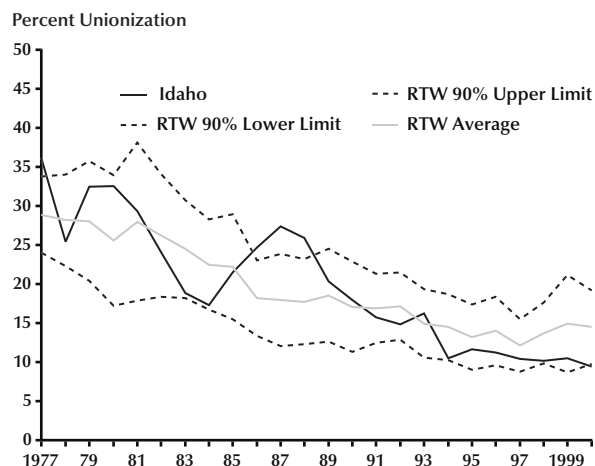
<sup>8</sup> We computed the following measure of *distance* ( $\bar{\Delta}_k$ ) to Idaho for each of the 50 states in terms of industrial mix and performed the comparison for the closest "neighbors":

$$\bar{\Delta}_k = \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^N (s_k^{it} - s^{it})^2,$$

where  $s_k^{it}$  is the employment share in industry  $i$  in state  $k$  in year  $t$ ,  $N$  is the number of industries,  $T$  is the total number of years in the sample period, and  $s^{it}$  is the index for Idaho, defined similarly. We used employment data from the following industry classifications: agricultural, mining, construction, manufacturing, transportation, wholesale trade, retail trade, finance insurance and real estate services, and personal services. The distribution of this measure had the following characteristics: the maximum value was 0.143, the mean was 0.019, and the 5th, 25th, 50th, 75th, and 90th percentiles were 0.002, 0.005, 0.014, 0.024, and 0.035. We selected states with a distance of less than 0.005.

**Figure 2****Evolution of Unionization in Manufacturing Industries**

Idaho vs. RTW States



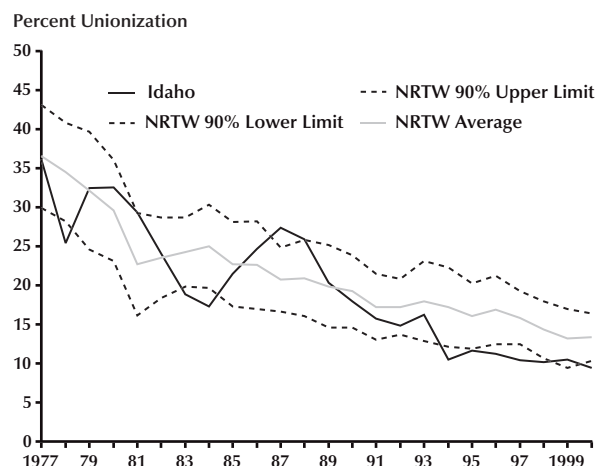
fied 11 such states: 5 RTW states (Kansas, Nebraska, Utah, Virginia, and Iowa) and 6 NRTW states (California, Colorado, Minnesota, Oklahoma, Oregon, and Washington).

So how do the patterns in unionization rates differ across industries? The manufacturing sector, being traditionally highly unionized, behaved quite differently compared with the nonmanufacturing sector. Figures 2 and 3, respectively, compare the unionization rate in manufacturing to the average of RTW and NRTW states. What is most interesting about the trend for Idaho's unionization is the relatively large decline that occurred between 1981 and 1984, prior to the passage of the law, and the pronounced recovery in the 1984-87 period, during which much of the debate about the passage of the law took place. We observe that the manufacturing unionization rate in Idaho gradually converged to the average unionization rate in RTW states. The convergence took place mostly after 1987, and this rate remained within the confidence bands and below the average for RTW states that had similar industrial composition prior to 1987. Figure 3 indicates that the manufacturing unionization rate in Idaho remained within the confidence bands for the average for NRTW states for most of the sample period, but fell below the lower confidence band in 1994 and remained away from the average thereafter.

The patterns observed in Idaho's manufacturing unionization rate do not seem to result from business

**Figure 3****Evolution of Unionization in Manufacturing Industries**

Idaho vs. NRTW States



cycles that affected all other states uniformly. However, since Idaho is a small state, its manufacturing unionization rate may have been subject to fluctuations in the unionization rate of a small number of industries, particularly in the period prior to the passage of the RTW law. Examining Idaho's unionization rates in narrowly defined manufacturing industries, we discovered that fluctuations in the years prior to 1987 were closely related to fluctuations in the food manufacturing industry.

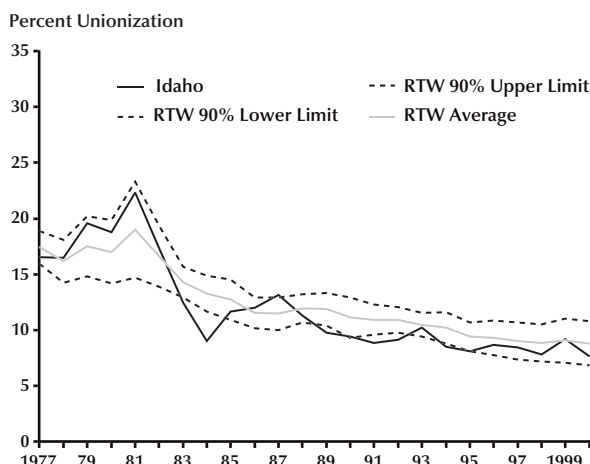
Figure 4 shows the evolution of the overall unionization rate in Idaho versus the average unionization rate in the five states with RTW laws and a similar industrial mix.<sup>9</sup> Idaho's unionization rate was around 17 percent in 1977; by 2000 it was down to about 9 percent, a decline of almost 50 percent. The average rate for RTW states also declined steadily, starting in 1981. Throughout the period of analysis, in 1983, 1984, and then again in 1987, 1989, 1991, 1992, and 1994, Idaho's unionization rate was significantly different from the average RTW state's unionization rate, at the 90 percent confidence level. In the years 1977-81 we observe that Idaho's unionization rate was close to the upper confidence band. In just three years, during the period 1981-84, the

<sup>9</sup> Note that there was no change in other states' RTW law status during 1977-2000. Idaho was the only state that changed status during this period. Louisiana became a RTW state in 1976 and is included with the other RTW states throughout the period. Excluding Louisiana did not change our conclusions.



**Figure 4****Evolution of Unionization in All Industries**

Idaho vs. RTW States

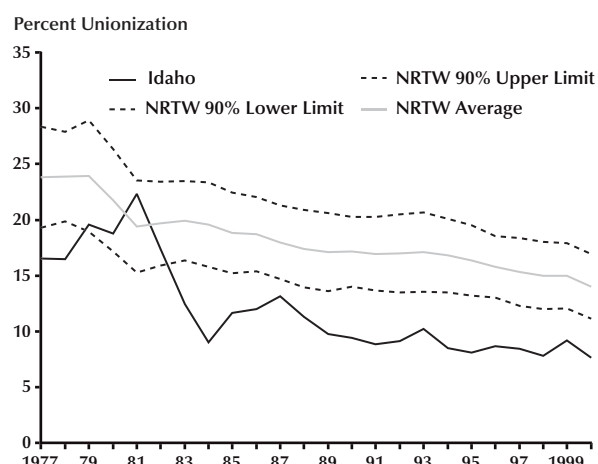


unionization rate fell from about 22 percent to almost 9 percent, a decline of about 60 percent. The decline observed for the average rate for RTW states was not as pronounced.<sup>10</sup> The pattern between 1984 and 1987 also exhibits a partial recovery in the unionization rate. After the law took effect in 1987, however, we observe a persistent decline in the unionization rate.

In Figure 5, we compare Idaho with the six closest NRTW states. First, note that on average, a NRTW state had a unionization rate of about 24 percent in 1977, compared with 17 percent for RTW states. These figures were about 14 percent and 9 percent, respectively, in 2000. The difference in unionization rates between the two groups of states persisted throughout the sample period. In the years 1979-82, Idaho's unionization rate is not statistically distinguishable from the average unionization rate in NRTW states. In the years following the 1981-84 decline, however, we can reject the equality of the two rates. Idaho's unionization rate hit the lower confidence bound for the NRTW states' average around 1982 and consistently remained below that bound for the rest of the analysis period. From the patterns observed in Figures 4 and 5, Idaho's unionization rate very early diverged from the NRTW states' average unionization rate and approached the RTW states' average. As shown in Figures 6 and 7, this behavior was largely due to the behavior observed in the nonmanufacturing sector. In both

**Figure 5****Evolution of Unionization in All Industries**

Idaho vs. NRTW States



figures, the dip during the 1981-84 period is visible and highly pronounced, and even as early as 1982 the unionization rate in nonmanufacturing industries had converged to the average unionization rate in RTW states and was statistically below the NRTW states' average. It is therefore likely that the quick convergence in Idaho's overall unionization rate was unrelated to the passage of the RTW.<sup>11</sup>

**Idaho's Neighbors**

To investigate the trends in the unionization rate further, we concentrate on Idaho's geographic neighbors and run a simple state-by-state regression of the form

<sup>10</sup> As explained in the appendix, prior to 1983, unionization rates were calculated based on samples that are roughly one-third of the samples that are used after 1983. The estimated unionization rates are less precise for the period before 1983 due to sample variability, especially for smaller states, and in particular for 1981, when the sample sizes were roughly one-third of the samples in 1977-80. Estimates of overall and nonmanufacturing unionization rates were less sensitive to sampling problems than those for the manufacturing sector. Still, when we discount 1981 and 1982, the decline observed in the manufacturing unionization rate from 1980 to 1983 is reliably estimated.

<sup>11</sup> As previously footnoted, the estimates of overall and nonmanufacturing unionization rates during the period 1977-86 were not likely to be seriously affected by the small sample sizes used by the CPS before 1983, even accounting for 1981, as the sample sizes used in the estimation exceeded the thresholds described in the appendix for reliability of the estimates. We are, however, silent on the driving factors of unionization in Idaho's nonmanufacturing industries, as the focus of our analysis is the manufacturing sector. We did verify, however, that the 1981-84 decline was not due to closures of large unionized firms.

Figure 6

### Evolution of Unionization in Nonmanufacturing Industries

Idaho vs. RTW States

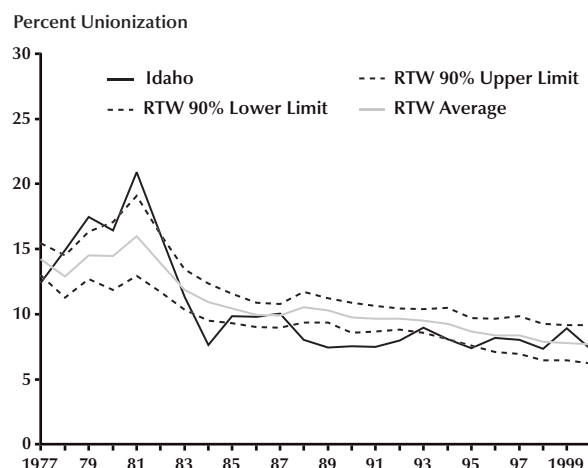


Figure 7

### Evolution of Unionization in Nonmanufacturing Industries

Idaho vs. NRTW States

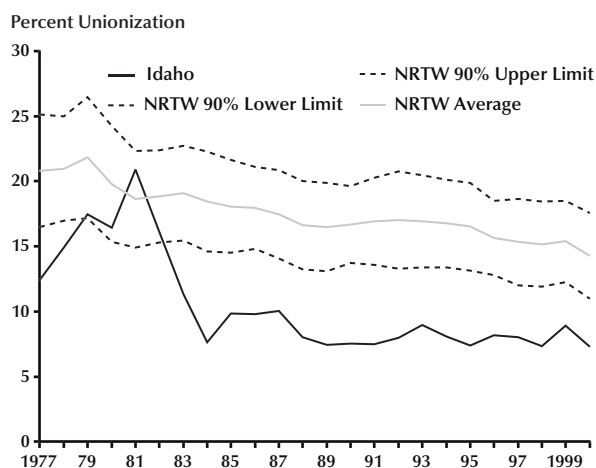


Table 1

### Change in Unionization Rate by State and Industry

|               | Overall               |                       |                    | Manufacturing        |                       |                    | Nonmanufacturing     |                       |                    |
|---------------|-----------------------|-----------------------|--------------------|----------------------|-----------------------|--------------------|----------------------|-----------------------|--------------------|
|               | 1977-86               | 1987-2000             | F (Prob)           | 1977-86              | 1987-2000             | F (Prob)           | 1977-85              | 1987-2000             | F (Prob)           |
| U.S.          | <b>-3.7</b><br>[-0.3] | <b>-1.8</b><br>[0.06] | 27.37***<br>(0.00) | <b>-4.7</b><br>[0.3] | <b>-3.3</b><br>[0.06] | 15.54***<br>(0.00) | <b>-2.8</b><br>[0.3] | <b>-1.2</b><br>[0.08] | 17.97***<br>(0.00) |
| Idaho         | <b>-6.4</b><br>[1.8]  | <b>-2.8</b><br>[0.7]  | 3.2*<br>(0.08)     | <b>-5.8</b><br>[2.0] | <b>-8.0</b><br>[0.8]  | 1.06<br>(0.31)     | -6.3<br>[2.5]        | -0.5<br>[0.7]         | 4.75**<br>(0.04)   |
| Washington    | <b>-3.0</b><br>[0.7]  | <b>-1.7</b><br>[0.3]  | 2.98*<br>(0.09)    | <b>-4.1</b><br>[0.9] | <b>-2.5</b><br>[0.3]  | 2.73<br>(0.11)     | <b>-2.4</b><br>[0.7] | <b>-1.2</b><br>[0.3]  | 2.25<br>(0.14)     |
| Oregon        | <b>-3.6</b><br>[1.0]  | <b>-2.1</b><br>[0.4]  | 1.71<br>(0.20)     | <b>-7.6</b><br>[0.7] | <b>-5.7</b><br>[0.7]  | 3.14*<br>(0.09)    | -1.6<br>[1.1]        | <b>-1.3</b><br>[0.4]  | 0.06<br>(0.81)     |
| Montana       | <b>-4.2</b><br>[0.7]  | <b>-2.1</b><br>[0.4]  | 6.48***<br>(0.01)  | -3.0<br>[2.7]        | <b>-5.3</b><br>[0.9]  | 0.65<br>(0.43)     | <b>-4.4</b><br>[0.6] | <b>-1.8</b><br>[0.3]  | 11.36***<br>(0.00) |
| Nevada (RTW)  | <b>-3.1</b><br>[0.9]  | 0.2<br>[0.4]          | 10.34***<br>(0.00) | -9.2<br>[3.9]        | -0.5<br>[2.5]         | 3.46*<br>(0.07)    | <b>-2.9</b><br>[0.8] | 0.2<br>[0.4]          | 10.22***<br>(0.00) |
| Utah (RTW)    | <b>-5.5</b><br>[1.4]  | <b>-3.7</b><br>[0.6]  | 1.26<br>(0.27)     | <b>-9.6</b><br>[1.7] | -2.6<br>[1.2]         | 11.49***<br>(0.00) | <b>-4.7</b><br>[1.4] | <b>-4.0</b><br>[0.7]  | 0.17<br>(0.68)     |
| Wyoming (RTW) | -3.5<br>[1.5]         | <b>-3.8</b><br>[0.2]  | 0.02<br>(0.88)     | 0.2<br>[2.1]         | 2.8<br>[3.3]          | 0.45<br>(0.51)     | -3.7<br>[1.6]        | <b>-4.0</b><br>[0.2]  | 0.04<br>(0.83)     |

NOTE: Heteroskedasticity-autocorrelation consistent standard errors are in brackets. Figures in bold indicate significance at 1 percent. "F" gives the F statistic for the test of equality of coefficients across two time periods. Probability values for the F statistic are in parentheses. \*, \*\*, and \*\*\* indicate significance of the F statistic at the 10, 5, and 1 percent levels, respectively.

$$(1) \log u_t = \alpha_{PRE} + \beta_{PRE}D(t-t_0) + \Delta\alpha_{POST}(1-D) + \beta_{POST}(1-D)(t-t_0) + \varepsilon_t,$$

where  $t_0 = 1977$ ,  $t = 1977, \dots, 2000$ , and  $D$  is a dummy variable that takes on a value of 1 if  $t < 1987$  and 0 otherwise. In this projection,  $\alpha_{PRE}$  is the intercept term for the pre-law period,  $\beta_{PRE}$  is the pre-law slope coefficient,  $\Delta\alpha_{POST}$  is the post-law increment in the intercept, and  $\beta_{POST}$  is the post-law slope coefficient. The estimated values of  $\beta_{PRE}$  and  $\beta_{POST}$  are multiplied by 100 and are presented in Table 1. With the log specification, the figures in the table can be interpreted as the annual percent rate of change in unionization. We also present the test results for the equality of the growth rates across the two periods  $\beta_{PRE} = \beta_{POST}$ .

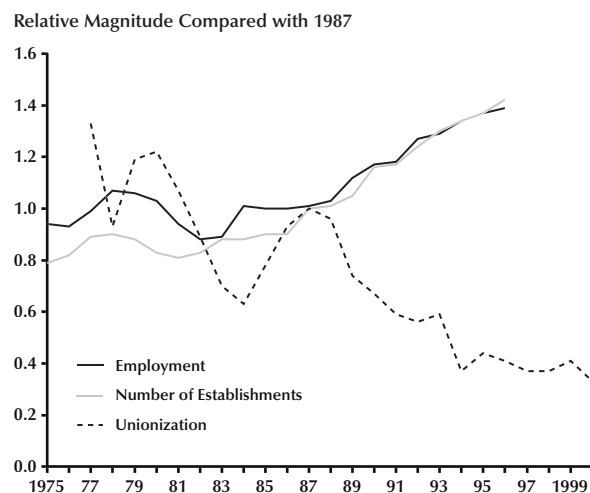
We observe a persistent decline in unionization rates. When all industries are considered, columns 1 and 2 reveal that, in general, the magnitude of the decline was higher in the 1977-86 period in all states in the region and in the United States, except for Wyoming. In Idaho, the rate of decline in overall unionization slowed down from 6.4 percent in the pre-law period to 2.8 percent in the post-law period. The difference between these two rates, however, is statistically significant only at the 10 percent level. Note also that, in both periods, Idaho's rates of decline were higher than the U.S. rates and most of those for its neighboring states.

When manufacturing is considered separately, columns 4 and 5 provide a different view. In fact, the decline in Idaho's manufacturing unionization rate accelerated somewhat in the post-law period, surpassing both the U.S. and its neighboring states, which, for the most part, exhibited a slowdown in the rate of decline. The difference between Idaho's unionization rates in manufacturing pre-law and post-law is not statistically significant because of the relatively high standard deviation for the pre-law period. Overall, the slowdown in the rate of decline of unionization did not apply to Idaho's manufacturing and was primarily driven by nonmanufacturing industries, as can be seen in the last two columns.

The findings in this section suggest that Idaho's unionization rate declined substantially over the sample period, approaching the average unionization rate in RTW states. While the decline in the unionization rate, especially in manufacturing, is persistent after 1987, a substantial part of the decline appears to have happened before 1987. The pattern between 1984 and 1987, during which much of the debate about the passage of the law took place, exhibits a partial recovery in the unionization rate. After the law took effect in 1987, we observe a con-

**Figure 8**

### Evolution of Key Indicators in Idaho's Manufacturing Industry



tinuing decline in the unionization rate, especially in manufacturing. Particularly during the period prior to 1987, large fluctuations in Idaho's unionization rate in manufacturing seem to be related to the behavior of individual industries.

## MANUFACTURING

We now turn to the industrial organization consequences of declining unionization in Idaho. We focus on two main indicators. First, we look at the growth in employment and the number of establishments in manufacturing industries and compare Idaho with its neighbors, in both the pre- and post-law periods. If the passage of the law has had an important positive effect on manufacturing growth, then we expect to observe an acceleration in the growth rate of employment and the number of establishments in Idaho. Second, we look at the changes in the importance of large establishments in manufacturing in Idaho, again, for both periods. As Holmes (1998) argues, large manufacturing establishments are more likely to be attracted to RTW states because larger plants are more likely to be unionized. This argument suggests that we might expect an influx of new large establishments into Idaho or an expansion of existing establishments.

### Employment Growth

Figure 8 is a preliminary look at the evolution of the three key variables in Idaho's manufacturing,

Table 2

**Manufacturing Growth Rates in Idaho and Its Neighbors (Simple Time Averages, Percent Annual Growth)**

|               | Employment      |                | No. of establishments |                | Average establishment size |                 |
|---------------|-----------------|----------------|-----------------------|----------------|----------------------------|-----------------|
|               | 1975-86         | 1987-96        | 1975-86               | 1987-96        | 1975-86                    | 1987-96         |
| Idaho         | 0.76<br>[6.38]  | 3.71<br>[2.56] | 1.27<br>[4.20]        | 3.99<br>[3.16] | -0.39<br>[6.73]            | -0.21<br>[3.17] |
|               | (1.36*)         |                | (1.98**)              |                | (0.08)                     |                 |
| Washington    | 1.57<br>[5.34]  | 2.18<br>[6.23] | 2.86<br>[4.16]        | 1.96<br>[2.44] | -1.04<br>[7.48]            | 0.29<br>[6.94]  |
|               | (0.25)          |                | (-0.59)               |                | (0.43)                     |                 |
| Oregon        | 1.18<br>[6.86]  | 1.67<br>[2.64] | 2.32<br>[4.23]        | 1.19<br>[2.52] | -0.98<br>[7.54]            | 0.55<br>[4.30]  |
|               | (0.21)          |                | (-0.74)               |                | (0.57)                     |                 |
| Montana       | -0.33<br>[6.80] | 1.35<br>[3.47] | 1.94<br>[5.43]        | 2.09<br>[4.34] | -2.10<br>[6.79]            | -0.51<br>[6.33] |
|               | (0.71)          |                | (0.07)                |                | (0.56)                     |                 |
| Nevada (RTW)  | 6.15<br>[9.22]  | 4.84<br>[5.02] | 5.46<br>[6.52]        | 6.01<br>[3.43] | 0.96<br>[10.42]            | -1.00<br>[5.64] |
|               | (-0.40)         |                | (0.24)                |                | (-0.53)                    |                 |
| Utah (RTW)    | 3.51<br>[4.52]  | 3.26<br>[2.64] | 3.01<br>[2.93]        | 3.57<br>[3.55] | 0.58<br>[5.69]             | -0.17<br>[4.96] |
|               | (-0.15)         |                | (0.41)                |                | (-0.32)                    |                 |
| Wyoming (RTW) | -1.42<br>[9.66] | 3.32<br>[3.39] | 2.19<br>[6.50]        | 2.69<br>[4.16] | -0.53<br>[9.63]            | 0.87<br>[7.26]  |
|               | (0.59)          |                | (0.21)                |                | (0.38)                     |                 |

NOTE: Standard deviations in brackets. Figures in parentheses are the t statistics associated with the difference of the variable's average across two periods of analysis. \* and \*\* indicate significance at the 10 and 5 percent levels, respectively, for a one-sided test; t statistics are based on unpaired comparisons with unequal variances.

where we have normalized each variable by its 1987 value. Before 1987, there is considerable fluctuation in both employment and the number of establishments, with no visible growth trend. Unionization exhibits a decline, but is also subject to wide fluctuations, as discussed before. The pattern after 1987 is remarkably stable for all three series. Employment and the number of establishments grew steadily in that period by about 40 percent compared with their 1987 level, and unionization declined by more than 60 percent.

Table 2 shows the simple average annual growth rates in employment, number of establishments, and average establishment size in manufacturing for Idaho and its neighbors. Consider employment and the number of establishments first. From 1975 to 1986, Idaho's manufacturing employment grew at a rate of 0.76 percent annually on average. The

average growth rate in the number of establishments was around 1.27 percent per year. However, there is a large standard deviation associated with both of these figures, a reflection of the fluctuating manufacturing growth in the state in that period, as depicted in Figure 8. Idaho's NRTW neighbors did not fare much better. Washington and Oregon appear to have experienced higher growth rates, but the standard deviations are so high that the differences with respect to Idaho are not statistically significant. Idaho's RTW neighbors appear to have fared much better in this period, except for Wyoming. Overall, it seems that the period before the law was a period of weak growth, especially for NRTW states.

This pattern changes dramatically in the post-law period. Idaho's growth rates were much higher compared with those in the pre-law period. Furthermore, the difference between the two periods' growth



**Table 3****Manufacturing Growth Rates: Results from State-by-State Regressions**

|               | Employment          |                     |                        | No. of establishments |                     |                        |
|---------------|---------------------|---------------------|------------------------|-----------------------|---------------------|------------------------|
|               | 1977-86             | 1987-2000           | F (Prob)               | 1977-85               | 1987-2000           | F (Prob)               |
| Idaho         | -0.03<br>[0.4]      | <b>3.7</b><br>[0.2] | <b>58.97</b><br>(0.00) | 0.6<br>[0.3]          | <b>4.1</b><br>[0.2] | <b>82.71</b><br>(0.00) |
| Washington    | 1.1<br>[0.5]        | 0.4<br>[0.8]        | 0.48<br>(0.49)         | <b>2.4</b><br>[0.2]   | <b>2.1</b><br>[0.2] | 0.93<br>(0.34)         |
| Oregon        | 0.02<br>[0.6]       | <b>1.2</b><br>[0.2] | 3.17<br>(0.09)         | <b>2.0</b><br>[0.2]   | <b>1.6</b><br>[0.2] | 2.04<br>(0.17)         |
| Montana       | -1.3<br>[0.6]       | <b>1.2</b><br>[0.1] | <b>15.74</b><br>(0.00) | <b>1.8</b><br>[0.6]   | <b>2.6</b><br>[0.2] | 2.58<br>(0.12)         |
| Nevada (RTW)  | <b>5.0</b><br>[0.7] | <b>4.3</b><br>[0.6] | 0.49<br>(0.49)         | <b>4.9</b><br>[0.5]   | <b>6.1</b><br>[0.2] | 4.39<br>(0.05)         |
| Utah (RTW)    | <b>3.3</b><br>[0.3] | <b>3.1</b><br>[0.1] | 0.28<br>(0.60)         | <b>2.9</b><br>[0.2]   | <b>4.0</b><br>[0.3] | <b>8.85</b><br>(0.00)  |
| Wyoming (RTW) | -0.03<br>[1.3]      | <b>2.7</b><br>[0.2] | 3.99<br>(0.06)         | <b>2.1</b><br>[0.4]   | <b>3.0</b><br>[0.3] | 2.07<br>(0.16)         |

NOTE: Heteroskedasticity-autocorrelation consistent standard errors are in brackets. Figures in bold indicate significance at 1 percent. "F" gives the F statistic for the test of equality of coefficients across two time periods. Probability values for the F statistic are in parentheses.

rates turns out to be statistically significant, unlike the case with neighboring states. Idaho's post-law growth rates also exceeded those of its NRTW neighbors and were similar to those of its RTW neighbors (although the pairwise comparisons are not always statistically significant due to large standard errors). Overall, the patterns of change in the growth of employment and the number of establishments point to a post-law acceleration of growth in Idaho, but not in any of the neighboring states.

Table 3 shows the results of a regression analogous to equation (1). The dependent variable is either the logarithm of employment or the number of establishments in manufacturing. The most notable result from this table is the exceptionally large growth rate of Idaho in the post-law period for both variables. The annual employment growth rate was about 3.7 percent post-law, compared with an almost zero annual average growth pre-law. The growth rate in the number of establishments was about seven times larger compared with that in the pre-law period. Idaho did much better after the RTW law was passed, compared with most other states in the region, both in employment and the number of establishments. The differences in these growth rates across the two periods have high statistical signifi-

cance for Idaho, but not for most of the other states.

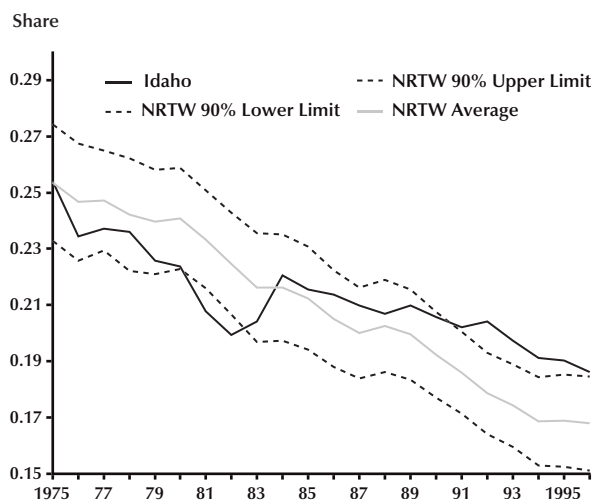
### **Manufacturing Employment Share**

Before turning to the analysis of establishment size, we report how the share of manufacturing as a fraction of total private employment evolved in Idaho. Again we compare Idaho against other states that had a similar industrial mix in the period prior to 1987. This analysis indicates that Idaho experienced a substantial change in industrial mix, especially after the passage of the RTW law.<sup>12</sup> Figure 9 compares Idaho's manufacturing share with the average manufacturing share in the six NRTW states we identified earlier. First, note that manufacturing's average employment share in NRTW states declined throughout the sample period, which is an indication of the steady decline in the manufacturing sector in the United States, especially during the last quarter of the twentieth century. Idaho's manufacturing share was far below the NRTW average

<sup>12</sup> Constructing a distance measure analogous to that of footnote 9, we observed that Idaho also experienced a substantial change during our sample period in the composition of its manufacturing industry. For brevity, we omit this analysis.

**Figure 9****Evolution of Manufacturing Share**

Idaho vs. NRTW States



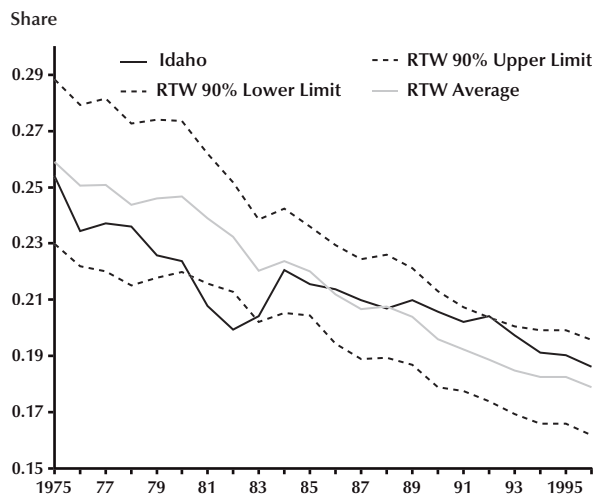
during the 1975-82 period and declined at a much faster rate than the average share in NRTW states. This trend slowly started to change around 1982; from 1984 onward, the manufacturing share in Idaho was above the NRTW average and declined much more slowly, which is consistent with accelerated growth in Idaho's manufacturing employment in this period. By 1987, Idaho's share exceeded the NRTW average, and the difference gradually became statistically significant. By the end of the analysis period, we can reject the hypothesis that Idaho had a manufacturing share similar to an "average" NRTW state with, initially, a similar industrial composition. The comparison with the RTW states' average share in Figure 10 is consistent with this finding. While Idaho's share was much lower than the average RTW states' share before 1982, it gradually became closer to the average afterward.<sup>13</sup>

### Average Establishment Size

Considering the results in Table 2 regarding the change in average establishment size, defined by the number of employees per establishment, we do not observe any definitive pattern. In all states, the difference in the average growth rates in this variable across the two periods was insignificant. This, however, does not necessarily mean that Idaho did not become an attractive location for larger plants or that existing plants had less incentive to

**Figure 10****Evolution of Manufacturing Share**

Idaho vs. RTW States



expand. It is well-known that there has been an ongoing nationwide trend toward smaller establishments.<sup>14</sup> It is possible that the increasing fraction of small plants in Idaho masked the increasing importance of larger establishments. To investigate this possibility, we look at the evolution of two measures: (i) the fraction of manufacturing employment in large establishments and (ii) the average size of large establishments. Following Holmes (1998), we define an establishment as "large" if it has at least 100 employees.<sup>15</sup> If large establishments became more important in Idaho's manufacturing sector after the law, then the first measure is expected to be higher in the post-law period. Similarly, if existing large establishments expanded, or if new large establishments that chose Idaho as a location after the law were larger than their pre-law counterparts on average, then we should see an increase in the second measure, too.

As Table 4 clearly indicates, the two variables

<sup>13</sup> The observations in this section also apply if we consider *all* RTW and NRTW states, not just those with an industrial mix similar to that of Idaho.

<sup>14</sup> See, for example, Davis (1990) and Davis and Haltiwanger (1990). The trend toward smaller establishment sizes might also be responsible for declining unionization, as explored by Even and Macpherson (1990).

<sup>15</sup> This choice is somewhat ad hoc, but as reported by Holmes (1998), 70 percent of all manufacturing establishments in 1992 were classified in this category. Outside manufacturing, the figure was 38 percent.

Table 4

## Large Establishments in Manufacturing: Idaho and Its Neighbors

|               | Average fraction of employment<br>in large establishments |                 | Average establishment size<br>in large establishments |                 |
|---------------|---|-----------------|---|-----------------|
|               | 1975-86   | 1987-96         | 1975-86   | 1987-96         |
| Idaho         | 0.66<br>[0.015]   | 0.68<br>[0.007] | 324.6<br>[16.3]                                       | 348.2<br>[10.8] |
|               | <b>(2.97)</b>   |                 | <b>(4.03)</b>   |                 |
| Washington    | 0.70<br>[0.013]   | 0.69<br>[0.018] | 444.1<br>[25.3]                                       | 451.6<br>[42.1] |
|               | (-1.06)   |                 | (0.49)  |                 |
| Oregon        | 0.63<br>[0.016]   | 0.61<br>[0.007] | 157.9<br>[7.5]  | 148.4<br>[3.7]  |
|               | <b>(-3.69)</b>  |                 | <b>(-3.84)</b>  |                 |
| Montana       | 0.51<br>[0.035]   | 0.43<br>[0.028] | 261.6<br>[22.6]                                       | 224.8<br>[11.1] |
|               | <b>(-5.84)</b>  |                 | <b>(-4.96)</b>  |                 |
| Nevada (RTW)  | 0.51<br>[0.020]   | 0.47<br>[0.035] | 236.5<br>[21.4]                                       | 236.6<br>[16.3] |
|               | <b>(-2.84)</b>  |                 | (0.01)  |                 |
| Utah (RTW)    | 0.68<br>[0.015]   | 0.68<br>[0.011] | 355.3<br>[30.0]                                       | 358.3<br>[11.0] |
|               | (0.01)  |                 | (0.32)  |                 |
| Wyoming (RTW) | 0.43<br>[0.035]   | 0.42<br>[0.020] | 198.9<br>[16.6]                                       | 184.3<br>[8.9]  |
|               | (-0.31)   |                 | (-2.48)   |                 |

NOTE: Standard deviations in brackets. Figures in parentheses are the t statistics associated with the test for the equality of the variable's average across two periods of analysis. Figures in bold indicate significance at 1 percent; t tests are based on unpaired comparisons with unequal variances.

measuring the importance of large establishments in the manufacturing sector experienced a significant increase in Idaho after the law was passed, but this did not occur in any of the neighboring states. There was about a 3 percent increase in the average fraction of employment in large establishments after the law, and the average establishment size for large establishments grew by about 24 employees, or by 7 percent. These results are consistent with the view (i) that Idaho became an attractive location for large establishments after the RTW law was passed and (ii) that the importance of large establishments in the manufacturing sector increased.

## CONCLUSION

We have examined the impact of RTW laws on a state's industrial performance using Idaho's recent experience. We have presented evidence that, even

as a late adopter of the law, Idaho experienced a strong decline in unionization and an acceleration in manufacturing growth. Evidence from Idaho's neighbors suggests that a similar pattern was not experienced by other states in the region, which indicates that a regional boom is not a likely explanation. We are cautious, however, in associating the increase in manufacturing growth with the passage of the law. The exact starting time of the decline in unionization and the narrow time frame of fluctuations in the unionization rate before the passage of the law suggest that the relation is not clear cut. The initial decline in unionization and its subsequent rebounding between 1984 and 1987 can potentially be related also to evolving expectations about the eventual ruling on the RTW law—because the bureaucratic process and political battles for the passing of the RTW law took almost two years, with several developments in favor of and against union-

ism. Adding to our skepticism is the Bunker Hill incidence mentioned earlier, which, by itself, may have been a turning point for the attitudes toward unions in Idaho. In summary, while we are tempted to associate the growth patterns and the decline in unionization with the passage of the law, we cannot rule out the possibility that the RTW law was a result of growing anti-unionism in Idaho and may not have been the cause of growth, per se.

In terms of policy implications, one has to be cautioned before prematurely claiming that Idaho's exceptional growth pattern would apply to every state considering the adoption of the law. Idaho's experience would definitely be more valuable than the evidence from other RTW legislation in the past because it took place in an environment where unionization had already lost considerable ground. As the analysis presented here suggests, even the process leading to the passage of the law may be quite important for the timing of events and the patterns of growth in key variables. Examining the behavior of union organizing activity through certification elections, as well as analyzing the effects on wages, can provide a more detailed analysis of the impact of the RTW law on unionization. The recent experience of Oklahoma, together with Idaho's, can be used for this purpose. The ongoing work by Dinlersoz and Hernández-Murillo (2001) aims to provide more evidence in this direction.

## REFERENCES

- Abraham, Steven E. and Voos, Paula B. "Right-to-Work Laws: New Evidence from the Stock Market." *Southern Economic Journal*, October 2000, 67(2), pp. 345-62.
- Davis, Steven J. "Size Distribution Statistics from County Business Patterns Data." Working paper, University of Chicago, 1990.
- \_\_\_\_\_ and Haltiwanger, John. "The Distribution of Employees by Establishment Size: Patterns of Change in the United States: 1962 to 1985." Unpublished manuscript, University of Chicago, 1990.
- Dinlersoz, Emin M. and Hernández-Murillo, Rubén. "A Recent Assessment of the RTW Laws' Effect in the Wake of Idaho's Experience." Working paper, 2001.
- Ellwood, David T. and Fine, Glenn. "The Impact of Right-to-Work Laws on Union Organizing." *Journal of Political Economy*, April 1987, 95(2), pp. 250-73.
- Even, William E. and Macpherson, David A. "Plant Size and the Decline of Unionism." *Economics Letters*, April 1990, 32(4), pp. 393-98.
- Galarneau, Diane. "Unionized Workers." *Perspectives on Labour and Income*, Spring 1996, 8(1), pp. 43-52.
- Goldfield, Michael. *The Decline of Organized Labor in the United States*. Chicago: The University of Chicago Press, 1987.
- Hirsch, Barry T.; Macpherson, David and Vroman, Wayne G. "Estimates of Union Density by State." *Monthly Labor Review*, July 2001, 124(7), pp. 51-55.
- Holmes, Thomas J. "The Effect of State Policies on the Location of Manufacturing: Evidence from State Borders." *Journal of Political Economy*, August 1998, 106(4), pp. 667-705.
- Kendrick, David. "Right-to-Work—The Idaho Experience." National Institute for Labor Relations Research, delivered at the Fraser Institute Conference on Right to Work, Toronto, Canada, 21 June 1996.
- Long, Richard J. "The Effect of Unionization on Employment Growth of Canadian Companies," *Industrial and Labor Relations Review*, July 1993, 46(4), pp. 691-703.
- Lowe, Graham S. "The Future of Work: Implications for Unions." *Relations Industrielles/Industrial Relations*, 1998, 53(2), pp. 1-25.
- Moore, William J. and Newman, Robert J. "The Effects of Right-to-Work Laws: A Review of the Literature." *Industrial and Labor Relations Review*, July 1985, 38(4), pp. 571-85.
- \_\_\_\_\_. "The Determinants and Effects of Right-to-Work Laws: A Review of the Recent Literature." *Journal of Labor Research*, Summer 1998, 19(3), pp. 445-69.
- National Association of State Development Agencies. *Directory of Incentives for Business Investment and Development in the United States: A State-by-State Guide*. Washington, DC: The Urban Institute Press, various years.



## Appendix

### DATA DESCRIPTION

#### *Unionization Rates*

Estimates of union membership rates by state and by state industry were obtained using the May files of the Census Bureau's Current Population Survey (CPS) for the period 1977-81, and from the Merged Outgoing Rotation Groups CPS files for the period 1983-2000, following the methodology of Hirsch, Macpherson, and Vroman (2001). The 1982 CPS did not include any questions pertaining to unions, and we set our estimate for 1982 to the average of the estimates in 1981 and 1983. For 1983 and onward, each year included all 12 months of the CPS, with each month including the outgoing rotation groups that were asked the union questions. Prior to 1981, the May surveys administered the union questions to *all* rotation groups; therefore the estimates before 1981 are based on samples that are one third of the size of the samples used after 1983. The May 1981 CPS administered the union questions only to the *outgoing* rotation groups, making sample sizes roughly one-third of the samples used in 1977-80. Union estimates for 1981 are, therefore, the least reliable.<sup>16</sup>

Due to the varying sample sizes, much of the year-to-year variation in the estimated unionization rates before 1983 can be attributed to sampling error. This would be a more serious problem if one wished to reliably estimate union earnings, for example, as opposed to simply estimate union membership rates as we did. The sample sizes of major industry groups in Idaho (overall, manufacturing, and nonmanufacturing) were within the standard measures used in the literature and the Census Bureau's guidelines (larger than 100 employ-

ees), except for manufacturing and particularly in 1981.

We were able to verify that our estimates of the proportion of union members from the employed population closely matched those of Hirsch, Macpherson, and Vroman (2001) at the national and state levels. Our estimates of state-industry rates use the same methodology, but there were no available series to verify accuracy.

#### *Data on Industries*

The data on industries come from the Census Bureau's County Business Patterns data series for the years 1975-96. The data covers all taxpaying establishments with one or more paid employees. The employment figures are taken from the mid-March period of every year. An establishment is defined as a single location where business is conducted or where services or industrial operations are performed. Establishment size designations are measured by paid employment in the mid-March pay period. Establishment counts for 1983 and onward are based on a determination of active status as of anytime during the year. For the years prior to 1983, establishment counts are based on whether the establishment was active in the fourth quarter. The data is available at the national, state, and county levels. Further details on this data set can be obtained from the Census Bureau's Web site, < [www.census.gov](http://www.census.gov) > .

<sup>16</sup> Every household that enters the CPS is interviewed each month for 4 months, then ignored for 8 months, then interviewed again for 4 more months. The union questions are asked only to households in their fourth and eighth interview. These are the outgoing rotation groups.



# Predicting Exchange Rate Volatility: Genetic Programming Versus GARCH and RiskMetrics™

Christopher J. Neely and Paul A. Weller

**I**t is well established that the volatility of asset prices displays considerable persistence. That is, large movements in prices tend to be followed by more large moves, producing positive serial correlation in squared returns. Thus, current and past volatility can be used to predict future volatility. This fact is important to both financial market practitioners and regulators.

Professional traders in equity and foreign exchange markets must pay attention not only to the expected return from their trading activity but also to the risk that they incur. Risk-averse investors will wish to reduce their exposure during periods of high volatility, and improvements in risk-adjusted performance depend upon the accuracy of volatility predictions. Many current models of risk management, such as Value-at-Risk (VaR), use volatility predictions as inputs.

The bank capital adequacy standards recently proposed by the Basel Committee on Banking Supervision illustrate the importance of sophisticated risk management techniques for regulators. These norms are aimed at providing international banks with greater incentives to manage financial risk in a sophisticated fashion, so that they might economize on capital. One such system that is widely used is RiskMetrics™, developed by J.P. Morgan.

A core component of the RiskMetrics system is a statistical model—a member of the large ARCH/GARCH family—that forecasts volatility. Such ARCH/

GARCH models are parametric. That is, they make specific assumptions about the functional form of the data generation process and the distribution of error terms. Parametric models such as GARCH are easy to estimate and readily interpretable, but these advantages may come at a cost. Other, perhaps much more complex models may be better representations of the underlying data generation process. If so, then procedures designed to identify these alternative models have an obvious payoff. Such procedures are described as nonparametric. Instead of specifying a particular functional form for the data generation process and making distributional assumptions about the error terms, a nonparametric procedure will search for the best fit over a large set of alternative functional forms.

This article investigates the performance of a genetic program applied to the problem of forecasting volatility in the foreign exchange market. Genetic programming is a computer search and problem-solving methodology that can be adapted for use in nonparametric estimation. It has been shown to detect patterns in the conditional mean of foreign exchange and equity returns that are not accounted for by standard statistical models (Neely, Weller, and Dittmar, 1997; Neely and Weller, 1999, 2001; Neely, 2001). These achievements suggest that a genetic program may also be a powerful tool for generating predictions of asset price volatility.

We compare the performance of a genetic program in forecasting daily exchange rate volatility for the dollar–Deutsche mark and dollar-yen exchange rates with that of a GARCH(1,1) model and a related RiskMetrics volatility forecast (described in the following section). These models are widely used by both academics and practitioners and thus are good benchmarks with which to compare the genetic program forecasts. While the overall forecast performance of the two methods is broadly similar, on some dimensions the genetic program produces significantly superior results. This encouraging finding suggests that more detailed investigation of this methodology applied to volatility forecasting would be warranted.

## THE BENCHMARK MODEL

Before discussing the genetic programming procedure, we will review the benchmark GARCH and RiskMetrics volatility models. Engle (1982) developed the autoregressive conditionally heteroskedastic (ARCH) model to characterize the observed serial correlation in asset price volatility. Suppose

---

Christopher J. Neely is a research officer at the Federal Reserve Bank of St. Louis. Paul A. Weller is a professor of finance at the Henry B. Tippie College of Business Administration of the University of Iowa. This paper is a revised and expanded version of a chapter entitled "Using a Genetic Program to Predict Exchange Rate Volatility," in *Genetic Algorithms and Genetic Programming in Computational Finance*, edited by Shu-Heng Chen, published by Kluwer Academic Publishers. We would like to thank Janis Zvingelis for excellent programming assistance. Charles Hokayem provided research assistance.

© 2002, The Federal Reserve Bank of St. Louis.

we assume that a price  $P_t$  follows a random walk,

$$(1) \quad P_{t+1} = P_t + \varepsilon_{t+1},$$

where  $\varepsilon_{t+1} \sim N(0, \sigma_t^2)$ . The variance of the error term depends upon  $t$ , and the objective of the model is to characterize the way in which this variance changes over time. The ARCH model assumes that this dependence can be captured by an autoregressive process of the form

$$(2) \quad \sigma_t^2 = \omega + \alpha_0 \varepsilon_t^2 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_m \varepsilon_{t-m}^2,$$

where the restrictions  $\omega \geq 0$  and  $\alpha_i \geq 0$  for  $i = 0, 1, \dots, m$  ensure that the predicted variance is always non-negative. This specification illustrates clearly how current levels of volatility will be influenced by the past and how periods of high or low price fluctuation will tend to persist.

Bollerslev (1986) extended the ARCH class to produce the generalized autoregressive conditionally heteroskedastic (GARCH) model, in which the variance is given by

$$(3) \quad \sigma_t^2 = \omega + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-2}^2 + \cdots + \beta_k \sigma_{t-k}^2 + \alpha_0 \varepsilon_t^2 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_m \varepsilon_{t-m}^2.$$

The simplest specification in this class, and the one most widely used, is referred to as GARCH(1,1) and is given by

$$(4) \quad \sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \alpha \varepsilon_t^2.$$

When  $\alpha + \beta < 1$ , the variance process displays mean reversion to the unconditional expectation of  $\sigma_t^2$ ,  $\omega/(1 - \alpha - \beta)$ . That is, forecasts of volatility in the distant future will be equal to the unconditional expectation of  $\sigma_t^2$ ,  $\omega/(1 - \alpha - \beta)$ .

The RiskMetrics model for volatility forecasting imposes the restrictions that  $\alpha + \beta = 1$  and that  $\omega = 0$ .<sup>1</sup> In addition, the parameter  $\beta$  is not estimated, but imposed to be equal to 0.94 (J.P. Morgan/Reuters, 1996). This value was found to minimize the mean-squared error (MSE) of volatility forecasts for asset prices. The RiskMetrics one-day-ahead volatility forecast is

$$(5) \quad \sigma_t^2 = \beta \sigma_{t-1}^2 + (1 - \beta) \varepsilon_t^2.$$

The GARCH model has been used to characterize patterns of volatility in U.S. dollar foreign exchange markets (Baillie and Bollerslev, 1989, 1991) and in the European Monetary System (Neely, 1999). However, initial investigations into the explanatory power of out-of-sample forecasts produced disappointing

results (West and Cho, 1995). Jorion (1995) found that volatility forecasts for several major currencies from the GARCH model were outperformed by implied volatilities generated from the Black-Scholes option-pricing model. These studies typically used the squared daily return as the variable to be forecast. However, the squared return is a very imprecise measure of true, unobserved volatility. For example, the exchange rate may move around a lot during the day, and yet end up close to its value the same time the previous day. In this case, the squared daily return would be small, even though volatility was high. More recently, it has been demonstrated that one can significantly improve the forecasting power of the GARCH model by measuring volatility as the sum of intraday squared returns (Andersen and Bollerslev, 1998). This measure is referred to as integrated, or realized, volatility. In theory, if the true underlying price path is a diffusion process, it is possible to obtain progressively more accurate estimates of the true volatility by increasing the frequency of intraday observation. Of course, there are practical limits to this; microstructural effects begin to degrade accuracy beyond a certain point.

## GENETIC ALGORITHMS AND GENETIC PROGRAMMING

*Genetic algorithms* are computer search procedures used to solve appropriately defined problems. The structure of the search procedure is based on the principles of natural selection. These procedures were developed for genetic algorithms by Holland (1975) and extended to *genetic programming* by Koza (1992). The essential features of both algorithms include (i) a means of representing potential solutions to a problem as character strings that can be split up and recombined to form new potential solutions and (ii) a fitness criterion that measures the “quality” of a candidate solution. Both types of algorithms produce successive “generations” of candidate solutions using procedures that mimic genetic reproduction and recombination. Each new generation is subjected to the pressures of “natural selection” by increasing the probability that candidate solutions scoring highly on the fitness criterion get to reproduce.

To understand the principles involved in genetic

<sup>1</sup> The restriction,  $\alpha + \beta = 1$ , implies that shocks to the volatility process persist forever; higher volatility today will lead one to forecast higher volatility indefinitely. It therefore falls into the class of integrated GARCH or IGARCH models.



programming, it is useful to understand the operation of the simpler genetic algorithm. Genetic algorithms require that potential solutions be expressed as fixed-length character strings. Consider a problem in which candidate solutions are mapped into binary strings  $s$ , with a length of five digits. One possible solution would be represented as (01010). Associated with this binary string would be a measure of fitness that quantifies how well it solves the problem. In other words, we need a fitness function  $m(s)$  that maps the strings into the real line and thus ranks the quality of the solutions. Next we introduce the *crossover operator*. Given two strings, a crossover point is randomly selected and the first part of one string is combined with the second part of the other. For example, given the two strings (00101) and (11010) and a crossover point between elements two and three, the new string (00010) is generated. The remaining parts of the original strings are discarded.

The algorithm begins by randomly generating an *initial population* of binary strings and then evaluating the fitness of each string by applying the fitness function  $m(s)$ . Next, the program produces a new (second) generation of candidate solutions by selecting pairs of strings at random from this initial population and applying the crossover operator to create new strings. The probability of selecting a given string is set to be proportional to its fitness. Thus a “selection pressure” in favor of progressively superior solutions is introduced. This process is repeated to produce successive generations of strings, keeping the size of each generation the same. The procedure “evolves” new generations of improved potential solutions.

Recall that genetic algorithms require that potential solutions be encoded as fixed length character strings. Koza’s (1992) extension, genetic programming, instead employs variable-length, hierarchical strings that can be thought of as decision trees or computer programs. However, the basic structure of a genetic program is exactly the same as described above. In particular, the crossover operator is applied to pairs of decision trees to generate new “offspring” trees.

The application in this paper represents forecasting functions as trees and makes use of the following function set in constructing them: plus, minus, times, divide, norm, log, exponential, square root, and cumulative standard normal distribution function. In addition, we supply the following set of data functions: *data*, *average*, *max*, *min*, and *lag*. The data functions can operate on any of the four

data series that we provide as inputs to the genetic program: (i) daily foreign exchange returns, (ii) integrated volatility (i.e., the sum of squared intraday returns), (iii) the sum of the absolute value of intraday returns, and (iv) the number of days until the next business day. For example,  $data(returns(t))$  is simply the identity function that computes the daily return at  $t$ . The other data functions operate in a similar fashion, but also take numerical arguments to specify the length of the window—the number of observations—over which the functions operate. The numerical arguments that the functions take are determined by the genetic program. Thus  $average(returns(t))(n)$  generates the arithmetic average of the return observations  $t, t-1, \dots, t-n+1$ .

The choice of elements to include in the function set is a potentially important one. While a genetic program can, in principle, produce a very highly complex solution from simple functions, computational limitations might make such solutions very difficult to find in practice. Providing specialized functions to the genetic program that are thought to be useful to a “good” solution to the problem can greatly increase the efficiency of the search by encouraging the genetic program to search in the area of the solution space containing those functions. On the other hand, this might bias the genetic program’s search away from other promising regions. To focus the search in promising regions of the solution space, we investigate the results of adding three additional complex data functions to the original set of functions. This is described below.

The expanded set of data functions consists of the original set plus *geo*, *mem*, and *arch5*. Each of these functions approximates the forecast of a known parametric model of conditional volatility. Thus, the genetic program might find them useful. The function *geo* returns the following weighted average of ten lags of past data:

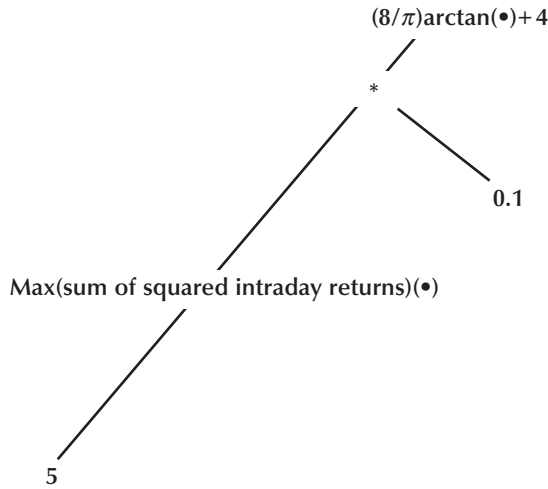
$$(6) \quad geo(data)(\alpha) \equiv \sum_{j=0}^9 \alpha \left[ (1-\alpha)^j \right] data_{t-j}.$$

This function can be derived from the prediction of an IGARCH specification with parameter  $\alpha$ , where we constrain  $\alpha$  to satisfy  $0.01 \leq \alpha \leq 0.99$  and ten lags of data are used. The function *mem* returns a weighted sum similar to that which would be obtained from a long memory specification for volatility. It takes the form

$$(7) \quad mem(data)(d) \equiv \sum_{j=0}^9 h_j data_{t-j},$$

Figure 1

## Example of a Hypothetical Forecast Function



where  $h_0 = 1$ ,  $h_j \propto (1/j!)(d+j-1)(d+j-2)\dots(d+1)d$  for  $j > 0$  and the sum of the coefficients  $h_j$  is constrained to equal 1 so that the output would be of the same magnitude as recent volatility. The parameter  $d$  is determined by the genetic program and constrained to satisfy  $-1 < d < 1$ . Finally, the function *arch5* permits a flexible weighting of the five most recent observations, where the values for  $h_j$  are provided by the genetic program and constrained to lie within  $\{-5, 5\}$  and to sum to 1. Again, the constraint on the sum of the coefficients ensures that the magnitude of the output will be similar to that of recent volatility. The function has the form

$$(8) \quad \text{arch5}(\text{data})(h) \equiv \sum_{j=0}^4 h_j \text{data}_{t-j}, \quad h = (h_0, h_1, \dots, h_4).$$

Figure 1 illustrates a simple example of a hypothetical tree determining a forecasting function. The function first computes the maximum of the sum of squared intraday returns over the last five days. This number is multiplied by 0.1, and the result is entered as the argument  $x$  of the function  $(8/\pi)\arctan(x) + 4$ . This latter function is common to all trees and maps the real line into the interval  $(0, 8)$ . It ensures that all forecasts are nonnegative and bounded above by a number chosen with reference to the characteristics of the in-sample period.

We now turn to the form of the fitness criterion. Because true volatility is not directly observed, it is necessary to use an appropriate proxy in order to assess the volatility forecasting performance of

the genetic program. One possibility is to use the ex post squared daily return. However, as Andersen and Bollerslev (1998) have pointed out, this is an extremely noisy measure of the true underlying volatility and is largely responsible for the apparently poor forecast performance of GARCH models. A better approach is to sum intraday returns to measure true daily volatility (i.e., integrated volatility) more accurately. We measure integrated volatility using five irregularly spaced intraday observations. If  $S_{i,t}$  is the  $i$ th observation on date  $t$ , we define

$$(9) \quad R_{i,t} = 100 \cdot \ln \left( \frac{S_{i+1,t}}{S_{i,t}} \right) \quad \text{for } i = 1, 2, 3, \text{ and } 4$$

$$= 100 \cdot \ln \left( \frac{S_{1,t+1}}{S_{5,t}} \right) \quad \text{for } i = 5$$

$$(10) \quad \sigma_{I,t}^2 = \sum_{i=1}^5 R_{i,t}^2.$$

Thus  $\sigma_{I,t}^2$  is the measure of integrated volatility on date  $t$ .<sup>2</sup> Using five intraday observations represents a compromise between the increase in accuracy generated by more frequent observations and the problems of data handling and availability that arise as one moves to progressively higher frequencies of intraday observation.

In constructing the rules, the genetic program minimized the mean-squared forecast error (MSE) as the fitness criterion. There are potential inefficiencies involved in using this criterion on heteroskedastic data. However, a heteroskedasticity-corrected fitness measure proved unsatisfactory in experiments. With three to five observations per day, there were instances where the integrated daily volatility was very small; the heteroskedasticity correction caused the measure to be inappropriately sensitive to those observations.<sup>3</sup>

<sup>2</sup> More precisely, daily volatility is calculated from 1700 Greenwich Mean Time (GMT) to 1700 GMT.

<sup>3</sup> A perennial problem with using flexible, powerful search procedures like genetic programming is overfitting—the finding of spurious patterns in the data. Given the well-documented tendency for the genetic program to overfit the data, it is necessary to design procedures to mitigate this (e.g., Neely, Weller, and Dittmar, 1997). Here, we investigated the effect of modifying the fitness criterion by adding a penalty for complexity. This penalty consisted of subtracting an amount  $(0.002 \times \text{number of nodes})$  from the negative MSE. Nodes are data and numerical functions. This modification is intended to bias the search toward functions with fewer nodes, which are simpler and therefore less prone to overfit the data. Unfortunately, this procedure produced no significant changes in performance, so we will report results only from the unmodified version.

**Table 1****Data Type and Source**

| Time | Source                           | Type of price                     |
|------|----------------------------------|-----------------------------------|
| 1000 | Swiss National Bank              | Triangular arbitrage on bid rates |
| 1400 | Federal Reserve Bank of New York | Midpoint of bid and ask           |
| 1600 | Bank of England                  | Triangular arbitrage, unspecified |
| 1700 | Federal Reserve Bank of New York | Midpoint of bid and ask           |
| 2200 | Federal Reserve Bank of New York | Midpoint of bid and ask           |

**DATA AND IMPLEMENTATION**

The object of this exercise is to forecast the daily volatility (the sum of intraday squared returns) of two currencies against the dollar, the German mark (DEM) and Japanese yen (JPY), over the period June 1975 to September 1999. Thus, the final nine months of data for the DEM represent the rate derived from that of the euro, which superseded the DEM in January 1999. The timing of observations was 1000, 1400, 1600, 1700, and 2200 GMT. Days with fewer than three valid observations or no observation at 1700 were treated as missing. In addition, weekends were excluded. The sources of the data for both exchange rates are summarized in Table 1. We provided the genetic program with three series in addition to the integrated volatility series: daily returns, sum of absolute intraday returns, and the number of days until the next trading day.

The full sample is divided into three subperiods: the training period June 1975 through December 1979; the selection period January 1980 through December 30, 1986; and the out-of-sample period December 31, 1986, through September 21, 1999. The role of these subperiods is described below.

In searching through the solution space of forecasting functions, the genetic program followed the procedures below.

1. Create an initial generation of 500 randomly generated forecast functions.
2. Measure the MSE of each function over the training period and rank according to performance.
3. Select the function with the lowest MSE and calculate its MSE over the selection period. Save it as the initial best forecast function.
4. Select two functions at random, using weights attaching higher probability to more highly ranked functions. Apply the crossover oper-

ator to create a new function, which then replaces an old function, chosen using weights attaching higher probability to less highly ranked functions. Repeat this procedure 500 times to create a new generation of functions.

5. Measure the MSE of each function in the new generation over the training period. Take the best function in the training period and evaluate the MSE over the selection period. If it outperforms the previous best forecast, save it as the new best forecast function.
6. Stop if no new best function appears for 25 generations, or after 50 generations. Otherwise, return to stage 4.

The stages above describe one trial. Each trial produces one forecast function. The results of each trial will generally differ as a result of sampling variation. For this reason it is necessary to run a number of trials and then to aggregate the results. The aggregation methods are described in the following section.

**RESULTS**

The benchmark results are those from the GARCH(1,1) and RiskMetrics models described in the Benchmark Model section, estimated over the in-sample period June 1975 to December 30, 1986. We forecast daily integrated volatility (defined in equations (9) and (10)) from these models, in and out of sample, at horizons of 1, 5, and 20 days.<sup>4</sup>

We also forecast with a genetic program whose training and selection periods coincide with the in-sample estimation period for the GARCH model. For each case of the genetic program we generated ten trials, each of which produced a forecast function.

<sup>4</sup> Note that the forecasted variable at the 5-day (20-day) horizon is the integrated volatility 5 (20) days in the future. It is not the sum of the next 5 (20) days of integrated volatility.

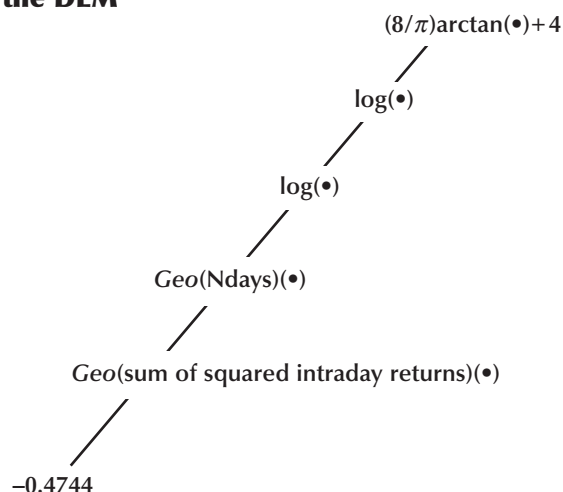
Table 2

**In-Sample Comparison of Genetic Program, GARCH, and RiskMetrics™: The Baseline Case**

| Exchange rate | Horizon | MSE   |       |       |      | MAE   |       |       |      | R <sup>2</sup> |       |       |      |
|---------------|---------|-------|-------|-------|------|-------|-------|-------|------|----------------|-------|-------|------|
|               |         | EW GP | MW GP | GARCH | RM   | EW GP | MW GP | GARCH | RM   | EW GP          | MW GP | GARCH | RM   |
| DEM           | 1       | 0.50  | 0.53  | 0.50  | 0.49 | 0.30  | 0.33  | 0.33  | 0.33 | 0.18           | 0.15  | 0.16  | 0.14 |
| DEM           | 5       | 0.56  | 0.59  | 0.56  | 0.52 | 0.31  | 0.34  | 0.37  | 0.34 | 0.12           | 0.11  | 0.10  | 0.10 |
| DEM           | 20      | 0.61  | 0.63  | 0.67  | 0.56 | 0.33  | 0.34  | 0.46  | 0.37 | 0.08           | 0.04  | 0.04  | 0.05 |
| JPY           | 1       | 0.56  | 0.58  | 0.60  | 0.62 | 0.32  | 0.32  | 0.38  | 0.37 | 0.22           | 0.20  | 0.14  | 0.08 |
| JPY           | 5       | 0.65  | 0.65  | 0.73  | 0.66 | 0.36  | 0.37  | 0.43  | 0.38 | 0.06           | 0.04  | 0.02  | 0.04 |
| JPY           | 20      | 0.66  | 0.67  | 0.71  | 0.69 | 0.38  | 0.39  | 0.51  | 0.40 | 0.05           | 0.03  | 0.01  | 0.02 |

NOTE: The in-sample mean-squared error (MSE), mean absolute error (MAE), and R<sup>2</sup> from GARCH(1,1), RiskMetrics™ (RM), and genetic program (GP) forecasts on DEM/USD and JPY/USD data at three forecast horizons: 1, 5, and 20 days. The GP forecast was generated using five data functions and without a penalty for complexity. In columns 3, 7, and 11 we report the forecast statistics—MSE, MAE, and R<sup>2</sup>—for the equally weighted (EW) genetic programming method. In columns 4, 8, and 12 we report the analogous statistics for the median-weighted (MW) genetic programming forecast. Columns 5, 9, and 13 contain the results for the GARCH forecast. Columns 6, 10, and 13 contain RiskMetrics™ forecast statistics. The in-sample period was June 1975 to December 30, 1986.

Figure 2

**One-Day-Ahead Forecasting Functions for the DEM**

The cases were distinguished by the following factors: (i) forecast horizon—1, 5, and 20 days; (ii) the number of data functions—five or eight. For each case, we generated ten rules. The forecasts from each set of ten rules were aggregated in two ways. The equally weighted forecast is the arithmetic average of the forecasts from each of the ten trials. The median-weighted forecast takes the median forecast from the set of ten forecasts at each date. We report six measures of out-of-sample forecast performance:

MSE, mean absolute error (MAE), R<sup>2</sup>, mean forecast bias, kernel estimates of the error densities, and generalized mean-squared forecast error matrix tests.

Before discussing the results, we first present a simple example of the forecasting rules produced by the genetic program. Figure 2 illustrates a one-day-ahead forecasting function for the DEM. Its out-of-sample MSE was 0.496. The function is interpreted as follows. The number  $-0.4744$  at the terminal node enters as the argument of  $geo(\text{sum of squared intraday returns})$ . Since the argument of  $geo$  is constrained to lie between 0.01 and 0.99, it is set to 0.01. The number generated by this function then enters as the argument in  $geo(Ndays)$ , where  $Ndays$  refers to the “number of days to the next trading day.” We caution that this example was chosen largely because of its relatively simple form; some trials generated rules that were considerably more complex, with as many as 10 levels and/or 100 nodes.

Table 2 reports in-sample results for the baseline case with five data functions. The figures for MSE for the DEM are very similar for the GARCH and equally weighted genetic program forecasts at the 1- and 5-day horizons, but the genetic program is appreciably better at the 20-day horizon. The median-weighted forecast is generally somewhat inferior to the equally weighted forecast, but follows the same pattern over the forecast horizons relative to the GARCH model. That is, its best relative performance is at the 20-day horizon. The RiskMetrics forecasts also are generally comparable to GARCH



**Table 3****Out-of-Sample Comparison of Genetic Program, GARCH, and RiskMetrics™: The Baseline Case**

| Exchange rate | Horizon | MSE   |       |       |      | MAE   |       |       |      | R <sup>2</sup> |       |       |      |
|---------------|---------|-------|-------|-------|------|-------|-------|-------|------|----------------|-------|-------|------|
|               |         | EW GP | MW GP | GARCH | RM   | EW GP | MW GP | GARCH | RM   | EW GP          | MW GP | GARCH | RM   |
| DEM           | 1       | 0.35  | 0.38  | 0.33  | 0.32 | 0.30  | 0.34  | 0.33  | 0.32 | 0.09           | 0.08  | 0.12  | 0.10 |
| DEM           | 5       | 0.38  | 0.42  | 0.36  | 0.34 | 0.31  | 0.35  | 0.35  | 0.33 | 0.06           | 0.06  | 0.08  | 0.07 |
| DEM           | 20      | 0.41  | 0.42  | 0.44  | 0.37 | 0.31  | 0.31  | 0.43  | 0.35 | 0.01           | 0.01  | 0.02  | 0.02 |
| JPY           | 1       | 1.35  | 1.35  | 1.29  | 1.33 | 0.42  | 0.44  | 0.47  | 0.47 | 0.14           | 0.13  | 0.16  | 0.11 |
| JPY           | 5       | 1.48  | 1.48  | 1.56  | 1.44 | 0.43  | 0.45  | 0.52  | 0.49 | 0.03           | 0.02  | 0.04  | 0.06 |
| JPY           | 20      | 1.48  | 1.48  | 1.43  | 1.46 | 0.45  | 0.46  | 0.55  | 0.51 | 0.02           | 0.02  | 0.05  | 0.05 |

NOTE: The out-of-sample MSE, MAE, and R<sup>2</sup> from GARCH(1,1), RiskMetrics™ (RM), and genetic program (GP) forecasts on DEM/USD and JPY/USD data at three forecast horizons: 1, 5, and 20 days. The GP forecast was generated using five data functions and without a penalty for complexity. The out-of-sample period was December 31, 1986, to September 21, 1999. See the notes to Table 2 for column definitions.

**Table 4****Out-of-Sample Results Using the Data Functions *Geo*, *Mem*, and *Arch5***

| Exchange rate | Horizon | MSE   |       |       |      | MAE   |       |       |      | R <sup>2</sup> |       |       |      |
|---------------|---------|-------|-------|-------|------|-------|-------|-------|------|----------------|-------|-------|------|
|               |         | EW GP | MW GP | GARCH | RM   | EW GP | MW GP | GARCH | RM   | EW GP          | MW GP | GARCH | RM   |
| DEM           | 1       | 0.37  | 0.44  | 0.33  | 0.32 | 0.29  | 0.37  | 0.33  | 0.32 | 0.12           | 0.05  | 0.12  | 0.10 |
| DEM           | 5       | 0.36  | 0.37  | 0.36  | 0.34 | 0.30  | 0.30  | 0.35  | 0.33 | 0.05           | 0.04  | 0.08  | 0.07 |
| DEM           | 20      | 0.38  | 0.38  | 0.44  | 0.37 | 0.30  | 0.30  | 0.43  | 0.35 | 0.01           | 0.01  | 0.02  | 0.02 |
| JPY           | 1       | 1.27  | 1.31  | 1.29  | 1.33 | 0.43  | 0.44  | 0.47  | 0.47 | 0.18           | 0.15  | 0.16  | 0.11 |
| JPY           | 5       | 1.45  | 1.46  | 1.56  | 1.44 | 0.46  | 0.46  | 0.52  | 0.49 | 0.04           | 0.03  | 0.04  | 0.06 |
| JPY           | 20      | 1.49  | 1.62  | 1.43  | 1.46 | 0.44  | 0.50  | 0.55  | 0.51 | 0.04           | 0.00  | 0.05  | 0.05 |

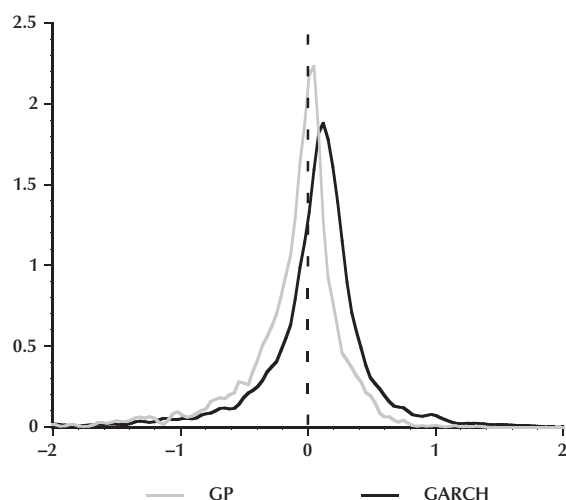
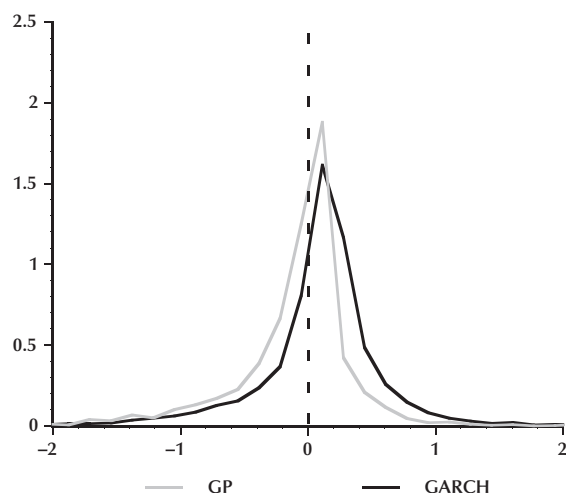
NOTE: The out-of-sample MSE, MAE, and R<sup>2</sup> from GARCH(1,1), RiskMetrics™ (RM), and genetic program (GP) forecasts on DEM/USD and JPY/USD data at three forecast horizons: 1, 5, and 20 days. The GP forecast was generated using eight data functions including *geo*, *mem*, and *arch5* (for descriptions see equations (6) through (8) in the text) and without a penalty for complexity. The out-of-sample period was December 31, 1986, to September 21, 1999. See the notes to Table 2 for column definitions.

forecasts at 1- and 5-day horizons, but a bit better at longer horizons. For the JPY, the genetic program produces equally weighted MSE figures that are in all cases lower than for the GARCH and RiskMetrics models. Similarly, the equally weighted genetic programming rules have higher R<sup>2</sup>s over each horizon than the GARCH and RiskMetrics models. This result is not especially surprising given the flexibility of the nonparametric procedure and its known tendency to overfit in-sample.

Table 3 presents a more interesting comparison—out-of-sample performance over the period

December 31, 1986, through September 21, 1999. The equally weighted genetic program MSE figures are usually slightly larger than those of the GARCH and RiskMetrics forecasts at all horizons for both currencies. Similarly, the genetic programming R<sup>2</sup>s are typically slightly smaller than those of the GARCH/RiskMetrics forecasts. However, the equally weighted genetic program has a lower MAE than do the GARCH/RiskMetrics models at all horizons and for both currencies.

Table 4 reports the out-of-sample performance of the genetic program forecasts using the augmented

**Figure 3****1-Day DEM Forecast Error Densities****1-Day JPY Forecast Error Densities**

NOTE: The kernel estimates of the densities of the 1-day forecast errors (forecast minus realized volatility) for the DEM and JPY for genetic program and GARCH(1,1) model over the out-of-sample period, December 31, 1986, through September 21, 1999. The dotted vertical line denotes zero.

set of data functions, which include *geo*, *mem*, and *arch5*. For ease of comparison Table 4 repeats the out-of-sample figures for the GARCH model. The MSE and  $R^2$  statistics from this table are more equivocal than those from Table 3. The equally weighted genetic program MSE for the DEM cases are slightly larger than those of the GARCH and RiskMetrics forecasts at the 1- and 5-day horizons, but the genetic

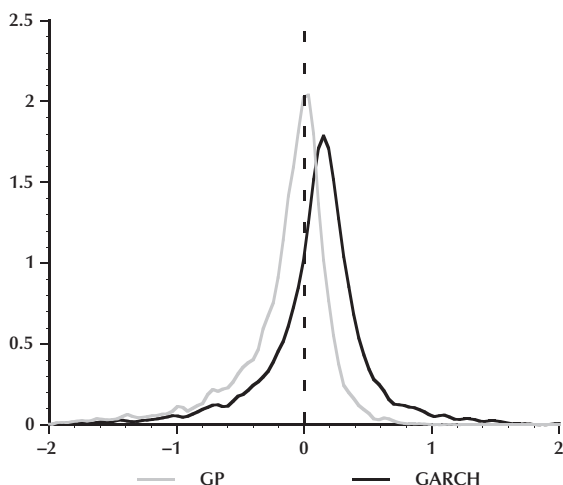
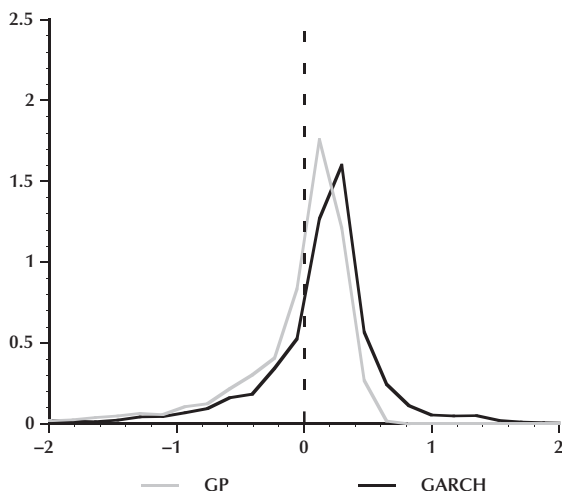
program performs somewhat better than GARCH at the 20-day horizon. This performance is not, however, reflected in the  $R^2$ , for which the GARCH/RiskMetrics models are better at longer horizons. For the JPY the situation is reversed; the equally weighted genetic programming MSE is lower than the GARCH/RiskMetrics figures at the 1-day horizon but larger at the 20-day horizon. The equally weighted genetic program also has a slight edge in  $R^2$  at the 1-day horizon. The figures for the MAE of the genetic program are not very different from Table 3 and are still substantially better than those of the GARCH/RiskMetrics predictions.

To summarize: With MSE as the performance criterion, neither the genetic programming procedure nor the GARCH/RiskMetrics model is clearly superior. The GARCH/RiskMetrics models do achieve slightly higher  $R^2$ s at longer horizons but the MAE criterion clearly prefers the genetic programming forecasts. In both tables, there is some tendency for the median-weighted genetic programming forecast to perform less well than its equally weighted counterpart. The out-of-sample RiskMetrics forecasts are usually marginally better than those of the estimated GARCH model by MSE and MAE criteria but marginally worse when judged by  $R^2$ .

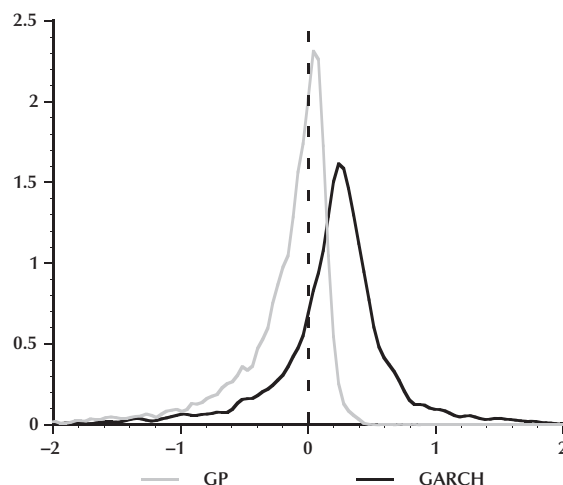
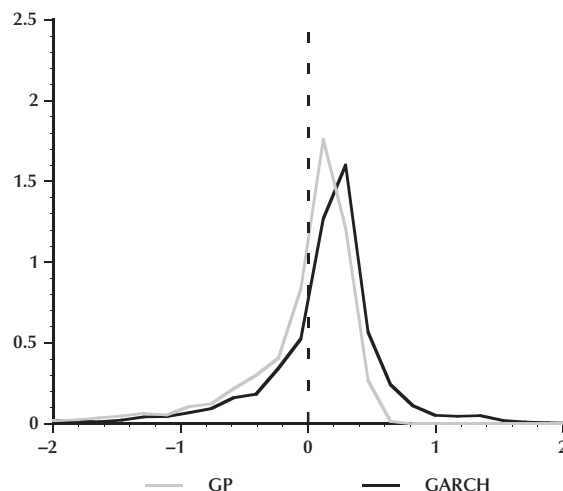
Comparing the genetic programming results in Table 4 with those of Table 3 shows that expanding the set of data functions leads to only a marginal improvement in the performance of the genetic program. Therefore further results will concentrate on out-of-sample forecasts in the baseline genetic programming case presented in Table 3, where only five data functions were used. We present kernel estimates of the densities of out-of-sample forecast errors at the various horizons in Figures 3 through 5.<sup>5</sup>

The most striking feature to emerge from these figures is the apparent bias in the GARCH forecasts when compared with their genetic program counterparts. At all forecast horizons and for both currencies, there is a positive shift in the error distributions of the GARCH forecasts that move the modes of the forecast densities away from zero. However, the relative magnitude of the bias in the mode does not carry over to the mean. Table 5 shows that, though both forecasts are biased in the mean, the magnitude of the bias is considerably greater for the genetic program. Tests for the bias—carried out with a Newey-West correction for serial correlation—show

<sup>5</sup> We choose to graph the density of the GARCH errors because the density of the RiskMetrics errors will have a mean of approximately zero by construction.

**Figure 4****5-Day DEM Forecast Error Densities****5-Day JPY Forecast Error Densities**

NOTE: The kernel estimates of the densities of the 5-day forecast errors (forecast minus realized volatility) for the DEM and JPY for genetic program and GARCH(1,1) model over the out-of-sample period, December 31, 1986, through September 21, 1999. The dotted vertical line denotes zero.

**Figure 5****20-Day DEM Forecast Error Densities****20-Day JPY Forecast Error Densities**

NOTE: The kernel estimates of the densities of the 20-day forecast errors (forecast minus realized volatility) for the DEM and JPY for genetic program and GARCH(1,1) model over the out-of-sample period, December 31, 1986, through September 21, 1999. The dotted vertical line denotes zero.

that all the forecasts are biased in a statistically significant way (Newey and West, 1987, 1994). The evidence from Figures 3 through 5—that the modes of the genetic programming error distribution are closer to zero than those of the GARCH model—indicates that the bias in the genetic programming forecasts is being substantially influenced by a small number of negative outliers.

The MSE and  $R^2$  evidence presented so far fails to indicate a clear preference for any of the four sets of forecasts. The best model varies by forecast horizon and by forecast evaluation criterion. This confused state of affairs leaves one wondering whether these disparate results can be reconciled to produce an unambiguous ranking of the two methodologies. One method by which multi-horizon

Table 5

## Tests for Mean Forecast Bias

| Exchange rate | Horizon | Mean $\sigma^2$ | Equally weighted GP  |       |         | Median-weighted GP   |       |         | GARCH                |      |         | RM                   |       |         |
|---------------|---------|-----------------|----------------------|-------|---------|----------------------|-------|---------|----------------------|------|---------|----------------------|-------|---------|
|               |         |                 | Predicted $\sigma^2$ | Bias  | p value | Predicted $\sigma^2$ | Bias  | p value | Predicted $\sigma^2$ | Bias | p value | Predicted $\sigma^2$ | Bias  | p value |
| DEM           | 1       | 0.43            | 0.29                 | -0.15 | 0.00    | 0.24                 | -0.19 | 0.00    | 0.46                 | 0.03 | 0.00    | 0.43                 | 0.00  | 0.92    |
| DEM           | 5       | 0.43            | 0.23                 | -0.20 | 0.00    | 0.16                 | -0.27 | 0.00    | 0.49                 | 0.05 | 0.00    | 0.43                 | 0.00  | 0.92    |
| DEM           | 20      | 0.43            | 0.19                 | -0.24 | 0.00    | 0.18                 | -0.25 | 0.00    | 0.59                 | 0.16 | 0.00    | 0.43                 | 0.00  | 0.92    |
| JPY           | 1       | 0.56            | 0.33                 | -0.22 | 0.00    | 0.32                 | -0.24 | 0.00    | 0.57                 | 0.02 | 0.06    | 0.55                 | 0.00  | 0.88    |
| JPY           | 5       | 0.56            | 0.37                 | -0.19 | 0.00    | 0.41                 | -0.15 | 0.00    | 0.59                 | 0.04 | 0.07    | 0.55                 | -0.01 | 0.87    |
| JPY           | 20      | 0.56            | 0.42                 | -0.14 | 0.00    | 0.44                 | -0.12 | 0.01    | 0.65                 | 0.09 | 0.02    | 0.55                 | -0.01 | 0.89    |

NOTE: In column 3, mean volatility is the mean daily integrated volatility over the out-of-sample period December 31, 1986, through September 21, 1999. Columns 4, 5, and 6 report the following statistics for the equally weighted genetic programming forecasts over the same period: mean forecast of integrated volatility, the bias in the forecast (predicted volatility minus realized volatility), and the p value for the test that the mean bias is zero. Columns 7 through 9 report the statistics for the median-weighted genetic programming forecasts, and columns 10 through 12 report the analogous results for GARCH forecasts. The RiskMetrics™ statistics are in columns 13 through 15. The genetic program forecasts are based on the five-function model described in Table 3. The p values are computed with Newey-West corrections for heteroskedasticity and serial correlation. The lag length was selected by the Newey and West (1994) procedure.

Table 6

## Test of Generalized Method of Second Forecast Error Moment Domination

|     | Eigenvalues |             |          |          |          |
|-----|-------------|-------------|----------|----------|----------|
|     | GARCH-EW GP | GARCH-MW GP | RM-EW GP | RM-MW GP | GARCH-RM |
| DEM | -0.090      | -0.148      | -0.021   | -0.037   | -0.017   |
|     | 0.012       | -0.003      | 0.003    | -0.138   | 0.086    |
|     | 0.082       | 0.079       | -0.084   | -0.002   | 0.036    |
| JPY | -0.369      | -0.359      | -0.028   | -0.035   | -0.295   |
|     | 0.145       | 0.127       | 0.055    | 0.059    | 0.200    |
|     | 0.199       | 0.203       | -0.101   | -0.102   | 0.144    |

NOTE: Table 6 provides sets of eigenvalues for the test of generalized method of second forecast error moment criterion. The first model dominates the second model if all the eigenvalues in a set are nonpositive and at least one is negative. GARCH-EW GP denotes the GARCH model versus the equally weighted genetic programming forecasts for the baseline case, as in Table 3. GARCH-MW GP denotes the GARCH model versus the median-weighted genetic programming forecasts for the baseline case, as in Table 3. RM-EW GP denotes the RiskMetrics™ model versus the equally weighted genetic programming forecasts for the baseline case. RM-MW GP denotes the RiskMetrics™ model versus the median-weighted genetic programming forecasts for the baseline case. GARCH-RM denotes the GARCH model versus RiskMetrics™ forecasts.

forecasts from two sources can be aggregated and compared is the generalized forecast error second moment (GFESM) method proposed by Clements and Hendry (1993). Unfortunately, this method has some drawbacks. For example, the GFESM can prefer

model 1 to model 2 based on forecasts from horizon 1 to horizon  $h$ , even if model 2's forecasts dominate at every forecast horizon up to  $h$ . To remedy the perceived weaknesses in the GFESM, Newbold, Harvey, and Leybourne (1999) proposed the generalized



mean-squared forecast error matrix (GMSFEM) criterion. This procedure prefers forecasting method 1 to method 2 if the magnitude of all linear combinations of forecast errors is at least as small under method 1 as method 2.

To explain the GMSFEM more fully, let us introduce some notation. The one-by-three vector of 1-, 5-, and 20-day GARCH forecast errors at time  $t$  is  $e_t^{GARCH} = \{e_{t,1}^{GARCH}, e_{t,5}^{GARCH}, e_{t,20}^{GARCH}\}$ , and the second moment matrix of these forecast errors is  $\Phi^{GARCH} = E(e_t^{GARCH} e_t^{GARCH'})$ . The RiskMetrics and genetic programming variables are defined analogously. The GMSFEM says that the GARCH model is preferred to the genetic programming forecasts if every linear combination of GARCH forecast errors is at least as small as every linear combination of genetic programming forecast errors. That is, if

$$(11) \quad d'(\Phi^{GARCH} - \Phi^{GP})d \leq 0 \quad \text{for all vectors } d \neq 0.^6$$

This condition is met if every eigenvalue of the matrix  $(\Phi^{GARCH} - \Phi^{GP})$  is nonpositive and at least one is negative. Clearly, the criterion prefers the genetic programming forecast if every eigenvalue is non-negative and at least one is positive.

Table 6 shows five sets of eigenvalues from the  $(\Phi^{GARCH} - \Phi^{GP})$  matrix, using both the equally weighted and median-weighted genetic program forecasts, for both exchange rates. It confirms the previous results. The only case in which there are all negative (or positive) eigenvalues is the comparison of the RiskMetrics forecast to the median-weighted genetic programming forecast. In that case, all the eigenvalues are negative, indicating that the RiskMetrics forecasts dominate the median-weighted genetic programming forecasts under the GMSFEM criterion. In every other set of eigenvalues there are both positive and negative values. Neither GARCH/RiskMetrics forecasts nor genetic programming forecasts dominate the other under the GMSFEM criterion.

## DISCUSSION AND CONCLUSION

We choose to use the problem of forecasting conditional volatility in the foreign exchange market to illustrate the strengths and weaknesses of genetic programming because it is a challenging problem with a well-accepted benchmark solution, the GARCH(1,1) model. The genetic program did reasonably well in forecasting out-of-sample volatility.

While the genetic programming rules did not usually match the GARCH(1,1) or RiskMetrics models' MSE or  $R^2$ , its performance on those measures was generally close. But the genetic program did consistently outperform the GARCH model on MAE and modal error bias at all horizons. The genetic programming solutions appeared to suffer from some in-sample overfitting, which was not mitigated, in this case, by an ad hoc penalty for rule complexity.

Our results suggest some interesting issues for further investigation. The superiority of the genetic program according to the MAE criterion is perhaps surprising given that we used MSE as the fitness criterion. This raises the possibility that further improvement in the forecasting performance of the genetic program relative to the GARCH model could be achieved by using MAE as the fitness criterion. Also, given that increasing the frequency of intra-day observations has been shown to improve the accuracy of forecasts based on the GARCH model (Andersen et al., 2001), it is important to discover whether the results of this investigation survive in that context.

## REFERENCES

- Andersen, Torben and Bollerslev, Tim. "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts." *International Economic Review*, 1998, 39(4), pp. 885-905.
- \_\_\_\_\_; \_\_\_\_\_. Diebold, Francis and Labys, Paul. "Modeling and Forecasting Realized Volatility." Working Paper W8160, National Bureau of Economic Research, 2001.
- Baillie, Richard and Bollerslev, Tim. "The Message in Daily Exchange Rates: A Conditional-Variance Tale." *Journal of Business and Economic Statistics*, July 1989, 7(3), pp. 297-305.
- \_\_\_\_\_. and \_\_\_\_\_. "Intra-Day and Inter-Market Volatility in Foreign Exchange Rates." *Review of Economic Studies*, May 1991, 58(3), pp. 565-85.
- Bollerslev, Tim. "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics*, April 1986, 31(3), pp. 307-27.
- Chen, Shu-Heng, ed. *Genetic Algorithms and Genetic*

<sup>6</sup> Note that the GMSFEM criterion implicitly favors models that do well in terms of MSE, rather than in terms of MAE.

- Programming in Computational Finance*. New York: Kluwer, 2002 (forthcoming).
- Clements, Michael P. and Hendry, David F. "On the Limitations of Comparing Mean Square Forecast Errors." *Journal of Forecasting*, December 1993, 12(8), pp. 617-76.
- Engle, Robert F. "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation." *Econometrica*, July 1982, 50(4), pp. 987-1007.
- Holland, John. *Adaptation in Natural and Artificial Systems*. Ann Arbor, MI: University of Michigan Press, 1975.
- J.P. Morgan/Reuters. *RiskMetrics™ Technical Document, Part II: Statistics of Financial Market Returns*. Fourth Edition. New York: 1996.
- Jorion, Philippe. "Predicting Volatility in the Foreign Exchange Market." *Journal of Finance*, June 1995, 50(2), pp. 507-28.
- Koza, John R. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge, MA: MIT Press, 1992.
- Neely, Christopher J. "Target Zones and Conditional Volatility: The Role of Realignments." *Journal of Empirical Finance*, April 1999, 6(2), pp. 177-92.
- \_\_\_\_\_. "Risk-Adjusted, Ex Ante, Optimal, Technical Trading Rules in Equity Markets." Working Paper WP 1999-015D, Federal Reserve Bank of St. Louis, 2001 (forthcoming in *International Review of Economics and Finance*).
- \_\_\_\_\_. and Weller, Paul A. "Technical Trading Rules in the European Monetary System." *Journal of International Money and Finance*, June 1999, 18(3), pp. 429-58.
- \_\_\_\_\_. and \_\_\_\_\_. "Technical Analysis and Central Bank Intervention." *Journal of International Money and Finance*, December 2001, 20(7), pp. 949-70.
- \_\_\_\_\_.; \_\_\_\_\_. and Dittmar, Robert. "Is Technical Analysis in the Foreign Exchange Market Profitable? A Genetic Programming Approach." *Journal of Financial and Quantitative Analysis*, December 1997, 32(4), pp. 405-26.
- Newbold, Paul; Harvey, David I. and Leybourne, Stephen L. "Ranking Competing Multi-Step Forecasts," in Robert F. Engle and Halbert White, eds., *Cointegration, Forecasting and Causality*. Chap. 4. Oxford: Oxford University Press, 1999.
- Newey, Whitney K. and West, Kenneth D. "A Simple Positive, Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*, May 1987, 55(3), pp. 703-08.
- \_\_\_\_\_. and \_\_\_\_\_. "Automatic Lag Selection in Covariance Matrix Estimation." *Review of Economic Studies*, October 1994, 61(4), pp. 631-53.
- West, Kenneth D. and Cho, Dongchul. "The Predictive Ability of Several Models of Exchange Rate Volatility." *Journal of Econometrics*, October 1995, 69(2), pp. 367-91.

All nonproprietary and nonconfidential data and programs for the articles written by Federal Reserve Bank of St. Louis staff and published in *Review* are available to our readers on our Web site: < [www.stls.frb.org/publications/review](http://www.stls.frb.org/publications/review) > . Also, you may request data and programs on either disk or hard copy: **Research Department**, Federal Reserve Bank of St. Louis, P.O. Box 442, St. Louis, MO 63166-0442. Please include the author, title, issue date, and page numbers with your request.

These data and programs are also available through **Inter-university Consortium for Political and Social Research (ICPSR)**. Member institutions may request data through the CDNet Order facility. Nonmembers may write to ICPSR, Institute for Social Research, P.O. Box 1248, Ann Arbor, MI 48106-1248; call 734-998-9900; or e-mail < [netmail@icpsr.umich.edu](mailto:netmail@icpsr.umich.edu) > .

General data can be obtained through **FRED (Federal Reserve Economic Data)**, a database providing U.S. economic and financial data and regional data for the Eighth Federal Reserve District. You may access FRED through our Web site: < [www.stls.frb.org/fred](http://www.stls.frb.org/fred) > .



■ [www.stls.frb.org/research](http://www.stls.frb.org/research) ■



**Federal Reserve Bank of St. Louis**

P.O. Box 442

St. Louis, MO 63166-0442

