Federal Reserve Bank of St. Louis

REVIEW

JANUARY/FEBRUARY 2006 VOLUME 88, NUMBER 1

The Fed's Monetary Policy Rule
William Poole

On the Size and Growth of Government
Thomas A. Garrett and Russell M. Rhine

The Evolution of the Subprime Mortgage Market
Souphala Chomsisengphet and Anthony Pennington-Cross

Are the Causes of Bank Distress Changing? Can Researchers Keep Up?
Thomas B. King, Daniel A. Nuxoll, and Timothy J. Yeager

Replicability, Real-Time Data, and the Science of Economic Research: FRED, ALFRED, and VDC
Richard G. Anderson

REVIEW

Director of Research
Robert H. Rasche

Deputy Director of Research
Cletus C. Coughlin

Review Editor
William T. Gavin

Research Economists
Richard G. Anderson
James B. Bullard
Riccardo DiCecio
Michael J. Dueker
Thomas A. Garrett
Massimo Guidolin
Hui Guo
Rubén Hernández-Murillo
Kevin L. Kliesen
Christopher J. Neely
Edward Nelson
Michael T. Owyang
Michael R. Pakko
Anthony N.M. Pennington-Cross
Jeremy M. Piger
Daniel L. Thornton
Howard J. Wall
Yi Wen
Christopher H. Wheeler
David C. Wheelock

Managing Editor
George E. Fortier

Assistant Editor
Lydia H. Johnson

Graphic Designer
Donna M. Stiller

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.

1 **The Fed's Monetary Policy Rule**
William Poole

13 **On the Size and Growth of Government**
Thomas A. Garrett and Russell M. Rhine

31 **The Evolution of the Subprime Mortgage Market**
Souphala Chomsisengphet and Anthony Pennington-Cross

57 **Are the Causes of Bank Distress Changing? Can Researchers Keep Up?**
Thomas B. King, Daniel A. Nuxoll, and Timothy J. Yeager

81 **Replicability, Real-Time Data, and the Science of Economic Research: FRED, ALFRED, and VDC**
Richard G. Anderson

Review is published six times per year by the Research Division of the Federal Reserve Bank of St. Louis and may be accessed through our web site: **research.stlouisfed.org/publications/review**. All non-proprietary and nonconfidential data and programs for the articles written by Federal Reserve Bank of St. Louis staff and published in *Review* also are available to our readers on this web site. These data and programs are also available through Inter-university Consortium for Political and Social Research (ICPSR) via their FTP site: www.icpsr.umich.edu/pra/index.html. Or contact the ICPSR at P.O. Box 1248, Ann Arbor, MI 48106-1248; 734-647-5000; **netmail@icpsr.umich.edu**.

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) 444-8809.

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: **research.stlouisfed.org/fred**.

Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are 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.

© 2006, Federal Reserve Bank of St. Louis.

ISSN 0014-9187

Sixth Annual Missouri Economics Conference

March 31–April 1, 2006
Memorial Union, University of Missouri–Columbia

The Missouri Economics Conference, jointly sponsored by the Department of Economics at the University of Missouri–Columbia and the Research Divisions of the Federal Reserve Banks of St. Louis and Kansas City, is held during the spring of each year at the University of Missouri–Columbia.

The purpose of the conference is to bring together leading research economists from the Federal Reserve, businesses, and academic institutions in the Eighth and Tenth Federal Reserve Districts. Conference sessions consist of presentations of three or four academic-level research papers, with discussants. Participants and presenters come from all over the world. A screening committee chooses the papers that are presented from among those submitted. Some papers, including those of the keynote speakers, may be invited by the committee.

Call for Papers: We are now accepting submissions of papers in all fields of economics to be presented at the conference. Completed manuscripts will be given priority, but abstracts will also be considered. To be considered for presentation at the conference, research papers or abstracts should be sent via email to moconf@kc.frb.org.

The deadline for submissions is February 15, 2006. Please include an abstract and JEL codes with all submissions. Additional information on the conference can be found at http://research.stlouisfed.org/conferences/moconf/6th_annual_info.html.

The Fed's Monetary Policy Rule

William Poole

This article was originally presented as a speech at the Cato Institute, Washington, D.C., October 14, 2005.

Federal Reserve Bank of St. Louis *Review*, January/February 2006, *88*(1), pp. 1-11.

In 1936, Henry Simons published a paper, "Rules Versus Authorities in Monetary Policy," that not only became a classic but also is still highly relevant to today's policy debates.[1] I rediscovered several important points in the paper while preparing this lecture.

In thinking about policy rules in recent years, I have tended to separate the political and economic cases for a rule. Simons argues for a much more integrated view of the issue:

> There are, of course, many special responsibilities which may wisely be delegated to administrative authorities with substantial discretionary power...The expedient must be invoked sparingly, however, if democratic institutions are to be preserved; and it is utterly inappropriate in the monetary field. An enterprise cannot function effectively in the face of extreme uncertainty as to the action of monetary authorities or, for that matter, as to monetary legislation.[2]

Thus, Simons argues that the rule of law that characterizes a democracy is also required to provide monetary policy predictability, which, in turn, is necessary for efficient operation of a market economy.

I've chosen a title designed to be provocative, for I suspect that few consider current Federal Reserve policy as characterized by a monetary rule. My logic is this: There is now a large body of evidence, which I'll review shortly, that Fed policy has been highly predictable over the past decade or so. If the market can predict the Fed's policy actions, then it must be the case that Fed policy follows a rule, or policy regularity, of some sort. My purpose is to explore the nature of that rule. Contrary to Simons's implication, the behavior of authorities can be predictable.

Before digging into specifics, consider what the "rules versus discretion" debate is about. Advocates of discretion, as I interpret them, are primarily arguing *against* a formal policy rule, and certainly against a legislated rule. They believe that policy will be more effective if characterized by "discretion."

Discretion surely cannot mean that policy is haphazard, capricious, random, or unpredictable. Advocates of discretion agree with Simons that "many special responsibilities...may wisely be delegated to administrative authorities with substantial discretionary power." However, they do not agree with Simons that discretion "is utterly inappropriate in the monetary field."

Interestingly, Simons argued that a fixed money stock would be the best rule, but only if

[1] Simons (1936).

[2] Simons (1936, pp. 1-2).

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

© 2006, The Federal Reserve Bank of St. Louis. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

substantial institutional reforms were in place in financial markets, such as 100 percent reserve requirements against bank deposits. Given the institutional structure, Simons argued for a rule focused on price-level stabilization, because "no monetary system can function effectively or survive politically in the face of extreme alternations of hoarding and dishoarding."[3] That is, Simons believed that large variations in the velocity of money would make a fixed money stock rule work poorly.

Despite the nature of his argument for a price-level stabilization rule, elsewhere in the same paper Simons argued that, "[o]nce well established and generally accepted as the basis of anticipations, any one of many different rules (or sets of rules) would probably serve about as well as another."[4] I think his first argument was correct—that different rules, even once fully understood, would have different operating properties in the economy, and that a choice among various possible rules should depend on which rule yields better economic outcomes.

My view has evolved over time to this general position: Monetary economists have not yet developed a formal rule that is likely to have better operating properties than the Fed's current practice. It is highly desirable that policy practice be formalized to the maximum possible extent. Or, more precisely, monetary economists should embark on a program of continuous improvement and enhanced precision of the Fed's monetary rule. It is possible to say a lot about the systematic characteristics of current Fed practice, even though I do not know how to write down the current practice in an equation. It is in this sense that I'll be describing the Fed's policy rule. And given that, as far as I know, there is no other effort to state in one place the main characteristics of the Fed's policy rule; I'm sure that subsequent work will refine and correct the way I characterize the rule. Thus, I am redefining the "rule" to fit current practice, which has yielded an environment in which policy actions are highly, though not perfectly, predictable in the markets.

[3] Simons (1936, p. 5).

[4] Simons (1936, p. 29).

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

POLICY PREDICTABILITY—A SUMMARY OF FINDINGS

I've discussed the predictability of Fed policy decisions on a number of occasions, most recently in a speech on October 4, 2005, entitled, "How Predictable Is Fed Policy?" Let me summarize the main findings.

Over the past decade, the Federal Open Market Committee (FOMC) has undertaken a number of steps toward greater transparency that have greatly improved the ability of markets to predict future policy actions. Among these steps are the announcement of policy actions at the conclusion of each FOMC meeting; the restriction of policy actions to regularly scheduled FOMC meetings, except under extraordinary conditions; the announcement of a specific numeric target for the federal funds rate in the post-FOMC meeting press releases and in the Directive to the Manager of the open market desk at the Federal Reserve Bank of New York; the inclusion of the individual votes at the FOMC meeting in the press release; and the expedited release of the minutes of the FOMC meetings. In addition, since 1989 all FOMC policy actions to change the target for the funds rate have been in multiples of 25 basis points. With the exception of one change of 75 basis points, all the changes have been either 25 or 50 basis points.

As I have noted previously, I believe that the evidence supports the conclusion that these steps toward increased transparency have brought the markets into much better "synch" with FOMC thinking about appropriate policy actions. My metric for judging how well markets have anticipated FOMC policy actions is the reaction of the yield on the 1-month-ahead federal funds futures contract between the close of business on the day

before the FOMC meets and the close of business on the day of the meeting. Our research suggests that changes of less than 5 basis points are "noise." Larger changes reflect surprises to market expectations.

Since the middle of 1995, when the FOMC has undertaken policy actions at regularly scheduled meetings, the markets have been surprised only 12 times, as measured by a change of 5 basis points or more in the 1-month-ahead federal funds futures contract. Since the middle of 2003, when the FOMC introduced "forward looking" language into the press release, there have been no surprises. In contrast, on all four occasions when the FOMC instituted intermeeting policy actions, the markets were taken by surprise.

On the other side of the coin, FOMC decisions to leave the funds rate target unchanged have also become largely predictable. Since the middle of 1995 there have been only two occasions when the markets expected a change in the funds rate target and the FOMC left it unchanged.

These findings open this question: What are the circumstances under which market expectations of FOMC actions are adjusted, so that, by the time the FOMC meets, the outcomes are generally correctly foreseen? There is a substantial literature documenting interest rate responses to arriving information. Given that the federal funds futures market predicts FOMC policy decisions quite accurately, that literature provides insight into how the FOMC responds to new information. What I'll do now is to step back from that level of detail to discuss policy regularities at a high level, starting with policy goals.

POLICY GOALS

On many occasions, dating back to Paul Volcker's confirmation hearing in 1979, Fed officials have stated that the goal of low and stable inflation is there because it maximizes the economy's sustainable rate of economic growth.[5]

The dual mandate in the Federal Reserve Act, as amended, and in other legislation provides for goals of maximum purchasing power, usually interpreted as price stability, and maximum employment. There are two aspects to achieving the employment goal. First, achieving low and stable inflation maximizes the economy's growth potential and, probably, maximizes the sustainable level of employment. Second, the Fed can enhance employment *stability* through timely adjustments in its policy stance. A subsidiary goal of general financial stability is closely related to both inflation and employment goals.

The Fed has gravitated to a specification of the inflation goal stated in terms of the core personal consumption expenditures (PCE) index. At the FOMC meeting of December 21, 1999, Chairman Greenspan provided a clear statement of the case for focusing on the PCE price index rather than on the consumer price index (CPI).

> The reason the PCE deflator is a better indicator in my view is that it incorporates a far more accurate estimate of the weight of housing in total consumer prices than the CPI. The latter is based upon a survey of consumer expenditures, which as we all know very dramatically underestimates the consumption of alcohol and tobacco, just to name a couple of its components. It also depends on people's recollections of what they spent, and we have much harder evidence of that in the retail sales data, which is where the PCE deflator comes from.[6]

There is evidence that the goal is effectively a 1 to 2 percent annual rate of change, averaged over a "reasonable" period whose precise definition depends on context. Evidence supporting

[5] See, for example Committee on Banking Housing and Urban Affairs, United States Senate, Ninety-sixth Congress, first session, *Hearings on the Nomination of Paul A. Volcker to be Chairman, Board of Governors of the Federal Reserve System*, July 30, 1979, p. 20; Committee on Banking, Housing and Urban Affairs, United States Senate, Ninety-eighth Congress, first session, *The Renomination of Paul A. Volcker to be Chairman, Board of Governors of the Federal Reserve System for a term of 4 years ending August 6, 1987*, July 14, 1983, p. 15; Committee on Banking, Housing and Urban Affairs, United States Senate, One Hundredth Congress, first session, *The Nomination of Alan Greenspan of New York, to be a member of the Board of Governors of the Federal Reserve System for the unexpired term of 14 years from February 1, 1978, vice Paul A. Volcker, resigned; and, to be Chairman, Board of Governors of the Federal Reserve System for a term of 4 years, vice Paul A. Volcker, resigned*, July 21, 1987, p. 29; Committee on Banking, Finance and Urban Affairs, United States House of Representatives, *Testimony of Alan Greenspan, February 23, 1988*, reprinted in the *Federal Reserve Bulletin*, April 1988, p. 227.

[6] FOMC *Transcript*, December 21, 1999, p. 49.

this view of the inflation goal appears in the minutes of the FOMC meetings of May 6, 2003, and August 9, 2005.[7]

I regard inflation stability as the primary goal not because it is more important in a welfare sense than maximum employment but because achieving low and stable inflation is prerequisite to achieving employment goals. Inflation stability also enhances, but does not guarantee, financial stability.

I take note, but will not further discuss here, the ongoing debate as to whether the inflation goal should be formalized as a particular numerical goal or range.

CHARACTERISTICS OF THE FED POLICY RULE

The Fed policy rule has a number of elements that can be identified and, in many cases, quantified. I'll now discuss the most important of these.

The Taylor Rule

Statements and testimony of Chairmen Volcker and Greenspan and other FOMC participants, supplemented by the transcripts and minutes of FOMC discussions over the past 25 years, clearly indicate that the long-run objective of Federal Reserve monetary policy is to maintain price stability, usually phrased as "low and stable inflation." In the short run, policy actions are undertaken with the intention of alleviating or moderating cyclical fluctuations, as Chairman Greenspan has noted:

> [M]onetary policy does have a role to play over time in guiding aggregate demand into line with the economy's potential to produce. This may involve providing a counterweight to major, sustained cyclical tendencies in private spending, though we can not be overconfident in our ability to identify such tendencies and to determine exactly the appropriate policy response.[8]

Over 10 years ago, John Taylor (1993) noted that these characteristics of FOMC policy actions could be summarized in a simple expression:

$$\begin{aligned} i &= p + .5\left(p - p^*\right) + .5y + r^* \\ &= 1.5\left(p - p^*\right) + .5y + \left(r^* + p^*\right), \end{aligned}$$

where i is the nominal federal funds rate, p is the inflation rate, p^* is the target inflation rate, y is the percentage deviation of real gross domestic product (GDP) from a target, and r^* is an estimate of the "equilibrium" real federal funds rate.

Under this characterization of the systematic or "rule like" character of FOMC policy actions, the funds rate is raised (lowered) when actual inflation exceeds (falls short of) the long-run inflation objective and is raised (lowered) when output exceeds (falls short of) a target level. In Taylor's example, the target for GDP was constructed from a 2.2 percent per annum trend of real GDP starting with the first quarter of 1984. In subsequent analyses this target has been interpreted as a measure of "potential GDP." When inflation and real GDP are on-target, then the policy setting of the real funds rate is the estimated equilibrium value of the real rate. This formulation of an interest rate monetary policy rule satisfies McCallum's properties for a rule that provides a "nominal anchor" to the economy.[9] Taylor showed that his equation closely tracked the actual federal funds rate from 1987 through 1992 except around the stock market crash in October 1987.

For such a rule to be operational, data on the inflation rate and GDP must be known to the FOMC. In practice, the equation can be specified with lagged data on inflation and GDP. More generally the equation can be written as follows:

$$i_t = \mathrm{a}\left(\bar{p}_{t-1} - p^*\right) + 100\mathrm{b}\ln\left(y_{t-1} / y_{t-1}^P\right) + \left(r^* + p^*\right),$$

where $\bar{p}_{t-1}$ is the previous quarter's PCE inflation rate measured on a year-over-year basis, y_{t-1} is the log of the previous quarter's level of real GDP, and y_{t-1}^P is the log of potential real GDP as estimated by the Congressional Budget Office. To

[7] FOMC *Minutes*, May 6, 2003, www.federalreserve.gov/fomc/minutes/20030506.htm; FOMC *Minutes*, August 9, 2005, www.federalreserve.gov/fomc/minutes/20050809.htm.

[8] Committee on Banking, Finance and Urban Affairs, United States House of Representatives, testimony of Alan Greenspan, July 13, 1988. Reprinted in *Federal Reserve Bulletin*, September 1988, p. 611.

[9] McCallum (1981).

Figure 1

Greenspan Years: Funds Rate and Taylor Rules ($p^* = 1.5$, $r^* = 2.0$) a = 1.5, b = 0.5

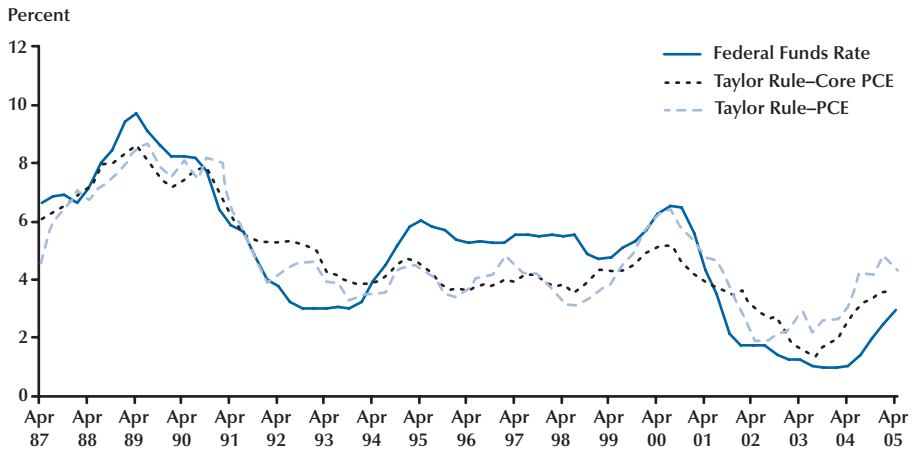


Figure 2

Greenspan Years: Funds Rate and Taylor Rules ($p^* = 1.5$, $r^* = 2.3$) a = 1.5, b = 0.8

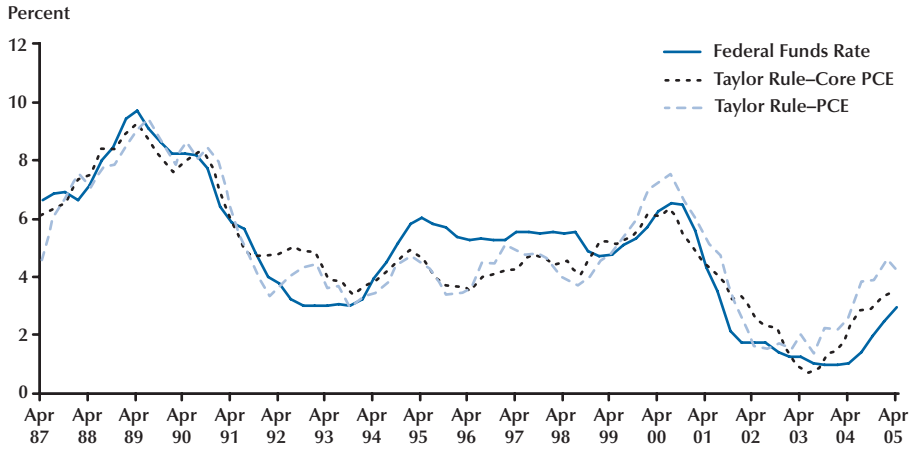
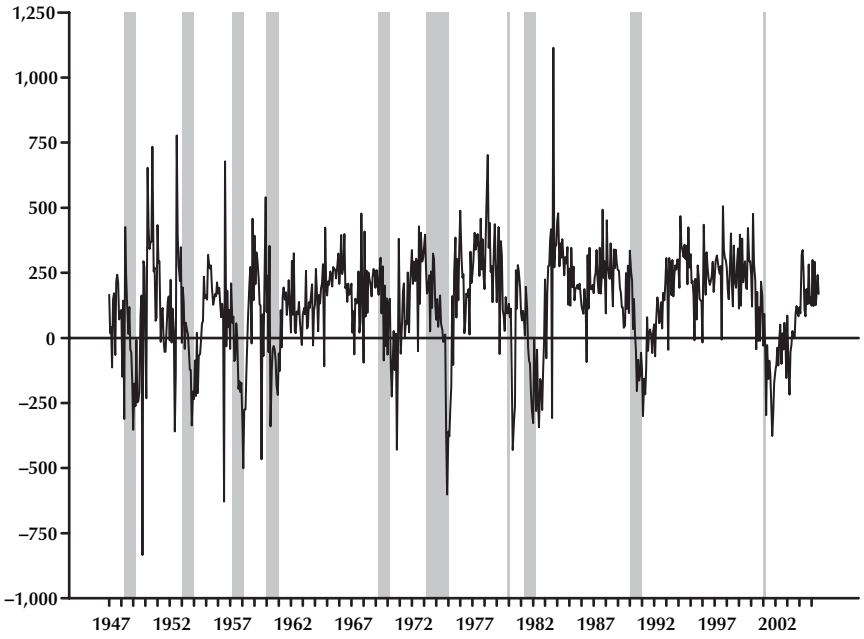


Figure 3

Monthly Changes in Nonfarm Payroll Employment: January 1947–August 2005



NOTE: Shaded bars indicate recessions.

ensure a "nominal anchor" for the economy, the coefficient a must be greater than 1.0.

Figure 1 shows the equation with the Taylor coefficients (a = 1.5, b = 0.5), an assumed equilibrium real rate of interest of 2.0, and an assumed inflation target of 1.5 percent. The solid blue line shows the actual federal funds rate and the dashed lines the two Taylor rule funds rates. The small-dash black line is the rule constructed with the core PCE inflation rate; the long-dash light blue line with the PCE inflation rate.[10] The average differences between the two "Taylor rules" and the actual funds rate over the entire period are 15 and 7 basis points, respectively. However, the volatility of each of the two Taylor rules is much less than that of the actual funds rate.

Figure 2 shows the comparison of the two Taylor rules with a larger coefficient on the output gap (b = 0.8) and a slightly higher assumed equilibrium real rate (r^* = 2.3). With these assumptions the average differences between the two equations and the funds rate over the entire period are 2 and –3 basis points, respectively, and the volatility of the two equations better approximates the volatility of the actual funds rate.

My purpose here is not to try to find the equation that reveals the policy rule of the Greenspan Fed; as I stated earlier, I do not know how to write down the current practice in an equation and the FOMC certainly does not view itself as implementing an equation. Rather, the illustrations should be viewed as evidence in support of the proposition that the general contours of FOMC policy actions are broadly predictable.

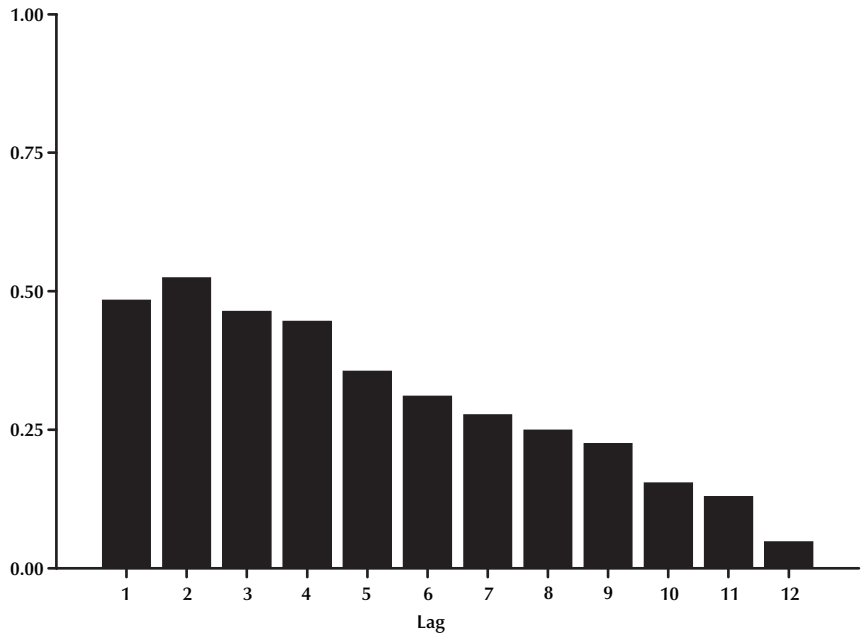
Policy Asymmetry

Under most circumstances the direction of FOMC policy actions is "biased" in a sense I'll explain. Policy bias exists because turning points

[10] Taylor originally specified his equation in terms of CPI inflation. Since the FOMC has stated a preference for PCE measures of inflation, those measures are used here.

Figure 4

Autocorrelations of Monthly Payroll Employment Changes: January 1947–August 2005



in economic activity—peaks and troughs of business cycles—are infrequent. Changes in economic activity as measured by output and employment are highly persistent. This persistence can be seen in Figure 3, which shows month-to-month changes in nonfarm payroll employment from January 1947 through August 2005. During expansions, employment changes are consistently positive; during recessions consistently negative. Changes opposite to the cyclical direction are rare and generally the consequence of identifiable transitory shocks such as those from strikes and weather disturbances. This pattern of business cycles generates strong autocorrelations in the month-to-month changes in payroll employment, as shown in Figure 4.[11]

Given such persistence, once it becomes apparent that a cyclic peak likely has occurred, the issue is never whether the Fed will raise the target funds rate but whether and how much the Fed will cut the target rate. Similarly, once it is apparent that an expansion is underway, the question is not whether the Fed will cut the target rate, but the extent and timing of increases.

Data Anomalies

Fed policy responds to incoming information, as it should. Sometimes data ought to be discounted because of anomalous behavior. For example, the FOMC has indicated that it monitors inflation developments as measured by the core rather than the total PCE inflation rate. This approach is appropriate because the impacts on inflation of food and energy prices are largely transitory; the difference between the inflation rate as measured by the total PCE index and as measured by the core PCE index fluctuates around zero.

Another example was the increase in tobacco prices in late 1998. Tobacco prices had a transitory impact on measured inflation, for both total and

[11] An estimate ARIMA model for monthly changes in nonfarm payroll employment over the period since 1947 indicates that

$$\Delta Payroll_Emp_t - 0.96\Delta Payroll_Emp_{t-1} = \varepsilon_t - 0.64\varepsilon_{t-1}.$$

core indices, during December 1998 and January 1999, but produced no lasting effect on trend inflation.[12] Similarly, information about real activity sometimes arrives that indicates transitory shocks to aggregate output and employment. An example of such a transitory shock is the strike against General Motors in June and July 1998.[13] Similarly, the September 2005 employment report reflects the impact of Hurricane Katrina.

Transitory and anomalous shocks to the data are ordinarily rather easy to identify. Both Fed and market economists develop estimates of these aberrations in the data shortly after they occur. The principle of looking through aberrations is easy to state but probably impossible to formalize with any precision. We know these shocks when we see them, but could never construct a completely comprehensive list of such shocks ex ante.

Policymakers piece together a picture of the economy from a variety of data, including anecdotal observations. When the various observations fit together to provide a coherent picture, the Fed can adjust the intended rate with some confidence. The market generally understands this process, as it draws similar conclusions from the same data.

Crisis Management

The above rules are suspended when necessary to respond to a financial crisis. The major examples of the Greenspan era are the stock market crash of 1987, the combination of financial market events in late summer and early fall 1998 that culminated in the near failure of Long Term Capital Management, crisis avoidance coming up to the century date change at the end of 1999, and the 9/11 terrorist attacks. In each case, the nature of the response was tailored to circumstances unique to each event. In all cases, crisis responses were helpful because markets had confidence in the Federal Reserve, including confidence that extra provision of liquidity would be withdrawn before risking an inflation problem. In the absence of such confidence, the Fed's ability to respond would be severely curtailed.

The history of Fed crisis management since World War II is generally a happy one. Before the Greenspan era, significant events include the failure of Penn Central in 1970 and the near failure of Continental Illinois in 1984. Perhaps just as important, the Fed has not responded to certain events where it was called to do so. Examples would include the New York City financial crisis in 1975 and failure of Drexel Burnham Lambert in 1990.[14]

Other Regularities in Policy Stance

Since August 1989, the FOMC has adjusted the intended federal rate in multiples of 25 basis points only. After February 1994, when the FOMC first began to announce its policy decision at the conclusion of its meeting, with few exceptions all adjustments have been made at regularly scheduled meetings. These exceptions were April 18, 1994, September 29, 1998, January 3, 2001, April 18, 2001, and September 17, 2001.

In general, the Fed can use intermeeting adjustments to respond to special circumstances, such as the rate cut on September 17, 2001, in response to 9/11, or to provide information to the market about a major change in policy thinking or direction, such as the rate cut on April 18, 2001. My own preference is to confine intermeeting adjustments to circumstances in which delaying action to the next meeting would have significant costs. In general, if the market believes that changed circumstances will lead to a changed decision at the next regularly scheduled meeting, then little is gained by acting between meetings.

[12] From the December 1998 CPI release in January 1999: "Three-fourths of the December rise in the index for all items less food and energy was accounted for by a 18.8 percent rise in the index for cigarettes, reflecting the pass-through to retail of the 45-cents-a-pack wholesale price increase announced by major tobacco companies in late November."

[13] From the July 16, 1998, Federal Reserve Statistical Release G.17 Industrial Production and Capacity Utilization press release: "Industrial production declined 0.6 percent in June after a revised gain of 0.3 percent in May. Ongoing strikes, which have curtailed the output of motor vehicles and parts, accounted for the decrease in industrial production." From the Employment Situation: July 1998, released August 7, 1998: "Nonfarm payroll employment edged up by 66,000 to 125.8 million, as growth was curtailed by strikes and plant shutdowns in automobile-related manufacturing."

[14] Drexel Burnham Lambert was first investigated by the Securities and Exchange Commission in late 1987 and charged with securities fraud in June 1988. A settlement was reached in December 1988, but the firm declared bankruptcy in February 1990.

By reserving almost all actions to regularly scheduled meetings, intermeeting actions have special force, which can be valuable in meeting financial crises.

ISSUES TO BE RESOLVED

The rules-versus-discretion debate historically was framed in terms of policy actions. The focus on policy actions was natural because, historically, central bankers were reticent to comment on the rationale for their policy actions and only rarely provided hints about the future course of policy actions. Over the past 15 years, as central bankers, including the FOMC, have striven for greater transparency in monetary policy, communication in the form of policy statements has moved to center stage. It is clear that policy statements are just as important as policy actions, at least in the short run, because significant market effects can flow from these statements. We need to face a new question: Can policy statements become predictable? I think the answer in principle is largely in the affirmative, although evidence on the issue is scanty and I do not believe that policy statements are currently highly predictable.

Two significant elements in FOMC policy statements are the "balance of risks" assessment introduced in January 2000 and the "forward looking" language introduced in August 2003. The balance-of-risks assessment was introduced to replace the long-standing "bias" statement in the Directive to the Open Market Desk. Historically, the bias statement had referred to the intermeeting period and was not even made public in timely fashion until May 1999. With the regularization of FOMC policy actions on scheduled meeting dates, and issuance of a statement following every meeting starting with May 1999 to indicate whether or not the funds rate target was changed, a consensus emerged among FOMC participants that the bias formulation did not provide a clear public communication. The balance-of-risks statement attempted to provide insight into the major policy concerns of FOMC members over the "foreseeable future."

Initially, the Committee sought to summarize the risks for policy in the foreseeable future in a single assessment covering the prospects for both real economic activity and inflation. In June 2003, the assessment of the risk for sustainable growth was unbundled from the risk for inflation, allowing the Committee to express concerns in different directions about the two risks. Until April 2005, the balance-of-risks was an unconditional statement; since then, the assessment has been conditioned upon "appropriate monetary policy action."

Over the 49 FOMC meetings since February 2000, there have been 10 substantive changes in the wording of the balance-of-risks statement.[15] One of these changes was a decision not to make a balance-of-risks assessment on March 18, 2003, in light of the uncertainty associated with the Iraq war. In the remaining 10 formulations of the statement, 5 assessed the risks as roughly balanced (or balanced conditional on appropriate policy), 3 indicated concern about economic weakness, 1 indicated concern about heightened inflation pressures, and 1 indicated a concern about the risk that inflation might become "undesirably low."

The switch in language on December 19, 2000, from a concern about heightened inflation pressures to economic weakness, was followed by a reduction in the federal funds target by 50 basis points at an unscheduled FOMC meeting on January 3, 2001. On August 13, 2002, the risk assessment was changed from balanced to weighted toward economic weakness, but the FOMC took no policy actions until it reduced the target for the funds rate by 50 basis points at its scheduled meeting on November 6, 2002—the second FOMC meeting after the change in language. The risk assessment was changed from balanced to weighted toward weakness at the May 6, 2003, scheduled FOMC meeting, and the federal funds rate target was reduced by 25 basis points at the subsequent FOMC meeting on June 25, 2003. Prior to August 2003, no policy actions were undertaken at a given FOMC meeting or its subsequent meeting when the risk assessment was balanced.

Beginning in August 2003, the FOMC added "forward looking" language to the press statement. Initially, the language indicated that "policy

[15] These changes occurred on December 19, 2000; March 19, 2002; August 13, 2002; November 6, 2002; March 18, 2003; May 6, 2003; June 25, 2003; December 9, 2003; May 4, 2004; and March 22, 2005.

accommodation can be maintained for a considerable period." In January 2004, the Committee changed the language to indicate that it could be "patient in removing its policy accommodation." The FOMC did not change the target federal funds rate while these statements were in effect. In May 2004, the Committee indicated that it "believes that policy accommodation can be removed at a pace that is likely to be measured." At its following meeting, the FOMC raised the federal funds rate target by 25 basis points. The Committee then raised the target rate by 25 basis points at all its subsequent meetings up to the time this speech was written. The most recent such meeting was September 20, 2005.

At a minimum, the FOMC can and should aspire to policy statements that are clear and do not themselves create uncertainty and ambiguity. The record since 2000 suggests that the balance-of-risks statement and more recently the forward-looking language included in the press releases have provided consistent signals about the direction of future policy actions.

In interpreting the FOMC's policy statements, it is important that each statement be read against previous ones. *Changes* in the wording are critical to understanding the perspective of the FOMC members about future policy actions.

RULE ENFORCEMENT

Obviously, there exists no legal enforcement mechanism of the current rule. Nevertheless, there are certainly incentives for the Fed Chairman to follow the rule, or work to define improvements.

The most powerful incentives arise from market reactions to Fed policy actions. The federal funds futures market provides a sensitive measure of near-term market expectations and the eurodollar futures market a sensitive measure of longer-term funds rate expectations. The spread between conventional and indexed Treasury securities provides information on inflation expectations or, more accurately, inflation compensation. Options in these markets provide information on the diffusion of investor expectations. Volatility of market rates and accompanying market commentary provide quick feedback as to market reactions to Fed policy actions and policy statements. It is not in the Fed's interest to confuse or whipsaw markets, and for this reason market reactions provide an incentive for the Fed to conduct policy in a predictable fashion that at the same time achieves policy goals. Policy actions should be unpredictable only in response to events that are themselves unpredictable. The response function itself should be as predictable as possible. That is, given the arrival of new information, the goal is that the market should be able to predict the policy action in response to that information.

Although market responses are the most important disciplining force, FOMC members other than the Chairman also provide input, including input through dissents when a member feels strongly that a different policy decision would be better. Reserve Bank directors weigh in through discount rate decisions. Since 1994, except in unusual circumstances, the FOMC has not changed the intended federal funds rate unless several Reserve Banks have proposed corresponding discount rate changes.[16]

Finally, the general role of public discussion, including the highly visible congressional hearings, bears on the process. Skillful public officials do not want to be forced into a defensive posture when confronting questions in hearings and in Q&A sessions following speeches. I'll leave it to political scientists to study the matter in detail, but will guess that public opinion plays a more important role than formal legal processes in enforcing many legislated and common law rules. If so, then public opinion can play an important role in enforcing extra-legal rules as well.

A SUMMING UP

Federal Reserve policy has become highly predictable in recent years; in the future this predictability will, I am sure, be seen as one of the

[16] During this period there were 7 occasions when the target funds rate was changed without an accompanying action by the Board of Governors to change the discount rate. Of the remaining 36 changes in the intended funds rate, 33 were accompanied by changes in the discount rate at four or more Federal Reserve Banks. On 24 of these occasions, the discount rate was changed at a majority of the Federal Reserve Banks.

hallmarks of the Greenspan era. Little has been institutionalized, and for this reason the current Federal Reserve policy rule must be regarded as somewhat fragile. Still, future Chairmen will want to extend Alan Greenspan's successful era and therefore it will be in their interest to commit to pursue policy regularities that work well.

I do not claim to have accurately identified all aspects of the Fed's current policy rule. I am tempted to call it the "Greenspan policy rule," for Alan Greenspan has surely had far more to do with its construction than anyone else. Nevertheless, I believe that most elements of the rule have become part of a general Fed culture, understood at least roughly by other FOMC members and by staff. While it is appropriate to refer to the "Greenspan rule," I believe that FOMC debates and staff contributions have had a lot to do with development of the rule. For this reason, I believe that we should be hopeful that consistent and predictable Fed policy is likely to continue into the future.

REFERENCES

Simons, Henry C. "Rules Versus Authorities in Monetary Policy." *Journal of Political Economy*, February 1936, *44*(1), pp. 1-30.

McCallum, Bennett T. "Price Level Indeterminacy with an Interest Rate Policy Rule and Rational Expectations." *Journal of Monetary Economics*, November 1981, *8*(3), pp. 319-29.

Poole, William. "How Predictable Is Fed Policy?" University of Washington, Seattle Oct. 4, 2005; www.stlouisfed.org/news/speeches/2005/10_04_05.htm; Federal Reserve Bank of St. Louis *Review*, November/December 2005, *87*(6), pp. 659-68.

Taylor, John B. "Discretion versus Policy Rules in Practice." *Carnegie-Rochester Conference Series on Public Policy*. Amsterdam: North-Holland, 1993, *39*(0), pp. 195-214.

On the Size and Growth of Government

Thomas A. Garrett and Russell M. Rhine

The size of the U.S. federal government, as well as state and local governments, increased dramatically during the 20th century. This paper reviews several theories of government size and growth that are dominant in the public choice and political science literature. The theories are divided into two categories: citizen-over-state theories and state-over-citizen theories. The relationship between the 16th Amendment to the U.S. Constitution and the timing of government growth is also presented. It is likely that portions of each theory can explain government size and growth, but the challenge facing economists is to develop a single unifying theory of government growth.

Federal Reserve Bank of St. Louis *Review*, January/February 2006, *88*(1), pp. 13-30.

Economists have long been divided on the role of government in a society.[1] John Maynard Keynes and John Kenneth Galbraith have argued that an economy needs to be continually fine-tuned by an activist government to operate efficiently[2]: Thus, as an economy grows, a growing government is also necessary to correct private-sector inefficiencies. This school of thought grew primarily out of the Great Depression, when markets seemed to fail and government intervention was viewed as the means to restore economic stability. Other 20th century economists, such as Frederick von Hayek and Milton Friedman, have argued that an activist government is the cause of economic instability and inefficiencies in the private sector.[3] Government should exist to ensure that a private market operates efficiently; it should not act to replace the market mechanism.

Various data clearly suggest that the size of the federal government in the United States has grown dramatically during the 20th century.[4] One measure of government growth is federal expenditures per capita. The history of real (2000 dollars) federal government expenditures per capita from 1792 to 2004 is shown in Figure 1. This growth did not occur gradually, however. In the early years of the United States, the federal government spent about $30 per person annually. By the 1910s, government expenditures per capita were about $129, or slightly more than four times the 1792 level. In 2004, the federal government spent $7,100 per capita, nearly 55 times more than was spent per capita in the 1910s. Spending growth did slow in the mid-1980s and actually decreased

[1] The evolution of this debate is presented in Yergin and Stanislaw (2002).

[2] John Maynard Keynes's book, *The General Theory of Employment, Interest, and Money*, is one of the most influential economic books of the 20th century. Keynes states the need for substantial increases in government spending during times of economic contractions. Similarly, John Kenneth Galbraith argued for an expansionary fiscal policy to increase economic activity and employment.

[3] Of the many publications of both these Nobel Prize–winning economists, the most influential are Hayek's *The Road to Serfdom* and Milton Friedman and Anna Schwartz's *A Monetary History of the United States 1867-1960*.

[4] All data on federal, state and local government expenditures are from the Office of Management and Budget (www.whitehouse.gov/omb) and the U.S. Census Bureau.

Thomas A. Garrett is a research officer at the Federal Reserve Bank of St. Louis, and Russell M. Rhine is an assistant professor at St. Mary's College of Maryland. Lesli Ott provided research assistance.

© 2006, The Federal Reserve Bank of St. Louis. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

Table 1

Cabinet Departments

Department	Year established
State	1789
Treasury	1789
Justice	1789
Defense*	1789
Interior	1849
Agriculture	1889
Commerce	1913
Labor	1913
Health and Human Services	1953
Housing and Urban Development	1965
Transportation	1966
Energy	1967
Education	1979
Veterans Affairs	1987
Environmental Protection Agency†	1990
Homeland Security	2002

NOTE: *The date refers to the Department of War; the Department of Defense was officially created in 1949. The Department of War (1789), the Department of the Navy (1798), the Department of the Army (1947) and the Department of the Air Force (1947) were all reorganized under the Department of Defense in 1949. See www.dod.gov.
†Cabinet-level rank under George W. Bush; see www.whitehouse.gov/government/cabinet.html.

SOURCE: Cabinet department websites.

in the mid-1990s. By the year 2000, however, per capita spending increased once again.

It is clear from Figure 1 that spending on national defense can have a substantial impact on the level of government spending. Figure 2 is a graph of total per capita expenditures with and without defense spending over the period 1947-2004. It is evident that the long-term growth in total per capita government spending is not solely a function of national defense.

Federal spending has also increased relative to gross domestic product (GDP) throughout much of this country's history, as seen in Figure 3. Expanded government during World War II is clearly evident in Figure 3, as is the slowdown in government growth during the 1980s and 1990s. Figure 1 shows that the federal government has historically spent more per person each year, but Figure 3 suggests that this growth in spending has been less than the growth in GDP at the end of the 20th century.

An examination of the components of federal government spending provides insight into which areas the government has increased activity. Figure 4 plots several components of federal government spending per capita from 1947 to 2004. Although total per capita spending increased following World War II, several components of federal government expenditures stayed relatively constant or even decreased slightly over the next 50 years: physical resources (e.g., transportation, energy), national defense, and "other functions" (e.g., agriculture, general government, international affairs). In fact, much of the reduction in federal expenditures per person occurring in the mid- to late 1990s can be attributed to a reduction in national defense spending. However, spending on national-debt net interest payments and human resources grew substantially over the same period. The dramatic increase in human resources that occurred reflects the growth in Social Security payments and the inception of entitlement programs such as Medicare (in 1965).

Another measure of the size of the federal government is the number of cabinet departments. Eight cabinet departments were created from 1788 to 1952. Since 1953, there have been an additional eight cabinet departments established. Table 1 provides a list of all executive cabinet departments and the dates they were each established. One can infer from Table 1 and Figure 1 that the increase in per capita expenditures during the 20th century was due to an increase in the physical size of government as well as an increase in spending by existing government agencies.

In addition to the increase in federal government expenditures, state and local government expenditures per capita have also increased since World War II, as seen in Figure 5. Inflation-adjusted expenditures per person were about $759 in 1948, compared with over $4,300 per person in 2004. The average annual growth rate in real per capita state and local government expenditures was 3.2 percent, compared with an average annual growth

Figure 1

Real Per Capita Federal Expenditures: 1792-2004

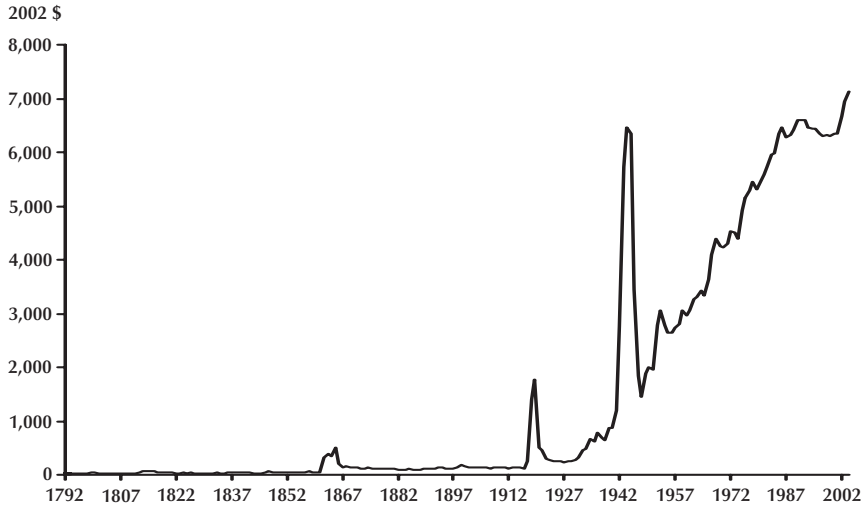


Figure 2

Real Per Capita Federal Expenditures: 1947-2004

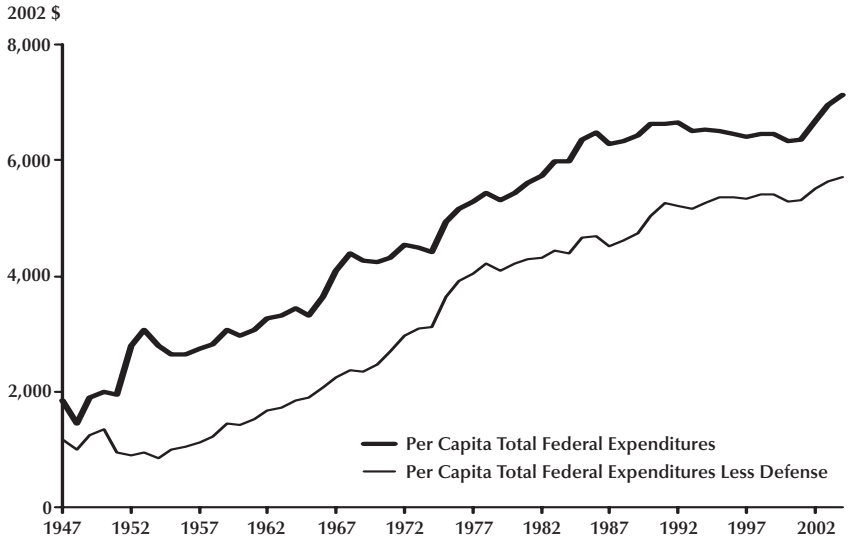


Figure 3

Total Federal Expenditures as a Percent of GDP: 1930-2004

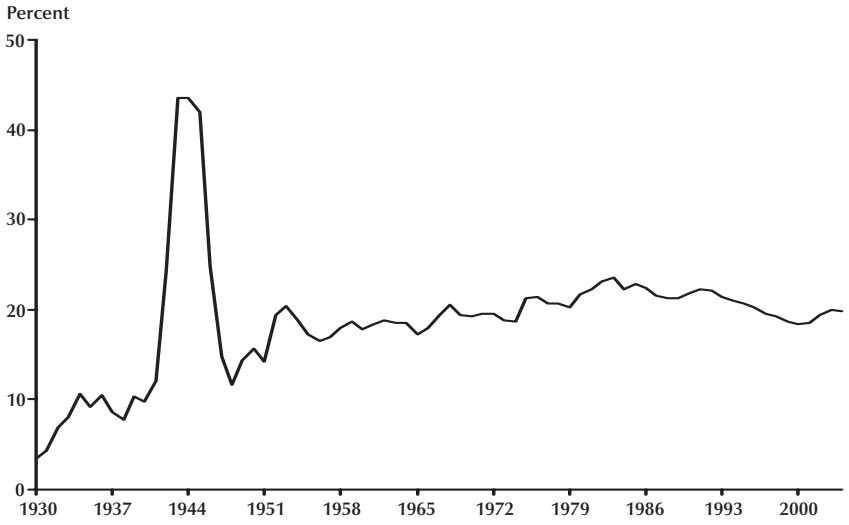


Figure 4

Real Per Capital Federal Expenditures by Component: 1947-2004

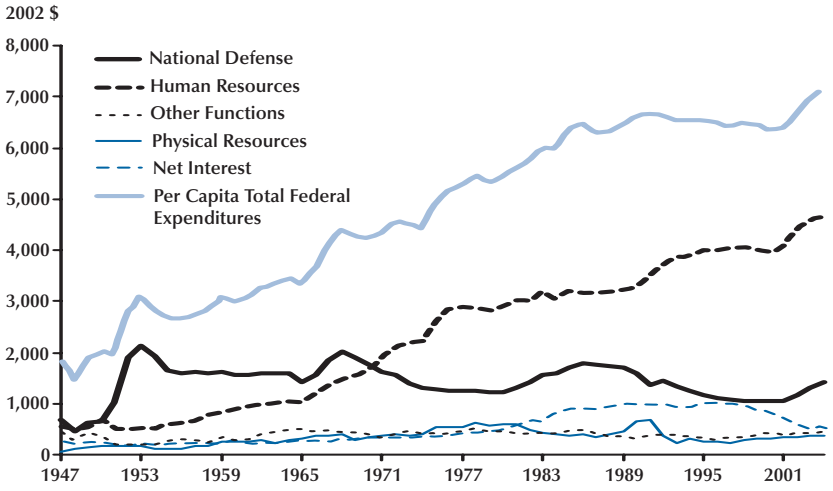
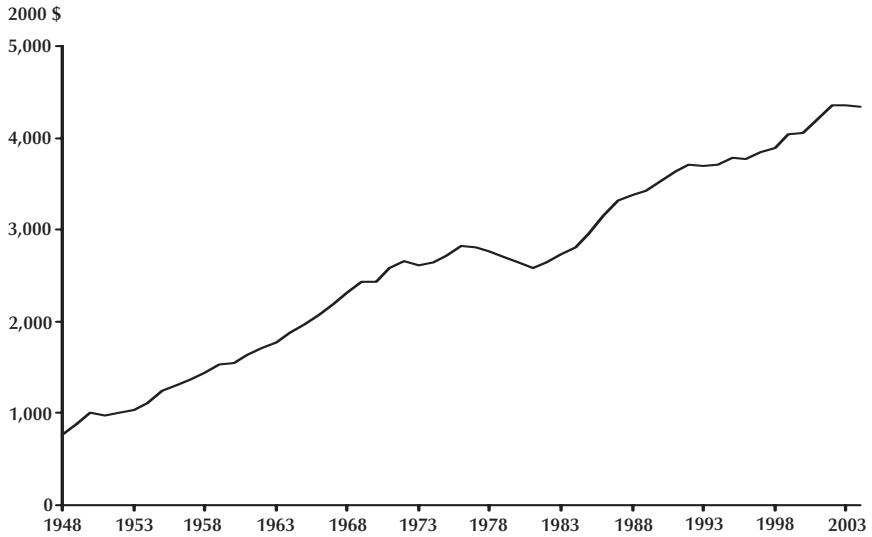


Figure 5

Real Per Capita State and Local Government Expenditures: 1948-2004



rate of 2.7 percent for real federal expenditures per person. Total government expenditures per person (federal + state + local) totaled $2,350 in 1948 and nearly $12,150 in 2004.

The data illustrated in Figures 1 through 5 provide convincing evidence that the size of government in the United States has grown throughout the 20th century.[5] An important question asked by economists and political scientists is why this growth has occurred. This paper presents several popular theories of government size and growth that have received attention in the economics and political science literature.[6] Since government and the citizenry are made up of individuals, all theories of government considered here approach the issue from a microeconomic perspective; specifically, they consider the incentives of voters, public officials, and the inherent inefficiencies that may arise in a representative democracy. Note that some theories are better suited to explain government size and others to explain government growth.

The theories of government size and growth fall into two distinct categories. The first category is citizen-over-state theories of government. These theories begin with the premise that citizens demand government programs and, as a republic, the government is simply responding to the will of the people. The other category is state-over-citizen theories of government growth. Here, the size of government is independent from citizen demand and government grows because of inherent inefficiencies in public sector activities and incentives facing government bureaucrats. The paper concludes with a discussion of the potential

[5] Another measure of government size is federal employment relative to total employment. Plotting this series over time reveals that federal employment is a diminishing share of total employment throughout the 20th century. A closer inspection of the data reveals that most of this decrease in federal employment is a result of a reduction in defense employment, which suggests that the number of federal government employees is not a good measure of the size of the government because subcontractors complete much of their work. For example, the federal government does not build military aircrafts; they pay subcontractors like Lockheed-Martin to build them. So, thousands of people working on the construction of aircrafts at Lockheed-Martin receive their pay indirectly from the federal government, and they are not included in government employment figures. In 2004, Lockheed-Martin had sales of $35.5 billion. Nearly 80 percent of sales were to the U.S. Department of Defense/Intelligence and Civil Government/Homeland Security (www.lockheedmartin.com).

[6] Kliesen (2003) discusses the increase in government size during the 20th century.

importance of the 16th Amendment to the U.S. Constitution, which allows the federal government to tax wage and business income. As will be discussed, the timing of the 16th Amendment and the start of government growth may be more than a coincidence.

CITIZEN-OVER-STATE THEORIES OF GOVERNMENT SIZE AND GROWTH

The citizen-over-state theories of government size and growth begin with the premise that government growth occurs because citizen demand for government programs has increased over time. It will become evident here that the demand for government can come from individual citizens or a collection of citizens organized into special interest groups. This section discusses three distinct citizen-over-state theories of government size and growth.

The Government as a Provider of Goods and a Reducer of Externalities

Voters decide which goods the government will provide and which negative externalities the government will correct.[7] The tool economists and other social scientists use to determine where the government will intervene is the median voter theorem. Hotelling (1929) and Downs (1957 and 1961) rank voters by political ideology and place the most conservative individual on the far right and the most liberal on the far left. Assuming a two-party system, the voters must choose the conservative candidate or the liberal candidate. Since the voter will choose the candidate with the views closest to his or her own views, whichever candidate wins the median voter will have a majority of votes and win the election.

An assumption of the median voter theorem is the use of majority rule voting. Additional assumptions are that citizens vote directly on government spending issues and that government spending is the only issue on the ballot. Thus, the median voter determines the demand for publicly provided goods, which is a function of income, the relative price of public goods to private goods, and tastes.

The price elasticity of demand for government and the price of government both determine whether government grows or contracts. Government will grow if the demand for government is price inelastic and the price of government increases. In other words, if the price of government goods or services increases and the quantity demanded of the goods or services does not decrease by a proportionate amount, total government spending increases. The other possibility for government growth is an elastic demand for government and a falling price of government. That is, if the price of government goods or services decreases and the quantity demanded of the goods or services increases by a more-than-proportionate amount, total spending increases.

The literature presents evidence in support of an increasing price and an inelastic demand for government. Baumol (1967) addresses the issue of relative private and public sector prices in terms of government growth. He shows that the increase in the price of the public sector goods and services relative to the price of the private sector goods and services is due to productivity gains in manufacturing. Since most government programs are services (i.e., national defense, education, and police), they have not experienced the same efficiency gains as manufacturing specifically and the private sector overall; thus, the relative price of public goods has been increasing.

Mueller (2003) presents additional evidence of the Baumol (1967) effect in OECD countries.[8] He found that 20 of the 25 OECD countries showed the expected growth in government expenditures as a percent of GDP from 1960 to 1995. Of the five countries that did not increase government expenditures by the amount predicted by the Baumol effect, four of the five still increased to some extent and only one decreased. It decreased

[7] A negative externality is a negative (costly) spillover from an activity onto a nonconsenting third party. An example is pollution from a factory that is dumped into a river and has an adverse affect on everyone downstream.

[8] OECD is the Organization of Economic Cooperation and Development: OECD countries are listed at www.oecd.org.

expenditures as a result of decreased defense spending after the end of the Cold War.

In addition to the productivity differences, Ferris and West (1999) found that wages in the public sector, which contribute to the price of government, are increasing faster than in the private sector. They find evidence of this in the salaries of unionized versus non-unionized public school teachers. The near-monopoly nature of publicly provided goods and services encourages the creation of unions, and they will demand higher wages. The government will appease the unions and simply pass these costs onto the taxpayers.

The remaining determinant that the literature uses to help explain the government as a provider of goods and services and reducer of externalities is citizen tastes and preferences. Over time taste for publicly provided goods and services changes and subsequently so will the demand for these goods and services. One such good is the redistribution of income and wealth for insurance purposes. Rodrik (1998) looks at the risk associated with open economies and presents evidence to support the hypothesis that the more open the economy, the larger the government. Specifically, he argues that the volatility of income and employment that corresponds with open economies is an insurable risk. The government programs that act as a form of insurance to protect workers are social programs (i.e., unemployment and social security).[9]

However, as pointed out by Mueller (2003), a problem with Rodrik's (1998) findings is that the large social programs in the United States grew at a time of significant slowdown in the domestic economy—the Great Depression—not the increased openness of the U.S. economy. Thus, social insurance programs are meant to reduce the risk of households' income volatility due, at least in part, to business cycles. The programs also attempt to smooth cash flow over a citizen's lifetime and across income levels.

[9] Ex post, not all citizens will benefit from social programs, as suggested by Garrett and Rhine (2005), which shows that less than 5 percent of 2003 retirees benefit from the Social Security system. Ex ante, however, the social program is publicly provided insurance.

The Government as a Redistributor of Income and Wealth

The second citizen-over-state theory of government surmises that government serves as a redistributor of income and wealth. All government programs are seen as mechanisms for redistribution. Meltzer and Richard (1978, 1981, 1983) present a model where leisure is inversely related to the fraction of total time worked, consumption is inversely related to the tax rate and is positively related to a lump-sum grant received from the government, and income is positively related to productivity. Their model produces a well-known result—a higher level of productivity equates to a higher level of income, and the higher income increases consumption and well-being.

Meltzer and Richard (1978, 1981, 1983) show that individuals will demand the combination of tax rates and lump sum payments that maximizes their well being. Individuals with a lower level of productivity, and subsequently a lower level of income, will demand a higher tax rate and a higher lump-sum payment from the government. The extreme case is individuals that do not work and pay no taxes; they will simply want to maximize their lump-sum payment and will demand a higher tax rate than that demanded by those individuals who are working. This model explains the growth in government in part because, over time, new entrants into the voting population are lower income workers. These lower income workers will cast votes for the candidate who will levy higher taxes and increase the amount of redistribution.

Kristov, Lindert, and McClelland (1992) explain that the amount of redistribution is based on social affinity. The closer the middle class feels to the poor, or the slower incomes are growing, the greater the amount of redistribution. The authors study the period immediately preceding and during the Great Depression as support for their claim. They explain that, when the economy was expanding, Americans voted not to increase taxes to fund relief for the poor. But, after the economy changed directions in the 1930s, social programs increased dramatically. Taxes on high earners increased, and the number of programs

that redistributed income and wealth increased as well.

Peltzman (1980) explains that candidates promise transfer payments to groups of citizens in order to gain their support. If the distribution of incomes over different socio- economic classes is similar, then the candidate must offer a greater amount of redistribution to gain supporters. With a trend toward more evenly distributed incomes in years prior to the Peltzman (1980) study, greater redistribution by the federal government was undertaken.

Interest Groups

Interest groups can increase the size of government by organizing members and applying political pressure more effectively than individual citizens (Olson, 1965, and Moe, 1980). Examples of interest groups mentioned frequently in the popular press include the Sierra Club, the National Organization of Women, and the National Rifle Association. One can think of an interest group as an organized collection of individual voters (or businesses) having the same preference for a specific policy. Through concentrated lobbying, an interest group can obtain a desired policy that has direct benefits for the interest group but the costs of the policy are spread across millions of taxpayers. Elected officials play a key role in this process as they weigh the political costs and benefits of each policy. Such disconnectedness between costs and benefits will result in inefficient levels of government expenditures—that is, the societal costs of the policy will be greater than the societal benefits.

Supply and demand analysis can be used to model an interest group economy (McCormick and Tollison, 1981). "Demanders" of a policy will be those groups that can organize and lobby for, say, $100 at a cost of less than $100. "Suppliers" (individual tax-payers) are those for whom it would cost more than $100 to lobby against losing that $100.[10] The incentives facing elected officials are such that they will target unorganized suppliers with low losses from any transfer while courting demanders who are organized and active in the political process. Thus, costs are spread across many taxpayers but the benefits are concentrated within the interest group. If too little or too much wealth is transferred, the political process will discipline the elected official at the polls.

Although economic theory can be used to explain how interest groups operate in a political market for transfers, economics has said little about how interest groups form (Olson, 1965). In fact, economic theory suggests that there would be little or no interest group formation because of the free-rider problem. Because the benefits of lobbying are nonrival and nonexcludable, it is rational for individuals who would benefit from lobbying to free-ride.[11] Despite a lack of theory for interest group formation, economics has produced dozens of papers that provide theoretical and empirical evidence on the link between interest groups and the size of government.[12]

Weingast, Shepsle, and Johnsen (1981) offer a rational explanation for the inefficiency (costs > benefits) of special interest projects. The authors focus on distributive policies that concentrate benefits within a geographic area and disperse the costs (taxes) over all constituencies. In the model of Weingast, Shepsle, and Johnsen, the national constituency is divided into districts that are each assumed to maximize its net benefits from any redistributive project and have only one representative. Because the district is only a fraction of the national constituency, the cost of the project is spread out over the entire constituency. Each district does not take into account the costs that are being placed on other constituencies when evaluating its own benefits. Thus, because the net benefits of a given project are overstated, the

[10] As suggested by Mueller (2003), the term suppliers should be taken loosely because individuals would only likely engage in the transfer under coercion.

[11] There are several ways in which interest groups can, at least partially, overcome the free-rider problem. One way is through coercion or mandatory membership, such as in the case of labor unions and state bar associations. Other interest groups may provide valuable private benefits to members, such as publications and educational material, at a relatively low cost of joining. The American Association of Retired Persons (AARP) is an example. Political entrepreneurs can also overcome the free-rider problem. Examples include many large corporations that have offices in Washington, D.C. The employees of large corporations also serve as informal lobbyists.

[12] See Ekelund and Tollison (2001) for a detailed overview of the literature on interest group theory.

project is larger than the efficient project size. Furthermore, because local projects are larger than the efficient level, the district's representative has even greater interest in acquiring projects that benefit his or her district.

Becker (1983) presents a theory of public policies that result from competition among special interest groups (or "pressure groups" according to Becker). Becker views political pressure as a public good.[13] An increase in interest group membership will increase pressure, but because pressure is a public good, free-riding (by would-be group members) will increase. Because free-riding increases, so do the costs of implementing pressure. Becker finds that efficiency in producing pressure is partly determined by the costs of controlling pressure—a greater control over free-riding increases the amount of pressure. With higher amounts of pressure, a special interest group is able to acquire more benefits (lower taxes or higher subsidies). Becker believes that efficiency is improved not only by controlling the free-rider problem, but also through the competition that occurs between tax groups and subsidy groups that consider their losses via taxes or subsidies. Therefore, interaction among competing special interest groups increases the power of the special interest lobby, and thus special interest spending.[14]

Sobel (2001) provides empirical evidence on the positive relationship between political action committees (PACs) and federal government spending.[15] He notes that the rise in federal government spending during the 1970s and 1980s and the subsequent slowdown in the 1990s parallel the increase and eventual decrease in the number of PACs over this same period. He finds that a 10 percent increase (decrease) in the number of PACs in time $t-1$ is associated with a 1.07 to 1.57 percent increase (decrease) in federal spending in time t. However, one issue is whether the number of registered PACs, as opposed to PAC membership, accurately represents the scope and power of the special interest lobby in the United States.

Although interest group theory may provide, at least in part, a reasonable explanation for the size of government, it is not without its theoretical and empirical challenges.[16] One issue is that of causality. Specifically, does interest group activity cause government spending, or do changes in the level of government spending influence interest group activity (Mueller and Murrell, 1985, 1986)? Another issue mentioned earlier is that of interest group formation. Although there are anecdotal explanations as to how interest groups can overcome the free-rider problem, such an idea has yet to be incorporated into a reasonable economic model. Finally, there is debate as to whether the interest group theory is in fact a citizen-over-state theory or a state-over-citizen theory given the pivotal role that elected officials play in the link between interest groups and government growth.

STATE-OVER-CITIZEN THEORIES OF GOVERNMENT SIZE AND GROWTH

The previous section discussed several citizen-over-state theories of government size and growth. Inherent in these theories is the idea that government is demand driven—that is, government size and growth occur because citizen demand for government has increased. This demand for government can come from individual citizens or groups of citizens (the interest group theory), with each party having a desire for some form of a publicly provided good, externality reduction, or redistribution of income.

The following section presents several theories of government growth that start from a

[13] A public good is nonrival (consumption by one person does not deny consumption by others) and nonexcludable (no price mechanism exists to deny consumption). National defense is a classic example of a public good.

[14] Note a key difference between Weingast, Shepsle, and Johnsen (1981) and Becker (1983): Weingast, Shepsle, and Johnsen believe interest groups arise as a result of concentrated benefits and dispersed costs that follow from the existence of independent districts (each district is an interest group), and it is this dispersion between costs and benefits that leads to larger government. Becker, however, believes that it is the competition among interacting interest groups that increases the power of the special interest lobby, and thus increases special interest spending.

[15] A PAC is an organization whose goal is to raise campaign funds for candidates seeking political office.

[16] Ekelund and Tollison (2001).

completely opposite premise from the previous theories—namely, that the size of government is supply driven rather than demand driven. These theories posit that the incentives facing public officials and the nature of our representative form of government provide an environment for government growth to occur in the absence of citizen demand. Government grows because of government and its inherent inefficiencies, structure (e.g., direct democracy versus a representative democracy), and incentives facing public officials. Appropriately, then, the following three theories are classified as state-over-citizen theories of government growth.

Bureaucracy Theory

Goods and services provided by the government do not arise out of thin air, but rather they must be created by a government agency. The supply of government output, then, may be a function not only of citizen demand (as the previous theories suggest), but also of the demand of government bureaucrats. Niskanen's (1971) theory of bureaucracy postulates that government bureaucrats maximize the size of their agencies' budgets in accordance with their own preferences and are able to do so because of the unique monopoly position of the bureaucrat. Because the bureaucrat provides output in response to his or her own personal preferences (e.g., the desire for salary, prestige, power), it is possible that the size of the bureaucrat's budget will be greater than the budget required to meet the demands of the citizenry. An important point is that bureaucracy theory does not deny the citizen demand models of government discussed in the previous section, but rather it suggests that bureaucrats can generate budgets that are in excess of what citizen demand warrants.

The ability of a bureaucrat to acquire a budget that is greater than the efficient level is dependent on several institutional assumptions (Niskanen, 1971, 2001). First, unlike private sector production, the public sector does not produce a specific number of units, but rather supplies a level of activity. As a result, this creates a monitoring problem for oversight agencies: It is difficult, if not impossible, for monitors to accurately judge the efficiency of production when no tangible or countable unit of output is available. Second, the monopoly nature of most bureaus shields them from competitive pressures necessary for efficiency and also denies funding agencies (Congress, the executive branch) comparable information on which to judge the efficiency of the bureau. Third, only the bureau knows its true cost schedule because bureau funding is provided by agents external to the bureau. This provides an opportunity for bureaucrats to overstate their costs in order to receive a larger budget. Finally, the bureaucrat can make take-it-or-leave-it budget proposals to the funding agency.

Niskanen (1971) shows that the bureaucrat will maximize a budget subject to the constraint that the budget must cover the costs of producing the good or service. The implication of the model is that the bureau's budget (and output) is expanded beyond the point where the marginal public benefits of the good or service equals the bureau's marginal costs of providing the good or service.[17]

Although the model presents clear reasoning on how a bureau can expand output and costs beyond the efficient level, in reality many bureaus cannot expand output beyond the level demanded by the citizenry. Examples of this at the local level include school districts and garbage collection: School districts cannot educate more students than those who are already attending school,

[17] A simple formulation of Niskanen's (1971) model of bureaucracy is:

- $B = B(Q)$, where B is the bureau's budget and Q is the perceived output of the bureau. The funding agency is aware of this public benefit schedule, $B(Q)$. It is assumed that $B' > 0$ and $B'' < 0$.
- $C = C(Q)$, where C represents the bureau's cost function, *which is known only to the bureau*. Also, $C' > 0$ and $C'' > 0$.

The bureaucrat is assumed to maximize his or her budget subject to the constraint that the budget must cover the costs of producing Q. Thus, the bureau's objective function is

$$O_B = B(Q) + \lambda(B(Q) - C(Q)).$$

Differentiating with respect to Q and λ and then rearranging terms gives

$$\frac{\partial B}{\partial Q} = \frac{\lambda}{(1+\lambda)}\frac{\partial C}{\partial Q} \quad (1)$$

$$B(Q) = C(Q). \quad (2)$$

Mueller (2003, Chap. 16) provides a detailed analysis of bureaucracy theory and presents extensions of the model presented here that relax several of the initial assumptions.

and garbage collectors cannot haul more garbage than is available for disposal. Even in these cases, however, a bureau may expand its budget beyond the efficient level—not by providing output beyond the efficient amount but by providing the services at a higher cost than necessary.

There has been ample literature that has compared the costs of public and private organizations that provide similar services. The activities or firms studied include, but are not limited to, hospitals (Clarkson, 1972), refuse collection (Bennett and Johnson, 1979, and Kemper and Quigley, 1976), water utilities (Morgan, 1977) and fire protection (Ahlbrandt, 1973). Mueller (2003, Chap. 16) provides a summary of 70 studies that examined the cost of public versus private sector provision of identical services. In all but five studies cited, the cost of public provision is significantly greater than private provision, thus lending support for the bureaucracy theory of government.

However, the cost difference between private and public organizations may simply be a result of a lack of competitive pressure rather than direct attempts by bureaucrats to maximize their budget. In addition, Mueller (2003) suggests that many of the assumptions necessary for the bureaucracy theory to hold may be too strong and actually weaken the ability of the bureaucrat to manipulate price and output.

For example, the ability of a bureau to present a take-it-or-leave-it budget proposal may be lessened if the funding agency or an oversight agency is aware of the advantage such a position affords the bureau. Thus, the funding agency may request that the bureau present several cost and output scenarios; if the bureau must present a cost schedule, it becomes more likely that the bureau will announce its true costs.[18] Also, several agencies exist, such as the U.S. General Accounting Office, that are set-up for the sole purpose of detecting excessive costs and inefficiencies in government bureaus. The possibility of an audit and the negative attention such an action brings creates an incentive for bureaucrats to limit their pricing power and, at least somewhat, promote an efficient organization.

Although the constraints on bureaucracy seem reasonable, they are somewhat limited given the number of local, state, and federal agencies that exist relative to the number of funding and oversight agencies. However, although the literature has presented strong evidence that bureaucracy may partly explain government size, much less work has been done on explaining how bureaucracy theory may explain government growth. One explanation put forth by Mueller (2003) is that the ability of a bureau to misrepresent its cost and/or output schedule is likely to be directly correlated with the bureau's size. Thus, larger bureaus can better manipulate their budgets relative to smaller bureaus, and any manipulation of the bureau's budget will increase the size of the bureau, which in turn increases the bureau's ability to manipulate the budget.

Despite the limits of bureaucracy theory, it remains a plausible explanation for the scope of government seen today. The common inefficiencies of large organizations, be they private or public, are not unknown by the general public, who often work in such organizations. In addition, it is not uncommon for the media to report waste or fraud that has occurred at large private and public organizations. The bureaucracy theory fits arguably well with the real-world experiences of many people.

Fiscal Illusion

The fiscal illusion theory assumes that government, specifically legislators and the executive branch, can deceive voters as to the true size of government. This theory is similar to the bureaucracy theory that postulated that bureaus can deceive legislators and funding agencies as to the true size of the bureau. The concept of fiscal illusion has been discussed in the economics literature for nearly a century, but Buchanan (1967) formulates the idea into a theory of government size and growth.

Fiscal illusion assumes that citizens measure the size of government by the quantity of taxes they pay. As such, taxes and tax collection meas-

[18] Bendor, Taylor, and Van Gaalen (1985) show that a bureau can charge a price higher than the efficient level (marginal costs = marginal benefits) only when the demand for the bureau's service is inelastic.

ures that are less obvious to citizens are more likely to be used by government. Examples include the federal withholding of income taxes and property tax collection through monthly mortgage payments. Although the income tax is considered a direct tax, versus indirect taxes such as gasoline or cigarette taxes, the ability of direct taxes to be disguised suggests that the collection method of some direct taxes may hide citizens' tax bills better than indirect taxes. Mueller (2003, p. 527) suggests that determining which taxes are hidden from citizens is largely an empirical issue.

Oates (1988) provides an overview of the empirical literature on fiscal illusion. He summarizes the empirical findings and develops five hypotheses to support the fiscal illusion theory of government. Oates (1988) concludes (i) tax burdens are more difficult to evaluate when the tax structure is more complicated, (ii) progressive tax structures that increase a citizen's tax bill according to income increases are less obvious than legislated changes to the tax code, (iii) homeowners are better able to judge their portion of property taxes than are renters, (iv) the issuance of debt (and thus the likelihood of future tax increases) appears less costly to voters than current tax increases, and (v) the "fly-paper effect" of government spending is real.

The fly-paper effect hypothesis deserves some explanation given the attention it has received in the literature (see Hines and Thaler, 1995). Economic theory predicts that a lump-sum increase in income to one level of government from another, say a lump-sum grant from the federal government to a state government, will increase government spending by the same amount as would an equal increase in citizen income in that state. Increases in income (revenue via taxes) or grants to the voter's government are identical because they both increase financial resources to the government. Government sets the level of expenditures desired by the median voter. Thus, when grant monies are obtained by the government, the voter can consider these grant funds as an increase in personal income via a reduction in taxes. Thus, through an efficient political process, any additional revenue from grants is offset by a decrease in tax revenue demanded by voters.

Typically, a $1 increase in personal income increases government spending by $0.05 to $0.10.[19] In the absence of a fly-paper effect, one should expect to see every $1 from a lump-sum grant to state governments (an income increase to state governments) increase government spending by the same amount, $0.05 to $0.10. However, the literature has shown that lump-sum grants increase government spending by $0.20 to $1 for every $1 in grant money, which is significantly greater than the $0.05 to $0.10 increase that would arise from an increase in median voter income.[20] Thus, the grant-money "sticks" to where it is sent, hence the term fly-paper effect. Inefficiencies in the political process and a disconnect between the preferences of the median voter and government are cited as reasons why the fly-paper effect may exist. If the fly-paper effect exists, then governments can increase spending without apparent tax increases. Increases in intergovernmental grants will need to be financed by taxes, but this tax revenue (and resulting tax burden on citizens) is not directly linked to the expenditures by the state governments.

The fly-paper effect and the broader issue of fiscal illusion are not without critics. Doubters of the fly-paper effect argue that incorrect modeling of empirical models and the political processes in the public sector as well as the failure to discern between numerous types of grants may explain the fly-paper effect found in the literature (Hamilton, 1983, and Chernick, 1979).[21] Regarding fiscal illusion, the literature does not explain exactly how government will grow if fiscal illusion is indeed present. Just because voters are unaware

[19] Hines and Thaler (1995) and Fisher (1996).

[20] Hines and Thaler (1995) summarize the results of numerous studies that present empirical estimates of the fly-paper effect.

[21] Grants can be lump-sum, matching (where the receiving government must match a certain percent of the expenditure), closed-ended, or open-ended. Whereas a lump-sum grant only creates an income effect, a matching grant creates both an income effect and a substitution effect. Economic theory predicts that matching grants will result in higher government spending than lump-sum grants (see Fisher, 1996, Chap. 9). Disentangling the effects of various forms of grants greatly complicates empirical analyses on the fly-paper effect.

Table 2

Vote Trading, Bundling, and Government Size

	Net Benefits (+) or Costs (–) to Each Voter in District			
Voters of district	Construction of post office in A	Dredging harbor in B	Construction of military base in C	Total
A	+$10	–$3	–$3	+$4
B	–$3	+$10	–$3	+$4
C	–$3	–$3	+$10	+$4
D	–$3	–$3	–$3	–$9
E	–$3	–$3	–$3	–$9
Total	–$2	–$2	–$2	–$6

SOURCE: Gwartney and Stroup (1997, p. 503).

of their true tax bill, that does not mean there is a clear method for government officials (legislators, bureaucrats) to take advantage of this situation to increase the size of government. Mueller (2003) argues that, for fiscal illusion to explain government size and growth, it must be combined with other theories of government growth discussed earlier to form a single model of government growth.

Monopoly Government and Leviathan

The idea that representative governments behave as monopolists was first suggested by Breton (1974). The party in control of the legislature has an objective function that includes the probability of reelection, personal pecuniary gain, and the pursuit of personal ideals. While providing basic public goods, such as police and fire protection (in the case of a local government), the monopoly government can obtain its objectives by bundling narrowly defined issues that benefit individual members of the government along with the more popular public-good services provided.

This idea stems from the neoclassical view of the monopolist, where a private monopolist can increase his profit by bundling other products that he does not monopolize with his monopolist product. Consumers will then buy the monopolist's package as long as their consumer surplus on the bundled products exceeds the cost of the individual packages.

In the case of governments, this bundling of services results in higher levels of government output. Tullock (1959) provides a comprehensive analysis of how the bundling of goods and vote-trading among legislators can increase the size of government. The example shown in Table 2 illustrates the point made by Tullock (1959): A five member legislature is considering the three projects, each of which is inefficient because the net costs outweigh the net benefits.[22] As a result, if each project was voted on separately (and each legislator voted according to the preferences of his constituency), then none of the projects would be implemented because each would lose by a 4-to-1 margin. But, bundling the three projects will garner "yes" votes from legislators representing districts A, B, and C, thus allowing the legislation to pass 3-to-2, thereby increasing the size of government expenditures.

The monopolist view of government has been extended further by Brennan and Buchanan (1977, 1980). In their model of a "leviathan" government, the monopoly government's sole objective is to maximize revenue. The citizenry is assumed to have lost all control over their govern-

[22] As noted in Gwartney and Stroup (1997, p. 503), vote-trading and bundling can also lead to efficient measures. The point made in the above example is that bundling can lead to greater government size.

ment, and political competition is seen as an ineffective constraint on the growth of government.[23] Their leviathan view of government is opposite of the government assumed in the citizen-over-state theories—the latter being a benevolent provider of goods, a reducer of externalities, and a redistributor of income. According to Brennan and Buchanan (1977), only constitutional constraints on the government's authority to tax and issue debt can limit a leviathan government.[24]

Empirical evidence for the monopoly view of government has provided mixed results. The studies are often conducted at the local rather than national level due to data availability. Many tests for monopoly government have a similar goal as those for the bureaucracy theory, namely that the cost of public services is greater than the costs of identical services provided by the private sector. Additional research has hypothesized that a constraint on a monopoly government is competition from neighboring governments (Martin and Wagner, 1978). This research on the monopoly power of government has shown that restrictions on incorporation raise the costs of existing local governments.

Tests for leviathan government begin with the premise that such a system should be less likely to occur when government is relatively smaller and there exists strong intergovernmental competition. As with the studies of monopoly, much of the literature on leviathan has focused on local governments (Oates, 1972, Nelson, 1987, and Zax, 1989). The mixed results obtained in these studies are due, at least in part, to the variety of methods authors use to proxy for government size. On the national level, Oates (1985) finds that countries having a federalist constitution (many levels of government) had a negative, but insignificant, effect on government growth. Much more empirical testing must be done before the leviathan view of government is broadly accepted as one plausible explanation for government growth.

A NOTE ON THE 16TH AMENDMENT TO THE U.S. CONSTITUTION

Prior to the adoption of the 16th amendment to the U.S. Constitution in 1913, the federal government was constrained from directly taxing personal income by Article 1, Section 9 of the U.S. Constitution, which reads as follows: "No Capitation, or other direct, Tax shall be laid, unless in Proportion to the Census or Enumeration herein before directed to be taken." A careful reading of this clause reveals that the federal government actually could levy a personal income tax (which is a direct tax) prior to the 16th amendment, but income tax collection had to be in apportionment to population.

The 16th amendment negated the apportionment clause written in Article 1, Section 9. The 16th amendment reads as follows: "The Congress shall have the power to lay and collect taxes on incomes, from whatever sources derived, without apportionment among the several States, and without regard to any census or enumeration." The 16th amendment was passed by Congress on July 2, 1909, and ratified on February 3, 1913.[25] What makes this amendment interesting with regard to government growth is that the dramatic rise in the size of the federal government (see Figure 1) began immediately following the ratification of the 16th amendment.

The option to levy a federal income tax that is made available to Congress does not itself imply

[23] This is a result of the rational ignorance of voters (voters don't care about the political process because the costs of doing so outweigh any benefit from their single vote) and collusion by elected officials.

[24] A revenue-maximizing government will typically not maximize revenue at a 100 percent tax rate because a tax base shrinks as tax rates increase. Consider the following: $T = r \cdot B(r)$, where T = tax revenue, r = tax rate, and B = tax base. Differentiating tax revenue (T) with respect to the tax rate (r) and manipulating terms gives the expression

$$\frac{\delta B}{\delta r} \cdot \frac{r}{B} = -1 .$$

This expression shows that tax revenues will be maximized when the elasticity of the tax base is equal to –1. If the elasticity is less than –1, then an increase in the rate will decrease the base by a larger amount, thereby decreasing revenue. On the other hand, if the elasticity is greater than –1, then an increase in the rate will decrease the base by a smaller amount, thereby increasing revenue.

[25] For an interesting history of the 16th amendment, see National Archives and Records Administration (1995) and www.ourdocuments.gov (keyword search "16th amendment").

that government will grow. The option to tax personal income means only that government has another source of revenue with on which to finance its growth. Explaining government growth must be done using the theories presented earlier—income taxes are simply a fuel that enables the engine of government growth to start.

However, the government has increased its reliance on federal income taxes over the past 90 years, the same time period in which government expenditures have increased dramatically. Personal income tax revenue as a percentage of all federal tax revenue increased from about 2 percent in 1913 to over 43 percent in 2004. Also, because of large exemptions, few people paid personal income taxes in 1913; if they did pay taxes, the rates were much lower than today. For example, in 1913 the lowest tax bracket was $0 to $20,000, with a 1 percent marginal tax rate; the highest bracket was on taxable income over $500,000, taxed at a 7 percent rate; and the personal married exemption was $4,000. In 2004 dollars, the lowest 1913 bracket and married exemption would be equal to $381,616 and $76,323, respectively, and the top 1913 bracket would be equal to a 2004 income of $10,495,000.[26] Compare this with actual 2004 tax statistics: The married exemption (no children) was $6,200, the lowest tax bracket was 10 percent on taxable income up to $7,150, and the top marginal tax rate was 35 percent on taxable income over $319,100.[27] Although the strength of any causality between the 16th amendment and later expansions of the income tax must be determined empirically, the strong correlation between these two events is compelling.[28]

SUMMARY AND CONCLUSIONS

The past 90 years has seen a dramatic rise in the size and growth of the government in the United States. This article presented various data illustrating this increase in government growth and then focused on several economic theories that attempt to explain this growth. The theories fit into one of two philosophies of government growth: either (i) the growth of government is driven by citizen demand or (ii) the growth in government is a result of government itself, brought on by inherent inefficiencies in the public sector, the personal incentives of public officials, and representative democracy.

The theories discussed in this article are not the only theories on government growth that have been raised. Researchers have suggested that electoral cycles, in conjunction with citizen demand, may play a role in the size and growth of government (Downs, 1957, and Coughlin, 1992). The expansion of the voting franchise, an arguably more controversial explanation for government growth, was suggested by Meltzer and Richard (1981); their idea is that groups of individuals that were given the right to vote were typically from the lower end of the income distribution and demanded greater government services.

Although each theory was presented here as a stand-alone explanation for government size and growth, the complexity of the public sector and the political process as well as the limits of empirical economic analysis suggest that government growth is likely to be a function of some or all of the above theories. In addition, many of the theories do a better job at either explaining size or growth, but do not adequately explain the current size of government or its growth over time. Some of the theories have not withstood empirical tests, and debate continues as to whether this is a result of incorrect theory or incorrect empirical modeling. The challenge for economists and political scientists is to formulate a single cohesive theory that accounts for all aspects of the citizen-over-state and state-over-citizen theories presented here.

[26] Calculations were made using the consumer price index (CPI).

[27] Internal Revenue Service: www.irs.gov/pub/irs-soi/02inpetr.pdf and 2004 Form 1040.

[28] Holcombe and LaCombe (1998) discuss the ratification of the 16th amendment and government growth.

REFERENCES

Ahlbrandt, Roger S. Jr. "Efficiency in the Provision of Fire Services." *Public Choice*, Fall 1973, *16*, pp. 1-15.

Baumol, William J. "The Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis."

American Economic Review, June 1967, *57*(3), pp. 415-26.

Becker, Gary S. "A Theory of Competition among Pressure Groups for Political Influence." *Quarterly Journal of Economics*, August 1983, *98*(3), pp. 371-400.

Bendor, Jonathan; Taylor, Serge and Van Gaalen, Roland. "Bureaucratic Expertise versus Legislative Authority: A Model of Deception and Monitoring in Budgeting." *American Political Science Review*, December 1985, *79*(4), pp. 1041-60.

Bennett, James T. and Johnson, Manuel H. "Public versus Private Provision of Collective Goods and Services: Garbage Collection Revisited." *Public Choice*, 1979, *34*(1), pp. 55-63.

Brennan, Geoffrey and Buchanan, James M. "Towards a Tax Constitution for Leviathan." J*ournal of Public Economics*, December 1977, *8*(3), pp. 255-73.

Brennan, Geoffrey and Buchanan, James M. *The Power to Tax: Analytical Foundations of a Fiscal Constitution*. Cambridge: Cambridge University Press, 1980.

Breton, Albert. *The Economic Theory of Representative Government*. Chicago: Aldine, 1974.

Buchanan, James. *Public Finance in Democratic Processes*. Chapel Hill, NC: University of North Carolina Press, 1967.

Chernick, Howard. "An Econometric Model of the Distribution of Project Grants," in P. Mieszkowski and W. Oakland, eds., Fiscal Federalism and Grants-in-Aid. Washington, DC: The Urban Institute, 1979.

Clarkson, Kenneth W. "Some Implications of Property Rights in Hospital Management." *Journal of Law and Economics*, October 1972, *15*(2), pp. 363-84.

Coughlin, Peter. *Probabilistic Voting Theory*. Cambridge: Cambridge University Press, 1992.

Downs, Anthony. *An Economic Theory of Democracy*. New York: Harper and Row, 1957.

Downs, Anthony. "Problems of Majority Voting: In Defense of Majority Voting," *Journal of Political Economy*, April 1961, *69*(2), pp. 192-99.

Ekelund, Robert and Tollison, Robert. "The Interest Group Theory of Government," in William Shughart and Laura Razzolini, eds., *The Elgar Companion To Public Choice*. Northhampton: Edward Elgar, 2001, pp. 357-78.

Ferris, J. Stephen and West, Edwin G. "Cost Disease verses Leviathan Explanations of Rising Government Costs: An Empirical Investigation." *Public Choice*, March 1999, *98*(3-4), pp. 307-16.

Fisher, Ronald. *State and Local Public Finance*. Chicago: Irwin, 1996.

Friedman, Milton and Schwartz, Anna J. *A Monetary History of the United States, 1867-1960*. Princeton: Princeton University Press, 1963.

Garrett, Thomas A. and Rhine, Russell M. "Social Security verses Private Retirement Accounts: A Historical Analysis." Federal Reserve Bank of St. Louis *Review*, March/April 2005, *87*(2), Part 1, pp. 103-21.

Gwartney, James and Stroup, Richard L. *Microeconomics: Private and Public Choice*. 8th Edition. Chicago: Dryden Press, 1997.

Hamilton, Bruce W. "The Fly Paper Effect and Other Anomalies." *Journal of Public Economics*, December 1983, *22*(3), pp. 347-61.

Hayek, Frederick A. von. *The Road to Serfdom*. London: George Routledge and Sons, 1944.

Hines, James R. Jr. and Thaler, Richard H. "The Fly Paper Effect." *The Journal of Economic Perspectives*, Fall 1995, *9*(4), pp. 217-26.

Holcombe, Randall G. and Lacombe, Donald J. "Interests versus Ideology in the Ratification of the 16th and 17th Amendments." *Economics and Politics*, July 1998, *10*(2), pp. 143-59.

Hotelling, Harold. "Stability in Competition." *Economic Journal*, March 1929, *39*(153), pp. 41-57.

Kemper, Peter and Quigley, John M. *The Economics of Refuse Collection*. Cambridge, MA: Balinger, 1976.

Keynes, John Maynard. *The General Theory of Employment, Interest, and Money*. New York: Harcourt Brace, 1936.

Kliesen, Kevin. "Big Government. The Comeback Kid?" Federal Reserve Bank of St. Louis *Regional Economist*, January 2003.

Kristov, Lorenzo; Lindert, Peter and McClelland, Robert. "Pressure Groups and Redistribution." *Journal of Public Economics*, July 1992, *48*(2), pp. 135-63.

Martin, Dolores T. and Wagner, Richard E. "The Institutional Framework for Municipal Incorporations: An Economic Analysis of Local Agency Formation Commissions in California." *Journal of Law and Economics*, October 1978, *21*(2), pp. 409-25.

McCormick, Robert and Tollison, Robert. *Politicians, Legislation, and the Economy: An Inquiry into the Interest Group Theory of Government.* Boston: Martinus Nijhoff, 1981.

Meltzer, Allan H. and Richard, Scott F. "Why Government Grows (and Grows) in a Democracy." *Public Interest*, Summer 1978, *52*, pp. 111-18.

Meltzer, Allan H. and Richard, Scott F. "A Rational Theory of the Size of Government." *Journal of Political Economy*, October 1981, *89*(5), pp. 914-27.

Meltzer, Allan H. and Richard, Scott F. "Tests of a Rational Theory of the Size of Government." *Public Choice*, 1983, *41*(3), pp. 403-18.

Moe, Terry M. *The Organization of Interests: Incentives and the Internal Dynamics of Political Interest Groups*. Chicago: University of Chicago Press, 1980.

Morgan, W. "Investor Owned vs. Publicly Owned Water Agencies: An Evaluation of the Property Rights Theory of the Firm." *Water Resources Bulletin*, 1977, *13*(4), pp. 775-81.

Mueller, Dennis C. *Public Choice III*. Cambridge: Cambridge University Press, 2003.

Mueller, Dennis and Murrell, Peter. "Interest Groups and the Political Economy of Government Size," in Francesco Forte and Alan Peacock, eds., *Public Expenditures and Government Growth*. Oxford: Basil Blackwell, 1985.

Mueller, Dennis C. and Murrell, Peter. "Interest Groups and the Size of Government." *Public Choice*, 1986, *48*(2), pp. 125-45.

National Archives and Records Administration. *Milestone Documents in the National Archives*. Washington, DC: 1995, pp. 69-73.

Nelson, Michael A. "Searching for Leviathan: Comment and Extension." *American Economic Review*, March 1987, *77*(1), pp. 198-204.

Niskanen, William. *Bureaucracy and Representative Government*. Chicago: Aldine-Atherton, 1971.

Niskanen, William. "Bureaucracy," in William Shughart and Laura Razzolini, eds., *The Elgar Companion To Public Choice*. Northhampton: Edward Elgar, 2001.

Oates, Wallace E. *Fiscal Federalism*. New York: Harcourt Brace Jovanovich, 1972.

Oates, Wallace E. "Searching for Leviathan: An Empirical Study." *American Economic Review*, September 1985, *75*(4), pp. 748-57.

Oates, Wallace E. "On the Nature and Measurement of Fiscal Illusion: A Survey," in G. Brennan et al., eds., *Taxation and Fiscal Federalism: Essays in Honour of Russell Mathews*. Sydney: Australian National University Press, 1988.

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

Peltzman, Sam. "The Growth of Government." *Journal of Law and Economics*, October 1980, *23*(2), pp. 209-87.

Rodrik, Dani. "Why Do More Open Economies Have Bigger Governments," *Journal of Political Economy*, October 1998, *106*(5), pp. 997-1032.

Sobel, Russell S. "The Budget Surplus: A Public Choice Explanation." Working Paper 2001-05, West Virginia University, 2001.

Tullock, Gordon. "Problems of Majority Voting." *Journal of Political Economy*, December 1959, *67*(6), pp. 571-79.

Weingast, Barry R.; Shepsle, Kenneth A. and Johnsen, Christopher. "The Political Economy of Benefits and Costs: A Neoclassical Approach to Distributive Politics." *Journal of Political Economy*, August 1981, *89*(4), pp. 642-64.

Yergin, Daniel and Stanislaw, Joseph. *The Commanding Heights: The Battle for the World Economy*. New York: Simon and Schuster, 2002.

Zax, Jeffrey S. "Is There a Leviathan in Your Neighborhood?" *American Economic Review*, June 1989, *79*(3), pp. 560-67.

The Evolution of the Subprime Mortgage Market

Souphala Chomsisengphet and Anthony Pennington-Cross

This paper describes subprime lending in the mortgage market and how it has evolved through time. Subprime lending has introduced a substantial amount of risk-based pricing into the mortgage market by creating a myriad of prices and product choices largely determined by borrower credit history (mortgage and rental payments, foreclosures and bankruptcies, and overall credit scores) and down payment requirements. Although subprime lending still differs from prime lending in many ways, much of the growth (at least in the securitized portion of the market) has come in the least-risky (A–) segment of the market. In addition, lenders have imposed prepayment penalties to extend the duration of loans and required larger down payments to lower their credit risk exposure from high-risk loans.

Federal Reserve Bank of St. Louis *Review*, January/February 2006, *88*(1), pp. 31-56.

INTRODUCTION AND MOTIVATION

Homeownership is one of the primary ways that households can build wealth. In fact, in 1995, the typical household held no corporate equity (Tracy, Schneider, and Chan, 1999), implying that most households find it difficult to invest in anything but their home. Because homeownership is such a significant economic factor, a great deal of attention is paid to the mortgage market.

Subprime lending is a relatively new and rapidly growing segment of the mortgage market that expands the pool of credit to borrowers who, for a variety of reasons, would otherwise be denied credit. For instance, those potential borrowers who would fail credit history requirements in the standard (prime) mortgage market have greater access to credit in the subprime market. Two of the major benefits of this type of lending, then, are the increased numbers of homeowners and the opportunity for these homeowners to create wealth.

Of course, this expanded access comes with a price: At its simplest, subprime lending can be described as high-cost lending.

Borrower cost associated with subprime lending is driven primarily by two factors: credit history and down payment requirements. This contrasts with the prime market, where borrower cost is primarily driven by the down payment alone, given that minimum credit history requirements are satisfied.

Because of its complicated nature, subprime lending is simultaneously viewed as having great promise and great peril. The promise of subprime lending is that it can provide the opportunity for homeownership to those who were either subject to discrimination or could not qualify for a mortgage in the past.[1] In fact, subprime lending is most

[1] See Hillier (2003) for a thorough discussion of the practice of "redlining" and the lack of access to lending institutions in predominately minority areas. In fact, in the 1930s the Federal Housing Authority (FHA) explicitly referred to African Americans and other minority groups as adverse influences. By the 1940s, the Justice Department had filed criminal and civil antitrust suits to stop redlining.

Souphala Chomsisengphet is a financial economist at the Office of the Comptroller of the Currency. Anthony Pennington-Cross is a senior economist at the Federal Reserve Bank of St. Louis. The views expressed here are those of the individual authors and do not necessarily reflect the official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System, the Board of Governors, the Office of Comptroller of the Currency, or other officers, agencies, or instrumentalities of the United States government.

© 2006, The Federal Reserve Bank of St. Louis. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

prevalent in neighborhoods with high concentrations of minorities and weaker economic conditions (Calem, Gillen, and Wachter, 2004, and Pennington-Cross, 2002). However, because poor credit history is associated with substantially more delinquent payments and defaulted loans, the interest rates for subprime loans are substantially higher than those for prime loans.

Preliminary evidence indicates that the probability of default is at least six times higher for nonprime loans (loans with high interest rates) than prime loans. In addition, nonprime loans are less sensitive to interest rate changes and, as a result, subprime borrowers have a harder time taking advantage of available cheaper financing (Pennington-Cross, 2003, and Capozza and Thomson, 2005). The Mortgage Bankers Association of America (MBAA) reports that subprime loans in the third quarter of 2002 had a delinquency rate $5^1/_2$ times higher than that for prime loans (14.28 versus 2.54 percent) and the rate at which foreclosures were begun for subprime loans was more than 10 times that for prime loans (2.08 versus 0.20 percent). Therefore, the propensity of borrowers of subprime loans to fail as homeowners (default on the mortgage) is much higher than for borrowers of prime loans.

This failure can lead to reduced access to financial markets, foreclosure, and loss of any equity and wealth achieved through mortgage payments and house price appreciation. In addition, any concentration of foreclosed property can potentially adversely impact the value of property in the neighborhood as a whole.

Traditionally, the mortgage market set minimum lending standards based on a borrower's income, payment history, down payment, and the local underwriter's knowledge of the borrower. This approach can best be characterized as using nonprice credit rationing. However, the subprime market has introduced many different pricing tiers and product types, which has helped to move the mortgage market closer to price rationing, or risk-based pricing. The success of the subprime market will in part determine how fully the mortgage market eventually incorporates pure price rationing (i.e., risk-based prices for each borrower).

This paper provides basic information about subprime lending and how it has evolved, to aid the growing literature on the subprime market and related policy discussions. We use data from a variety of sources to study the subprime mortgage market: For example, we characterize the market with detailed information on 7.2 million loans leased from a private data provider called LoanPerformance. With these data, we analyze the development of subprime lending over the past 10 years and describe what the subprime market looks like today. We pay special attention to the role of credit scores, down payments, and prepayment penalties.

The results of our analysis indicate that the subprime market has grown substantially over the past decade, but the path has not been smooth. For instance, the market expanded rapidly until 1998, then suffered a period of retrenchment, but currently seems to be expanding rapidly again, especially in the least-risky segment of the subprime market (A– grade loans). Furthermore, lenders of subprime loans have increased their use of mechanisms such as prepayment penalties and large down payments to, respectively, increase the duration of loans and mitigate losses from defaulted loans.

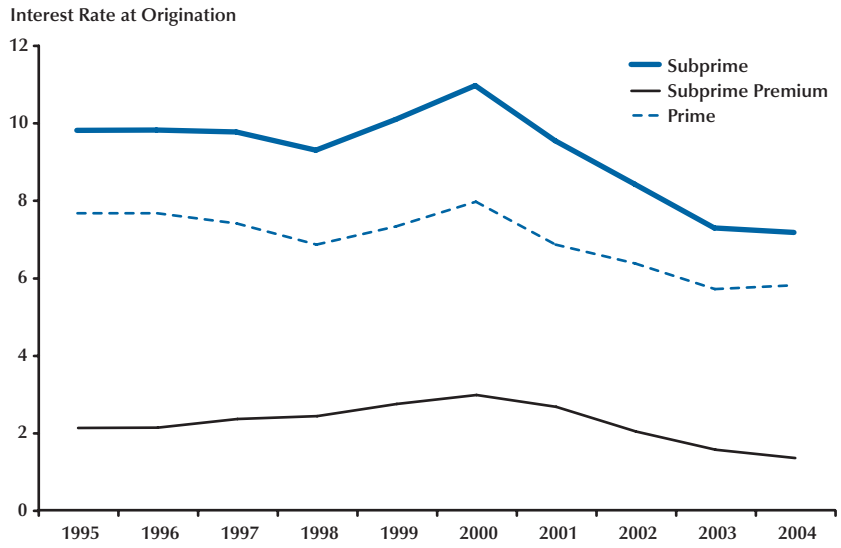
WHAT MAKES A LOAN SUBPRIME?

From the borrower's perspective, the primary distinguishing feature between prime and subprime loans is that the upfront and continuing costs are higher for subprime loans. Upfront costs include application fees, appraisal fees, and other fees associated with originating a mortgage. The continuing costs include mortgage insurance payments, principle and interest payments, late fees and fines for delinquent payments, and fees levied by a locality (such as property taxes and special assessments).

Very little data have been gathered on the extent of upfront fees and how they differ from prime fees. But, as shown by Fortowsky and LaCour-Little (2002), many factors, including borrower credit history and prepayment risk, can substantially affect the pricing of loans. Figure 1 compares interest rates for 30-year fixed-rate loans in the prime and the subprime markets. The

Figure 1

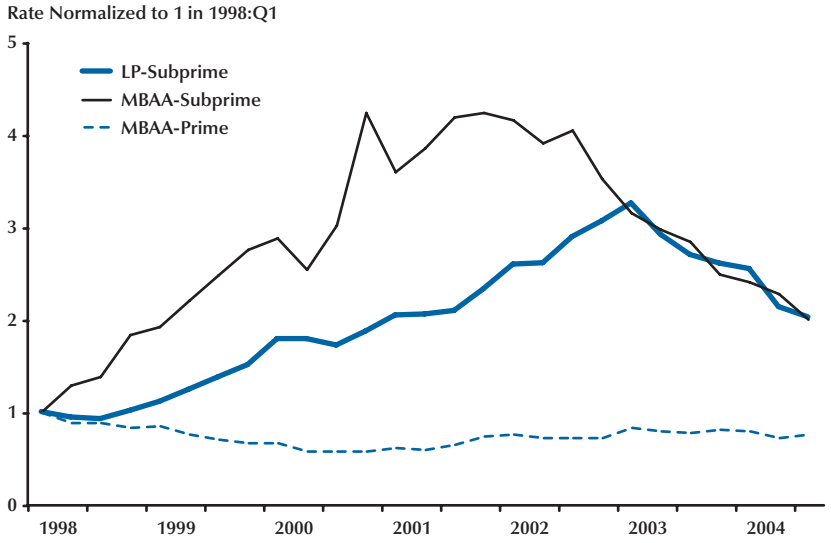
Interest Rates



NOTE: Prime is the 30-year fixed interest rate reported by the Freddie Mac Primary Mortgage Market Survey. Subprime is the average 30-year fixed interest rate at origination as calculated from the LoanPerformance data set. The Subprime Premium is the difference between the prime and subprime rates.

Figure 2

Foreclosures In Progress



NOTE: The rate of foreclosure in progress is normalized to 1 in the first quarter of 1998. MBAA indicates the source is the Mortgage Bankers Association of America and LP indicates that the rate is calculated from the LoanPerformance ABS data set.

Table 1

Underwriting and Loan Grades

Credit history	Premier Plus	Premier	A–	B	C	C–
Mortgage delinquency in days	0 x 30 x 12	1 x 30 x 12	2 x 30 x 12	1 x 60 x 12	1 x 90 x 12	2 x 90 x 12
Foreclosures	>36 months	>36 months	>36 months	>24 months	>12 months	>1 day
Bankruptcy, Chapter 7	Discharged >36 months	Discharged >36 months	Discharged >36 months	Discharged >24 months	Discharged >12 months	Discharged
Bankruptcy, Chapter 13	Discharged >24 months	Discharged >24 months	Discharged >24 months	Discharged >18 months	Filed >12 months	Pay
Debt ratio	50%	50%	50%	50%	50%	50%

SOURCE: Countrywide, downloaded from www.cwbc.com on 2/11/05.

prime interest rate is collected from the Freddie Mac Primary Mortgage Market Survey. The subprime interest rate is the average 30-year fixed-rate at origination as calculated from the LoanPerformance data set. The difference between the two in each month is defined as the subprime premium. The premium charged to a subprime borrower is typically around 2 percentage points. It increases a little when rates are higher and decreases a little when rates are lower.

From the lender's perspective, the cost of a subprime loan is driven by the loan's termination profile.[2] The MBAA reports (through the MBAA delinquency survey) that 4.48 percent of subprime and 0.42 percent of prime fixed-rate loans were in foreclosure during the third quarter of 2004. According to LoanPerformance data, 1.55 percent of fixed-rate loans were in foreclosure during the same period. (See the following section "Evolution of Subprime Lending" for more details on the differences between these two data sources.) Figure 2 depicts the prime and subprime loans in foreclosure from 1998 to 2004. For comparison, the rates are all normalized to 1 in the first quarter of 1998 and only fixed-rate loans are included.

The figure shows that foreclosures on prime loans declined slightly from 1998 through the third quarter of 2004. In contrast, both measures of subprime loan performance showed substantial increases. For example, from the beginning of the sample to their peaks, the MBAA measure increased nearly fourfold and the LoanPerformance measure increased threefold. Both measures have been declining since 2003. These results show that the performance and termination profiles for subprime loans are much different from those for prime loans, and after the 2001 recession it took nearly two years for foreclosure rates to start declining in the subprime market. It is also important to note that, after the recession, the labor market weakened but the housing market continued to thrive (high volume with steady and increasing prices). Therefore, there was little or no equity erosion caused by price fluctuations during the recession. It remains to be seen how subprime loans would perform if house prices declined while unemployment rates increased.

The rate sheets and underwriting matrices from Countrywide Home Loans, Inc. (download from www.cwbc.com on 2/11/05), a leading lender and servicer of prime and subprime loans, provide some details typically used to determine what type of loan application meets subprime underwriting standards.

Countrywide reports six levels, or loan grades, in its B&C lending rate sheet: Premier Plus, Premier, A–, B, C, and C–. The loan grade is determined by the applicant's mortgage or rent payment history, bankruptcies, and total debt-to-income ratio. Table 1 provides a summary of the four

[2] The termination profile determines the likelihood that the borrower will either prepay or default on the loan.

Table 2

Underwriting and Interest Rates

		LTV				
Loan grade	**Credit score**	60%	70%	80%	90%	100%
Premier Plus	680	5.65	5.75	5.80	5.90	7.50
	660	5.65	5.75	5.85	6.00	7.85
	600	5.75	5.80	5.90	6.60	8.40
	580	5.75	5.85	6.00	6.90	8.40
	500	6.40	6.75	7.90		
Premier	680	5.80	5.90	5.95	5.95	7.55
	660	5.80	5.90	6.00	6.05	7.90
	600	5.90	5.95	6.05	6.65	8.45
	580	5.90	6.00	6.15	6.95	
	500	6.55	6.90	8.05		
A–	680					
	660	6.20	6.25	6.35	6.45	
	600	6.35	6.45	6.50	6.70	
	580	6.35	6.45	6.55	7.20	
	500	6.60	6.95	8.50		
B	680					
	660	6.45	6.55	6.65		
	600	6.55	6.60	6.75		
	580	6.55	6.65	6.85		
	500	6.75	7.25	9.20		
C	680					
	660					
	600	6.95	7.20			
	580	7.00	7.30			
	500	7.45	8.95			
C–	680					
	660					
	600					
	580	7.40	7.90			
	500	8.10	9.80			

NOTE: The first three years are at a fixed interest rate, and there is a three-year prepayment penalty.

SOURCE: Countrywide California B&C Rate Sheet, downloaded from www.cwbc.com on 2/11/05.

underwriting requirements used to determine the loan grade. For example, to qualify for the Premier Plus grade, the applicant may have had no mortgage payment 30 days or more delinquent in the past year (0 x 30 x 12). The requirement is slowly relaxed for each loan grade: the Premier grade allows one payment to be 30-days delinquent; the A– grade allows two payments to be 30-days delinquent; the B grade allows one payment to be 60-days delinquent; the C grade allows

one payment to be 90-days delinquent; and the C– grade allows two payments to be 90-days delinquent. The requirements for foreclosures are also reduced for the lower loan grades. For example, whereas the Premier Plus grade stipulates no foreclosures in the past 36 months, the C grade stipulates no foreclosures only in the past 12 months, and the C– grade stipulates no active foreclosures. For most loan grades, Chapter 7 and Chapter 13 bankruptcies typically must have been discharged at least a year before application; however, the lowest grade, C–, requires only that Chapter 7 bankruptcies have been discharged and Chapter 13 bankruptcies at least be in repayment. However, all loan grades require at least a 50 percent ratio between monthly debt servicing costs (which includes all outstanding debts) and monthly income.

Loan grade alone does not determine the cost of borrowing (that is, the interest rate on the loan). Table 2 provides a matrix of credit scores and loan-to-value (LTV) ratio requirements that determine pricing of the mortgage within each loan grade for a 30-year loan with a 3-year fixed interest rate and a 3-year prepayment penalty. For example, loans in the Premier Plus grade with credit scores above 680 and down payments of 40 percent or more would pay interest rates of 5.65 percentage points, according to the Countrywide rate sheet for California. As the down payment gets smaller (as LTV goes up), the interest rate increases. For example, an applicant with the same credit score and a 100 percent LTV will be charged a 7.50 interest rate. But, note that the interest rate is fairly stable until the down payment drops below 10 percent. At this point the lender begins to worry about possible negative equity positions in the near future due to appraisal error or price depreciation.

It is the combination of smaller down payments and lower credit scores that lead to the highest interest rates. In addition, applicants in lower loan grades tend to pay higher interest rates than similar applicants in a higher loan grade. This extra charge reflects the marginal risk associated with missed mortgage payments, foreclosures, or bankruptcies in the past. The highest rate quoted is 9.8 percentage points for a C– grade loan with the lowest credit score and a 30 percent down payment.

The range of interest rates charged indicates that the subprime mortgage market actively price discriminates (that is, it uses risk-based pricing) on the basis of multiple factors: delinquent payments, foreclosures, bankruptcies, debt ratios, credit scores, and LTV ratios. In addition, stipulations are made that reflect risks associated with the loan grade and include any prepayment penalties, the length of the loan, the flexibility of the interest rate (adjustable, fixed, or hybrid), the lien position, the property type, and other factors.

The lower the grade or credit score, the larger the down payment requirement. This requirement is imposed because loss severities are strongly tied to the amount of equity in the home (Pennington-Cross, forthcoming) and price appreciation patterns.

As shown in Table 2, not all combinations of down payments and credit scores are available to the applicant. For example, Countrywide does not provide an interest rate for A– grade loans with no down payment (LTV = 100 percent). Therefore, an applicant qualifying for grade A– but having no down payment must be rejected. As a result, subprime lending rations credit through a mixture of risk-based pricing (price rationing) and minimum down payment requirements, given other risk characteristics (nonprice rationing).

In summary, in its simplest form, what makes a loan subprime is the existence of a premium above the prevailing prime market rate that a borrower must pay. In addition, this premium varies over time, which is based on the expected risks of borrower failure as a homeowner and default on the mortgage.

A BRIEF HISTORY OF SUBPRIME LENDING

It was not until the mid- to late 1990s that the strong growth of the subprime mortgage market gained national attention. Immergluck and Wiles (1999) reported that more than half of subprime

Table 3

Total Originations—Consolidation and Growth

Year	Total B&C originations (billions)	Top 25 B&C originations (billions)	Top 25 market share of B&C	Total originations	B&C market share of total
1995	$65.0	$25.5	39.3%	$639.4	10.2%
1996	$96.8	$45.3	46.8%	$785.3	12.3%
1997	$124.5	$75.1	60.3%	$859.1	14.5%
1998	$150.0	$94.3	62.9%	$1,450.0	10.3%
1999	$160.0	$105.6	66.0%	$1,310.0	12.2%
2000	$138.0	$102.2	74.1%	$1,048.0	13.2%
2001	$173.3	$126.8	73.2%	$2,058.0	8.4%
2002	$213.0	$187.6	88.1%	$2,680.0	7.9%
2003	$332.0	$310.1	93.4%	$3,760.0	8.8%

SOURCE: *Inside B&C Lending*. Individual firm data are from *Inside B&C Lending* and are generally based on security issuance or previously reported data.

refinances[3] originated in predominately African-American census tracts, whereas only one tenth of prime refinances originated in predominately African-American census tracts. Nichols, Pennington-Cross, and Yezer (2005) found that credit-constrained borrowers with substantial wealth are most likely to finance the purchase of a home by using a subprime mortgage.

The growth of subprime lending in the past decade has been quite dramatic. Using data reported by the magazine *Inside B&C Lending*, Table 3 reports that total subprime or B&C originations (loans) have grown from $65 billion in 1995 to $332 billion in 2003. Despite this dramatic growth, the market share for subprime loans (referred to in the table as B&C) has dropped from a peak of 14.5 percent in 1997 to 8.8 percent in 2003. During this period, homeowners refinanced existing mortgages in surges as interest rates dropped. Because subprime loans tend to be less responsive to changing interest rates (Pennington-Cross, 2003), the subprime market share should tend to drop during refinancing booms.

The financial markets have also increasingly securitized subprime loans. Table 4 provides the securitization rates calculated as the ratio of the total number of dollars securitized divided by the number of dollars originated in each calendar year. Therefore, this number roughly approximates the actual securitization rate, but could be under or over the actual rate due to the packaging of seasoned loans.[4] The subprime loan securitization rate has grown from less than 30 percent in 1995 to over 58 percent in 2003. The securitization rate for conventional and jumbo loans has also increased over the same time period.[5] For example, conventional securitization rates have increased from close to 50 percent in 1995-97 to more than 75 percent in 2003. In addition, all or almost all of the loans insured by government loans are securitized. Therefore, the subprime mortgage market has become more similar to the prime market over time. In fact, the 2003 securitization rate of subprime loans is comparable to that of prime loans in the mid-1990s.

[3] A refinance is a new loan that replaces an existing loan, typically to take advantage of a lower interest rate on the mortgage.

[4] Seasoned loans refers to loans sold into securities after the date of origination.

[5] Conventional loans are loans that are eligible for purchase by Fannie Mae and Freddie Mac because of loan size and include loans purchased by Fannie Mae and Freddie Mac, as well as those held in a portfolio or that are securitized through a private label. Jumbo loans are loans with loan amounts above the government-sponsored enterprise (conventional conforming) loan limit.

Table 4

Securitization Rates

	Loan type			
Year	FHA/VA	Conventional	Jumbo	Subprime
1995	101.1%	45.6%	23.9%	28.4%
1996	98.1%	52.5%	21.3%	39.5%
1997	100.7%	45.9%	32.1%	53.0%
1998	102.3%	62.2%	37.6%	55.1%
1999	88.1%	67.0%	30.1%	37.4%
2000	89.5%	55.6%	18.0%	40.5%
2001	102.5%	71.5%	31.4%	54.7%
2002	92.6%	72.8%	32.0%	57.6%
2003	94.9%	75.9%	35.1%	58.7%

NOTE: Subprime securities include both MBS and ABS backed by subprime loans. Securitization rate = securities issued divided by originations in dollars.

SOURCE: *Inside MBS & ABS*.

Many factors have contributed to the growth of subprime lending. Most fundamentally, it became legal. The ability to charge high rates and fees to borrowers was not possible until the Depository Institutions Deregulation and Monetary Control Act (DIDMCA) was adopted in 1980. It preempted state interest rate caps. The Alternative Mortgage Transaction Parity Act (AMTPA) in 1982 permitted the use of variable interest rates and balloon payments.

These laws opened the door for the development of a subprime market, but subprime lending would not become a viable large-scale lending alternative until the Tax Reform Act of 1986 (TRA). The TRA increased the demand for mortgage debt because it prohibited the deduction of interest on consumer loans, yet allowed interest deductions on mortgages for a primary residence as well as one additional home. This made even high-cost mortgage debt cheaper than consumer debt for many homeowners. In environments of low and declining interest rates, such as the late 1990s and early 2000s, cash-out refinancing[6] becomes a popular mechanism for homeowners to access the value of their homes. In fact, slightly over one-half of subprime loan originations have been for cash-out refinancing.[7]

In addition to changes in the law, market changes also contributed to the growth and maturation of subprime loans. In 1994, for example, interest rates increased and the volume of originations in the prime market dropped. Mortgage brokers and mortgage companies responded by looking to the subprime market to maintain volume. The growth through the mid-1990s was funded by issuing mortgage-backed securities (MBS, which are sometimes also referred to as private label or as asset-backed securities [ABS]). In addition, subprime loans were originated mostly by nondepository and monoline finance companies.

During this time period, subprime mortgages were relatively new and apparently profitable, but the performance of the loans in the long run was not known. By 1997, delinquent payments and defaulted loans were above projected levels and an accounting construct called "gains-on sales

[6] Cash-out refinancing indicates that the new loan is larger than the old loan and the borrower receives the difference in cash.

[7] One challenge the subprime industry will face in the future is the need to develop business plans to maintain volume when interest rates rise. This will likely include a shift back to home equity mortgages and other second-lien mortgages.

Table 5

Top Ten B&C Originators, Selected Years

Rank	2003	2002
1	Ameriquest Mortgage, CA	Household Finance, IL
2	New Century, CA	CitiFinancial, NY
3	CitiFinancial, NY	Washington Mutual, WA
4	Household Finance, IL	New Century, CA
5	Option One Mortgage, CA	Option One Mortgage, CA
6	First Franklin Financial Corp, CA	Ameriquest Mortgage, DE
7	Washington Mutual, WA	GMAC-RFC, MN
8	Countrywide Financial, CA	Countrywide Financial, CA
9	Wells Fargo Home Mortgage, IA	First Franklin Financial Corp, CA
10	GMAC-RFC, MN	Wells Fargo Home Mortgage, IA

	2001	2000
1	Household Finance, IL	CitiFinancial Credit Co, MO
2	CitiFinancial, NY	Household Financial Services, IL
3	Washington Mutual, WA	Washington Mutual, WA
4	Option One Mortgage, CA	Bank of America Home Equity Group, NC
5	GMAC-RFC, MN	GMAC-RFC, MN
6	Countrywide Financial, CA	Option One Mortgage, CA
7	First Franklin Financial Corp, CA	Countrywide Financial, CA
8	New Century, CA	Conseco Finance Corp. (Green Tree), MN
9	Ameriquest Mortgage, CA	First Franklin, CA
10	Bank of America, NC	New Century, CA

	1996
1	Associates First Capital, TX
2	The Money Store, CA
3	ContiMortgage Corp, PA
4	Beneficial Mortgage Corp, NJ
5	Household Financial Services, IL
6	United Companies, LA
7	Long Beach Mortgage, CA
8	EquiCredit, FL
9	Aames Capital Corp., CA
10	AMRESCO Residential Credit, NJ

NOTE: B&C loans are defined as less than A quality non-agency (private label) paper loans secured by real estate. Subprime mortgage and home equity lenders were asked to report their origination volume by *Inside B&C Lending*. Wholesale purchases, including loans closed by correspondents, are counted.

SOURCE: *Inside B&C Lending*.

accounting" magnified the cost of the unanticipated losses. In hindsight, many lenders had underpriced subprime mortgages in the competitive and high-growth market of the early to mid-1990s (Temkin, Johnson, and Levy, 2002).

By 1998, the effects of these events also spilled over into the secondary market. MBS prices dropped, and lenders had difficulty finding investors to purchase the high-risk tranches. At or at about the same time, the 1998 Asian financial crisis greatly increased the cost of borrowing and again reduced liquidity in the all-real-estate markets. This impact can be seen in Table 4, where the securitization rate of subprime loans drops from 55.1 percent in 1998 to 37.4 percent in 1999. In addition, the volume of originations shown in Table 3 indicates that they dropped from $105.6 billion in 1999 to $102.2 billion in 2000. Both of these trends proved only transitory because both volume and securitization rates recovered in 2000-03.

Partially because of these events, the structure of the market also changed dramatically through the 1990s and early 2000s. The rapid consolidation of the market is shown in Table 3. For example, the market share of the top 25 firms making subprime loans grew from 39.3 percent in 1995 to over 90 percent in 2003.

Many firms that started the subprime industry either have failed or were purchased by larger institutions. Table 5 shows the top 10 originators for 2000-03 and 1996. From 2000 forward the list of top originators is fairly stable. For example, CitiFinancial, a member of Citigroup, appears each year, as does Washington Mutual and Countrywide Financial. The largest firms increasingly dominated the smaller firms from 2000 through 2003, when the market share of the top 25 originators increased from 74 percent to 93 percent.

In contrast, many of the firms in the top 25 in 1996 do not appear in the later time periods. This is due to a mixture of failures and mergers. For example, Associated First Capital was acquired by Citigroup and at least partially explains Citigroup's position as one of the top originators and servicers of subprime loans. Long Beach Mortgage was purchased by Washington Mutual, one of the nation's largest thrifts. United Companies filed for bankruptcy, and Aames Capital Corporation was delisted after significant financial difficulties. Household Financial Services, one of the original finance companies, has remained independent and survived the period of rapid consolidation. In fact, in 2003 it was the fourth largest originator and number two servicer of loans in the subprime industry.

THE EVOLUTION OF SUBPRIME LENDING

This section provides a detailed picture of the subprime mortgage market and how it has evolved from 1995 through 2004. We use individual loan data leased from LoanPerformance. The data track securities issued in the secondary market. Data sources include issuers, broker dealers/deal underwriters, servicers, master servicers, bond and trust administrators, trustees, and other third parties.

As of March 2003, more than 1,000 loan pools were included in the data. LoanPerformance estimates that the data cover over 61 percent of the subprime market. Therefore, it represents the segment of the subprime market that is securitized and could potentially differ from the subprime market as a whole. For example, the average rate of subprime loans in foreclosure reported by the LoanPerformance data is 35 percent of the rate reported by the MBAA. The MBAA, which does indicate that their sample of loans is not representative of the market, classifies loans as subprime based on lender name. The survey of lenders of prime and subprime loans includes approximately 140 participants. As will be noted later in the section, the LoanPerformance data set is dominated by the A–, or least risky, loan grade, which may in part explain the higher rate of foreclosures in the MBAA data. In addition, the demand for subprime securities should impact product mix.

The LoanPerformance data set provides a host of detailed information about individual loans that is not available from other data sources. (For example, the MBAA data report delinquency and foreclosure rates but do not indicate any information about the credit score of the borrower, down

payment, existence of prepayment penalties, or interest rate of the loan.[8]) The data set includes many of the standard loan application variables such as the LTV ratio, credit score, loan amount, term, and interest rate type. Some "cleaning" of the data is conducted. For example, in each tabulation, only available data are used. Therefore, each figure may represent a slightly different sample of loans. In addition, to help make the results more comparable across figures, only adjustable- and fixed-rate loans to purchase or refinance a home (with or without cash out) are included from January 1995 through the December of 2004. But because of the delay in data reporting, the estimates for 2004 will not include all loans from that year.

Volume

Although the subprime mortgage market emerged in the early 1980s with the adoption of DIDMCA, AMTPA, and TRA, subprime lending rapidly grew only after 1995, when MBS with subprime-loan collateral become more attractive to investors. Figure 3 illustrates this pattern using our data (LoanPerformance) sample. In 1995, for example, the number of subprime fixed-rate mortgages (FRMs) originated was just slightly above 62,000 and the number of subprime adjustable-rate mortgages (ARMs) originated was just above 21,000. Since then, subprime lending has increased substantially, with the number of FRM originations peaking at almost 780,000 and ARM originations peaking (and surpassing FRMs) at over 866,000.[9]

The subprime market took a temporary downturn when the total number of FRM subprime originations declined during the 1998-2000 period; this observation is consistent with our earlier brief history discussion and the downturn in originations reported by Inside Mortgage Finance (2004) and shown in Table 3. Since 2000, however, the subprime market has resumed its momentum. In fact, from 2002 to 2003 the LoanPerformance data show a 62 percent increase and the Inside Mortgage Finance data show a 56 percent increase in originations.

During the late 1990s, house prices increased and interest rates dropped to some of the lowest rates in 40 years, thus providing low-cost access to the equity in homes. Of the total number of subprime loans originated, just over one-half were for cash-out refinancing, whereas more than one-third were for a home purchase (see Figure 4). In 2003, for example, the total number of loans for cash-out refinancing was over 560,000, whereas the number of loans for a home purchase totaled more than 820,000, and loans for no-cash-out refinancing loans amounted to just under 250,000. In the prime market, Freddie Mac estimated that, in 2003, 36 percent of loans for refinancing took at least 5 percent of the loan in cash (downloaded from the Cash-Out Refi Report at www.freddiemac.com/news/finance/data.html on 11/4/04). This estimate is in contrast with typical behavior in the subprime market, which always has had more cash-out refinancing than no-cash-out refinancing.

Given the characteristics of an application, lenders of subprime loans typically identify borrowers and classify them in separate risk categories. Figure 5 exhibits four risk grades, with A– being the least risky and D being the riskiest grade.[10] The majority of the subprime loan origi-

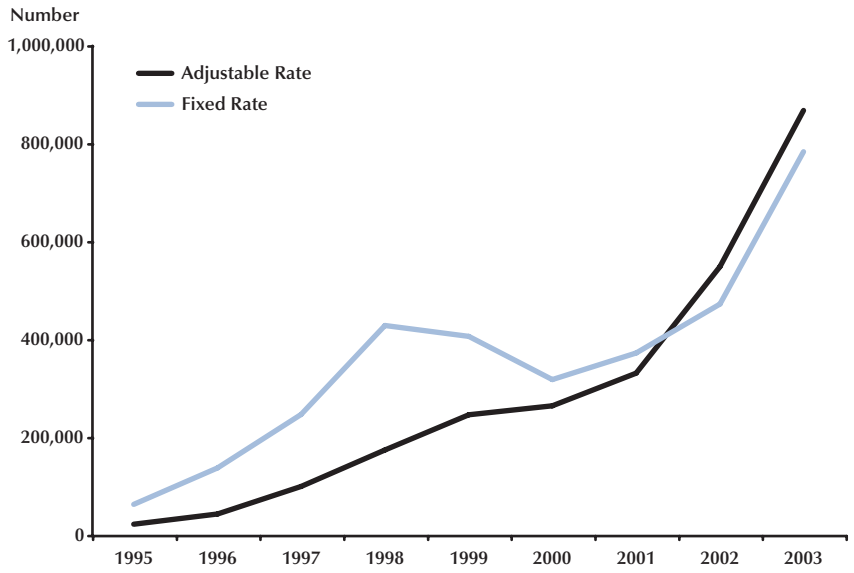
[8] An additional source of information on the subprime market is a list of lenders published by the United States Department of Housing and Urban Development (HUD) Policy Development and Research (PD&R). This list has varied from a low of 51 in 1993 to a high of 256 in 1996; in 2002, the last year available, 183 subprime lenders are identified. The list can then be matched to the Home Mortgage Disclosure Act (HMDA) data set. The list is compiled by examining trade publications and HMDA data analysis. Lenders with high denial rates and a high fraction of home refinances are potential candidates. The lenders are then called to confirm that they specialize in subprime lending. As a result, loans identified as subprime using the HUD list included only firms that specialize in subprime lending (not full-service lenders). As a result, many subprime loans will be excluded and some prime loans will be included in the sample. Very little detail beyond the interest rate of the loan and whether the rate is adjustable is included. For example, the existence of prepayment penalties is unknown—a unique and key feature of subprime lending. Still this lender list has proved useful in characterizing the neighborhood that these loans are originated in. See, for example, Pennington-Cross (2002) and Calem, Gillen, and Wachter (2004).

[9] Similarly, Nichols, Pennington-Cross, and Yezer (2005) note that the share of subprime mortgage lending in the overall mortgage market grew from 0.74 percent in the early 1990s to almost 9 percent by the end of 1990s.

[10] Loan grades are assigned by LoanPerformance and reflect only the rank ordering of any specific firm's classifications. Because these classifications are not uniform, there will be mixing of loan qualities across grades. Therefore, these categories will likely differ from the Countrywide examples used earlier.

Figure 3

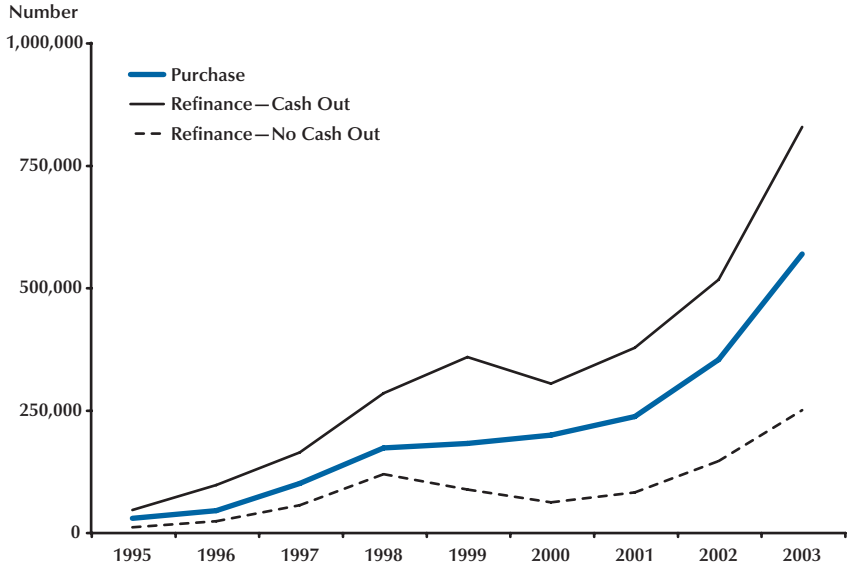
Number of Loans Originated



SOURCE: LoanPerformance ABS securities data base of subprime loans.

Figure 4

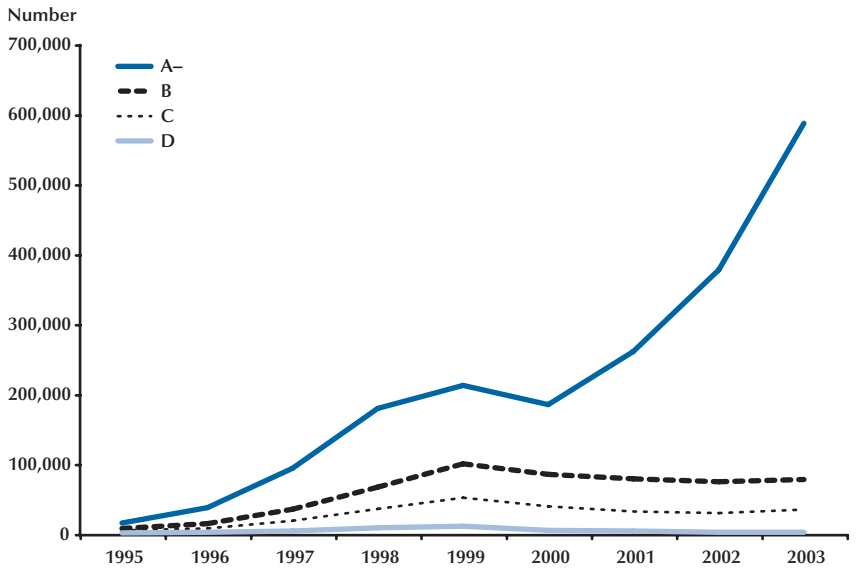
Number of Loans Originated by Purpose



SOURCE: LoanPerformance ABS securities data base of subprime loans.

Figure 5

Number of Loans Originated by Grade



SOURCE: LoanPerformance ABS securities data base of subprime loans.

nations in this data set are classified into the lowest identified risk category (grade A–), particularly after 1998. In addition, the proportion of grade A– loans to the total number of loans has continuously increased from slightly over 50 percent in 1995 to approximately 84 percent in 2003. On the other hand, the shares of grades B, C, and D loans have all declined since 2000. Overall, these observations illustrate that, since 1998-99, the subprime market (or at least the securitized segment of the market) has been expanding in its least-risky segment. It seems likely then that the move toward the A– segment of subprime loans is in reaction to (i) the events of 1998, (ii) the difficulty in correctly pricing the higher-risk segments (B, C, and D credit grades), and, potentially, (iii) changes in the demand for securities for subprime loans in the secondary market.

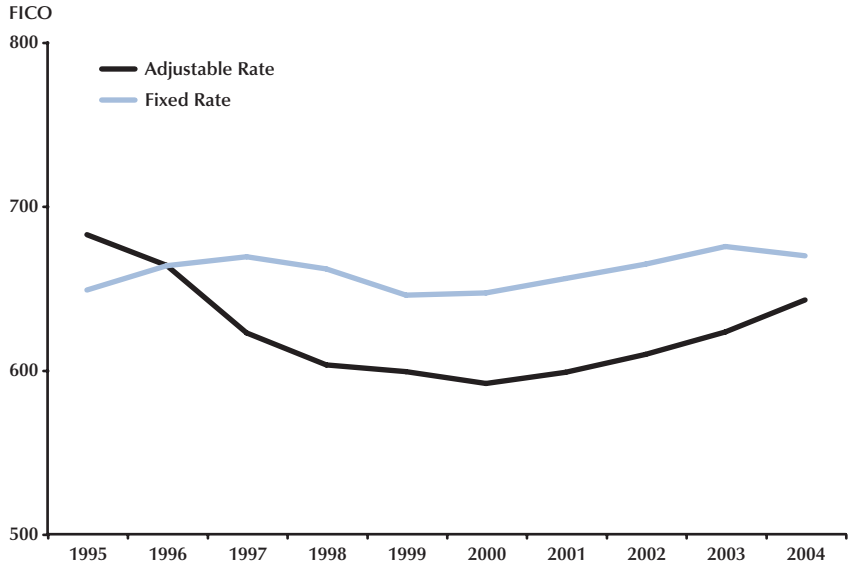
Credit Scores

On average, ARM borrowers have lower credit scores than FRM borrowers (see Figure 6). In 2003, for example, the average FICO (a credit score created by Fair Isaac Corporation to measure consumer credit worthiness) for FRMs is almost 50 points lower than for ARMs (623 versus 675). During the 1990s, average credit scores tended to decline each year, particularly for ARM borrowers; but since 2000, credit scores have tended to improve each year. Hence, it appears that subprime lenders expanded during the 1990s by extending credit to less-credit-worthy borrowers. Subsequently, the lower credit quality unexpectedly instigated higher delinquency and default rates (see also Temkin, Johnson, and Levy, 2002).

With the improved credit quality since 2000, the average FICO has jumped from just under 622 in 2000 to just over 651 in 2004 (closing in on the 669 average conventional FICO reported by Nichols, Pennington-Cross, and Yezer, 2005). As shown in Figure 7, lenders of subprime loans are increasing the number of borrowers with scores in the 500-600 and 700-800 ranges and decreasing the number with scores below 500. Specifically, from 2000 to 2003, the share of borrowers with FICO scores between 700 and 800 rose from approximately 14 percent to 22 percent.

Figure 6

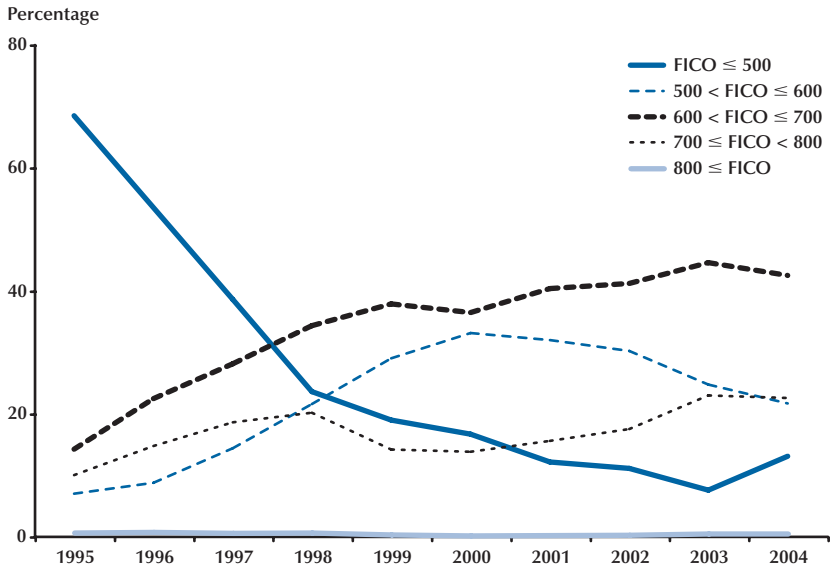
Average Credit Score (FICO)



SOURCE: LoanPerformance ABS securities data base of subprime loans.

Figure 7

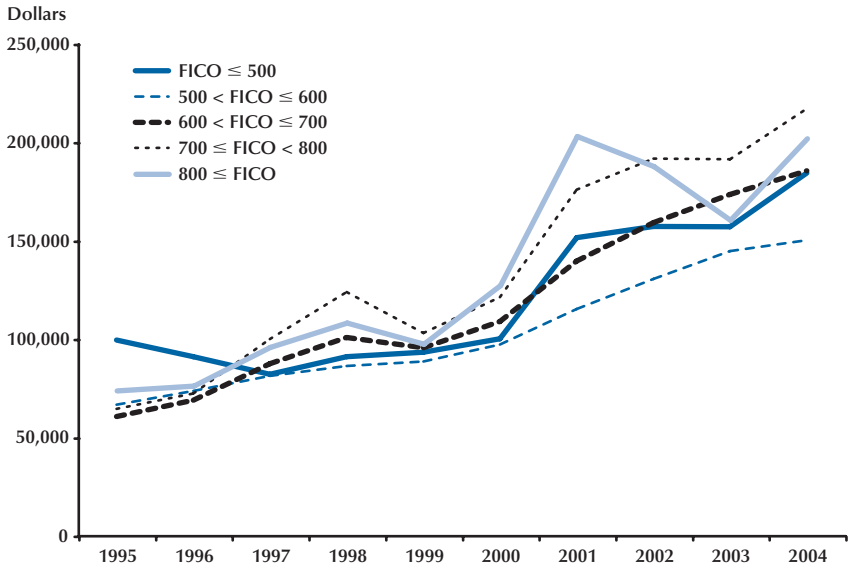
Share of Loans by Credit Score



SOURCE: LoanPerformance ABS securities data base of subprime loans.

Figure 8

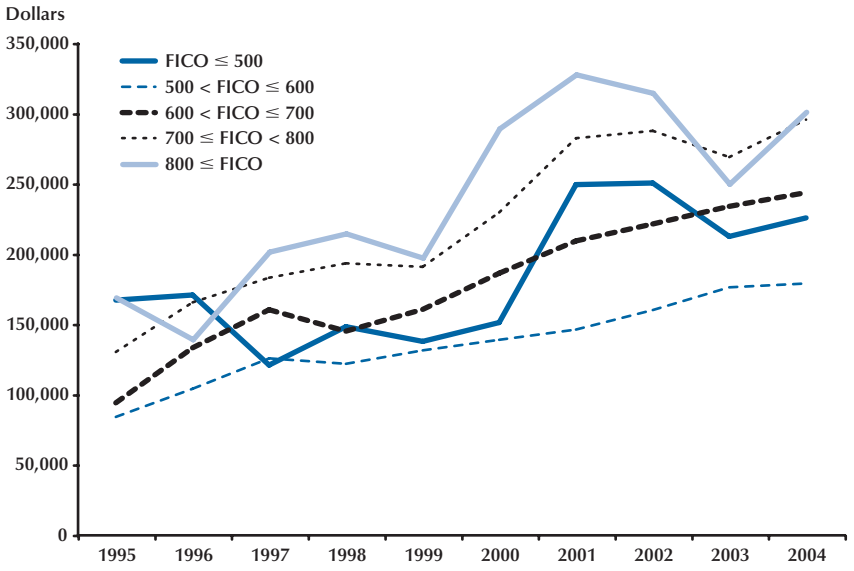
Loan Amounts by Credit Score



SOURCE: LoanPerformance ABS securities data base of subprime loans.

Figure 9

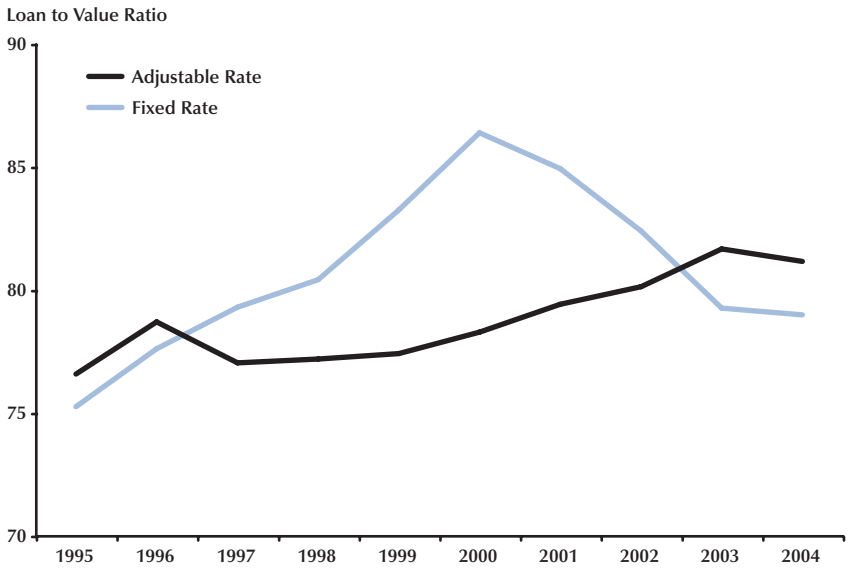
House Prices by Credit Score



SOURCE: LoanPerformance ABS securities data base of subprime loans.

Figure 10

Loan to Value Ratio (LTV)



SOURCE: LoanPerformance ABS securities data base of subprime loans.

Moreover, lenders have on average provided smaller loans to higher-risk borrowers, presumably to limit risk exposure (see Figure 8). As noted previously, these changes in underwriting patterns are consistent with lenders looking for new ways to limit risk exposure. In addition, although loan amounts have increased for all borrowers, the amounts have increased the most, on average, for borrowers with better credit scores. Also, as expected, borrowers with the best credit scores purchased the most expensive houses (see Figure 9).

Down Payment

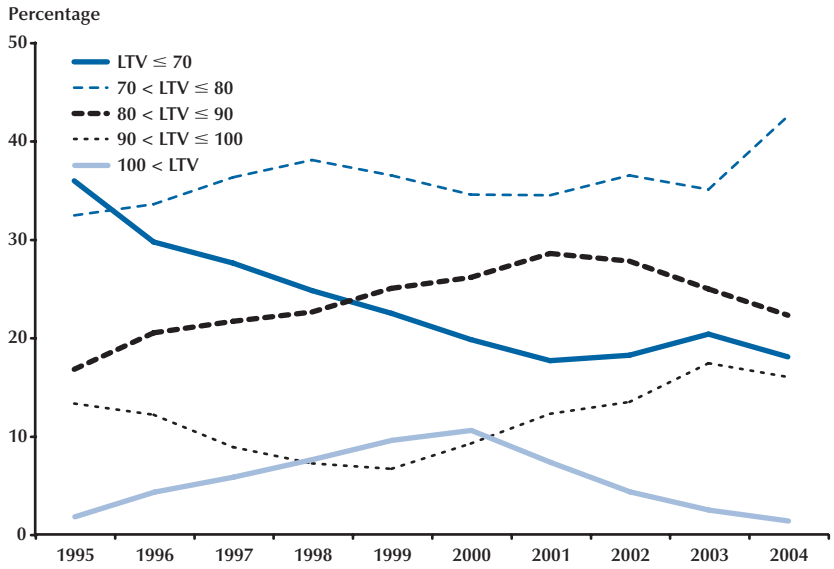
Figure 10 depicts average LTV ratios for subprime loan originations over a 10-year period. The primary finding here is that down payments for FRMs were reduced throughout the 1990s but have increased steadily since. (Note that the change in business strategy occurs just after the 1998 crisis.) In contrast, over the same period, down payments for ARMs were reduced. On first inspection, it may look like lenders are adding more risk by originating more ARMs with higher LTVs; however, this change primarily reflects borrowers with better credit scores and more loans classified as A–. Therefore, this is additional evidence that lenders of subprime loans reacted to the losses sustained in 1998 by moving to less-risky loans—primarily to borrowers with higher credit scores.

As shown in Figure 11, this shift in lending strategy was accomplished by (i) steadily reducing loans with a large down payment (LTV $\leq$ 70), (ii) decreasing loans with negative equity (LTV > 100), and (iii) increasing loans with a 10 percent down payment. Overall, lenders of subprime loans have been increasing loan amounts, shifting the distribution of down payments, and increasing credit score requirements, on average, since 2000.

In general, borrowers with larger down payments tend to purchase more expensive homes (Figure 12). By tying the amount of the loan to the size of the down payment, lenders limit their exposure to credit risk.

Figure 11

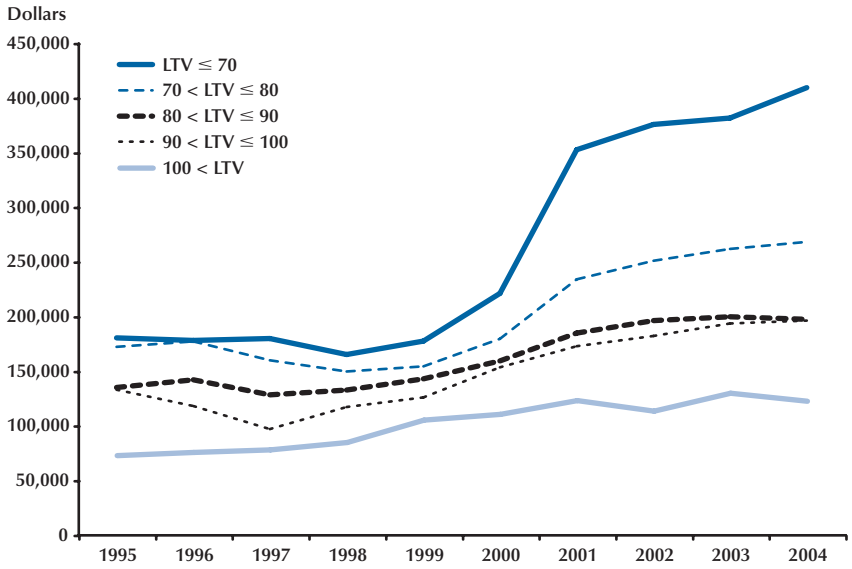
Share of Loans by LTV



SOURCE: LoanPerformance ABS securities data base of subprime loans.

Figure 12

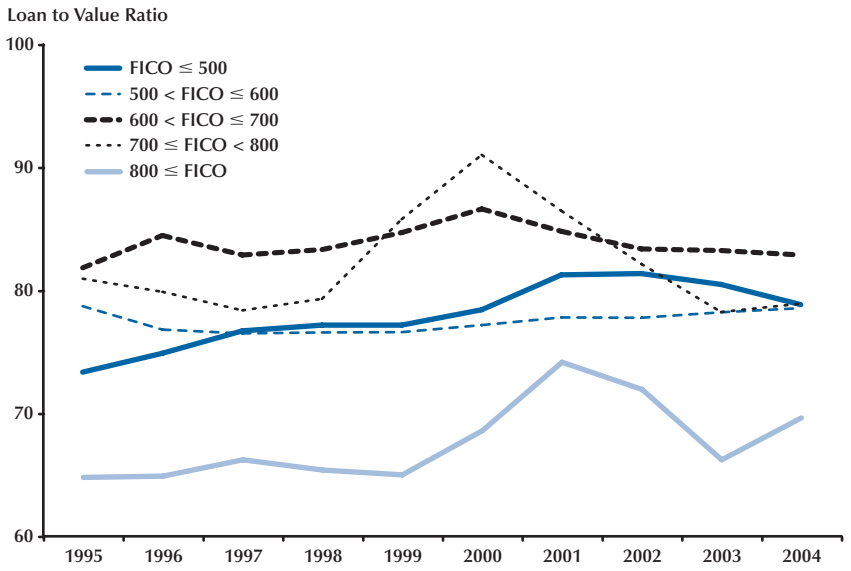
House Prices by LTV



SOURCE: LoanPerformance ABS securities data base of subprime loans.

Figure 13

LTV by Credit Score



SOURCE: LoanPerformance ABS securities data base of subprime loans.

The LTV-FICO Trade-off

In Figure 13, we observe that borrowers with the best credit scores tend to also provide the largest down payments. But, beyond this observation, there seems little correlation between credit scores and down payments.

In contrast, Figure 14 shows a clear ordering of down payments (LTV ratios) by loan grade. Loans in higher loan grades have smaller down payments on average. In fact, over time, especially after 2000, the spread tends to increase. This finding is consistent with the philosophy that loans identified as being more risky must compensate lenders by providing larger down payments. This helps to reduce credit risk associated with trigger events, such as periods of unemployment and changes in household structure, which can make it difficult for borrowers to make timely payments.

Consistent with the loan grade classifications, Figure 15 shows that lower-grade loans have lower credit scores. Therefore, as loans move to better grades, credit scores improve and down payments decrease.

INTEREST RATES

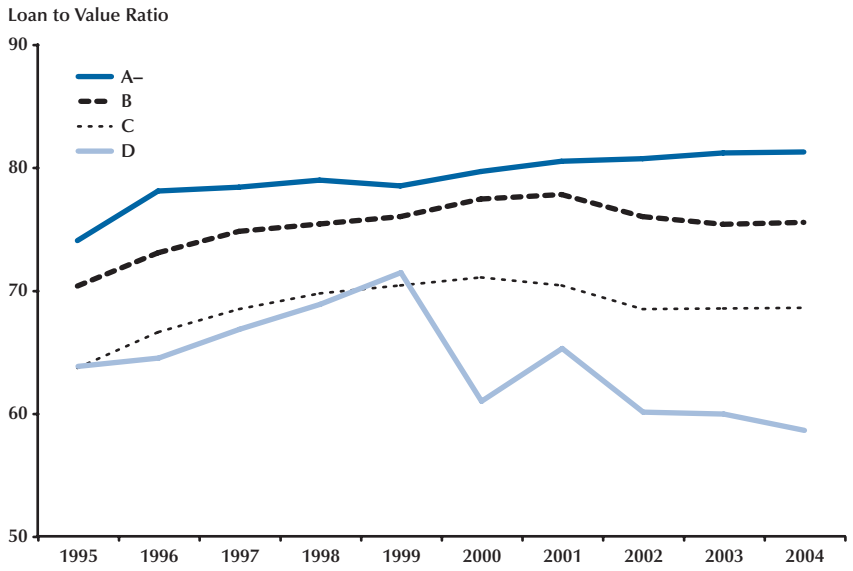
This section examines patterns in the interest rate that borrowers are charged at the origination of the loan. This does not reflect the full cost of borrowing because it does not include any fees and upfront costs that are borne by the borrower. In addition, the borrower can pay extra fees to lower the interest rate, which is called paying points.

Despite these stipulations, we are able to find relationships between the observed interest rates and underwriting characteristics. There is not much difference in the average interest rate (the interest rate on the loan excluding all upfront and continuing fees) at origination for FRMs and ARMs (see Figure 16). But, both product types have experienced a large drop in interest rates, from over 10 percent in 2000 to approximately 7 percent in 2004.

Underwriting standards usually rely heavily on credit history and LTVs to determine the appropriate risk-based price. In Figures 17 and 18 we see evidence of risk-based pricing based on bor-

Figure 14

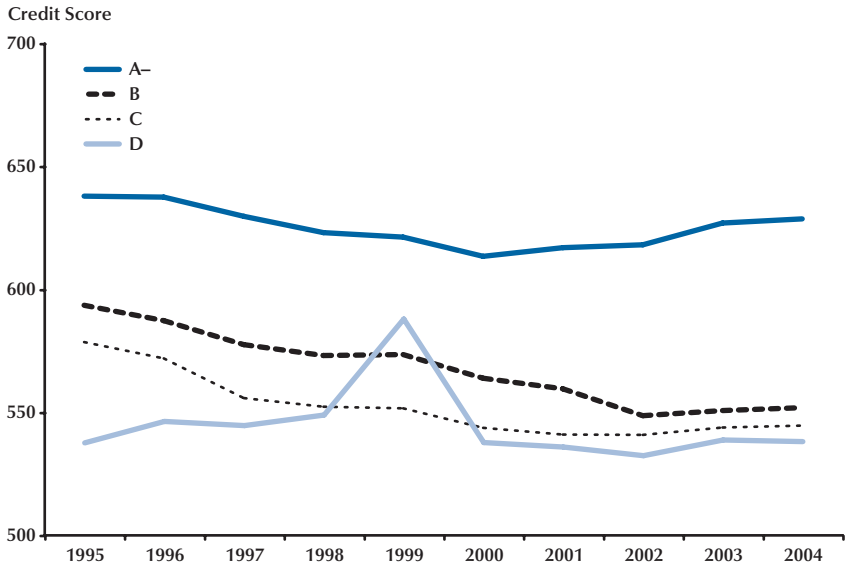
LTV by Loan Grade



SOURCE: LoanPerformance ABS securities data base of subprime loans.

Figure 15

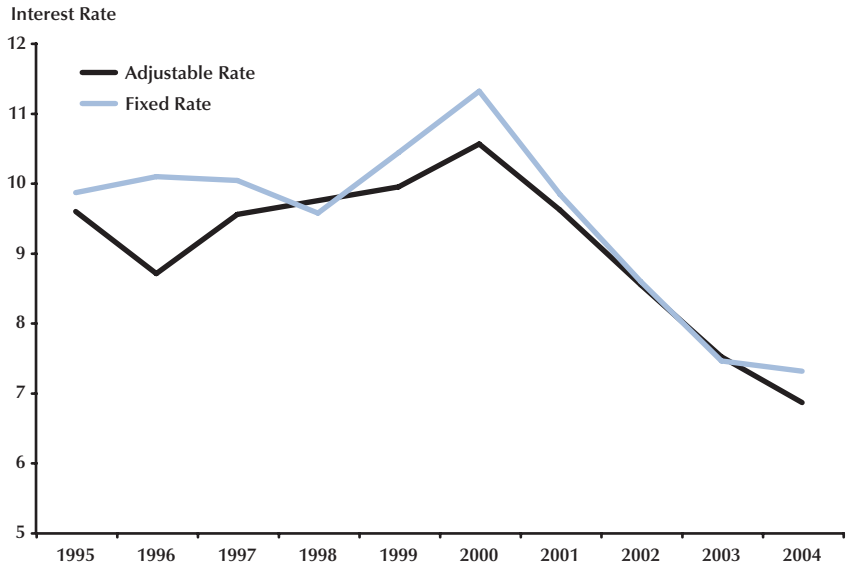
Credit Score by Loan Grade



SOURCE: LoanPerformance ABS securities data base of subprime loans.

Figure 16

Interest Rates



SOURCE: LoanPerformance ABS securities data base of subprime loans.

rower credit scores and, to some small extent, on borrower down payments. For example, borrowers with the highest FICO scores tend to receive a lower interest rate. In 2004, average interest rates vary by over 2 percentage points from the highest to the lowest FICO scores.

This range of interest rates does not hold when pricing is based solely on down payments. In fact, the striking result from Figure 18 is that, on average, the pricing of subprime loans is very similar for all down-payment sizes, except for loans with LTVs greater than 100, which pay a substantial premium. One way to interpret these results is that lenders have found good mechanisms to compensate for the risks of smaller down payments and, as a result, down payments in themselves do not lead to higher borrower costs. However, if the equity in the home is negative, no sufficient compensating factor can typically be found to reduce expected losses to maintain pricing parity. The borrower has a financial incentive to default on the loan because the loan amount is larger than the value of the home. As a result, the lender must increase the interest rate to decrease its loss if a default occurs.

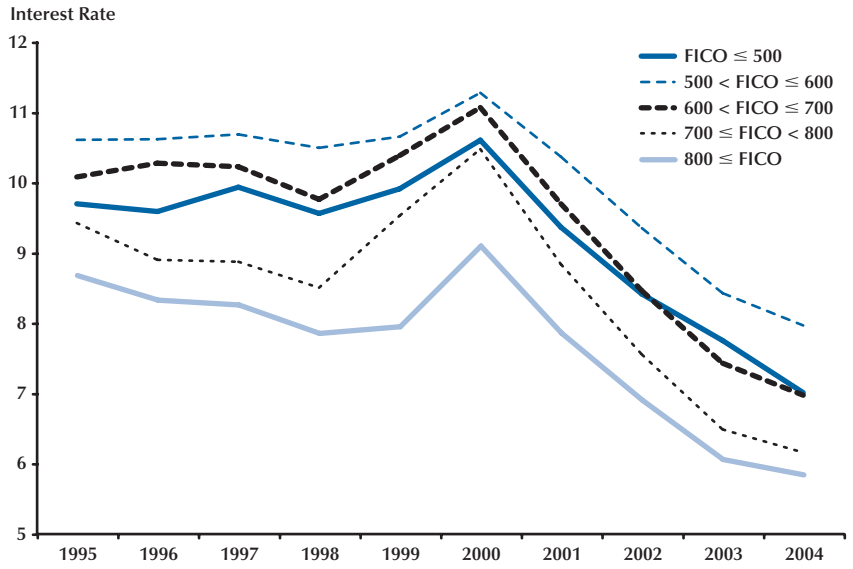
Figure 19 shows the average interest rate by loan grade. The riskiest borrowers (Grade D) receive the highest interest rate, whereas the least-risky borrowers (Grade A–) receive the lowest interest rate. Interestingly, although interest rates overall changed dramatically, the spread between the rates by grade have remained nearly constant after 1999. This may indicate that the risks, and hence the need for risk premiums, are in levels, not proportions, across risk grades.

Prepayment Penalties

It is beyond the scope of this paper to define specific examples of predatory lending, but prepayment penalties have been associated with predatory practices. A joint report by the U.S. Department of Housing and Urban Development (HUD) and the U.S. Department of Treasury (Treasury) (2002) defined predatory lending as lending that strips home equity and places borrowers at an increased risk of foreclosure. The

Figure 17

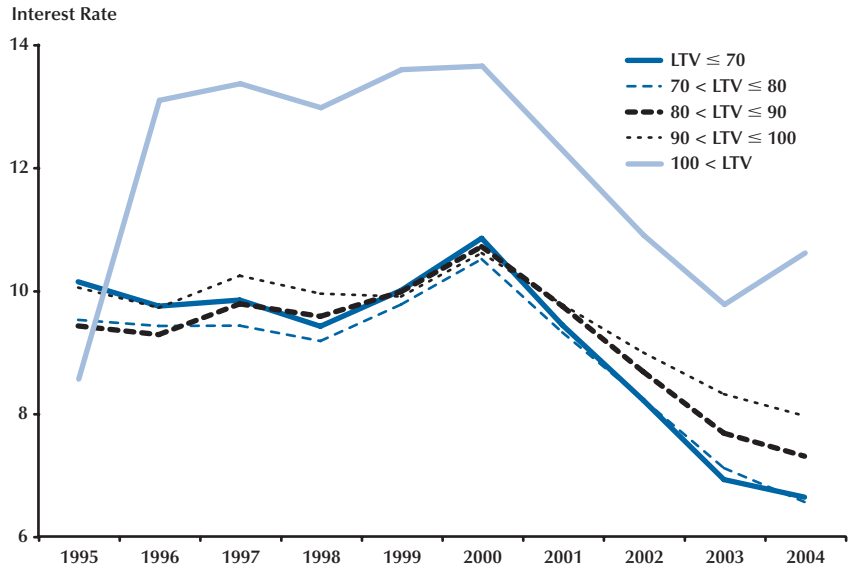
Interest Rates by Credit Score



SOURCE: LoanPerformance ABS securities data base of subprime loans.

Figure 18

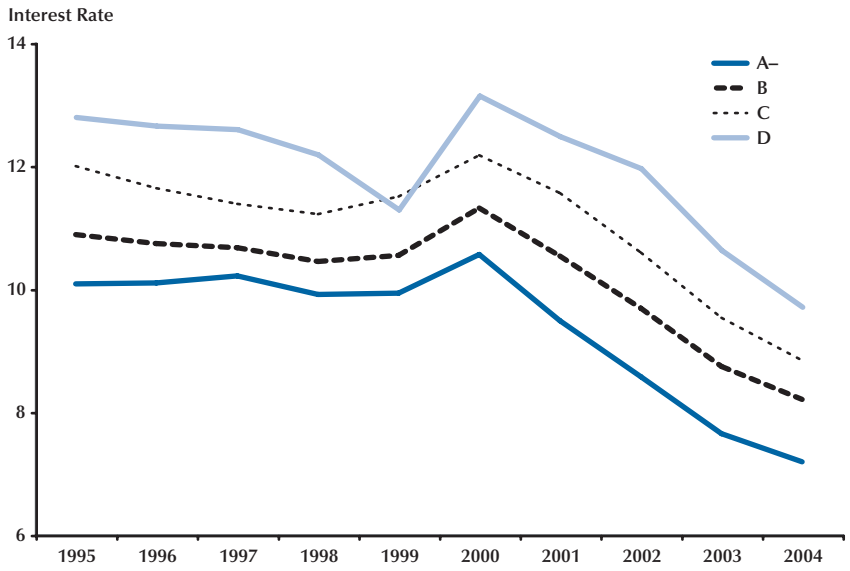
Interest Rates by LTV



SOURCE: LoanPerformance ABS securities data base of subprime loans.

Figure 19

Interest Rates by Loan Grade



SOURCE: LoanPerformance ABS securities data base of subprime loans.

characteristics include excessive interest rates and fees, the use of single-premium credit life insurance, and prepayment penalties that provide no compensating benefit, such as a lower interest rate or reduced fees. In addition, some public interest groups such as the Center for Responsible Lending believe that prepayment penalties are in their very nature predatory because they reduce borrower access to lower rates (Goldstein and Son, 2003).

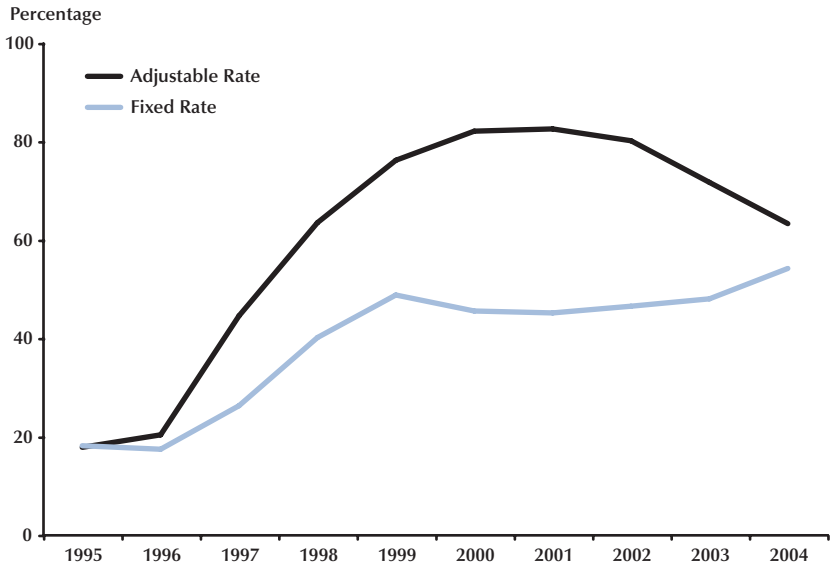
Both Fannie Mae and Freddie Mac changed their lending standards to prohibit loans (i.e., they will not purchase them) that include some types of prepayment penalties. On October 1, 2002, Freddie Mac no longer allowed the purchase of subprime loans with a prepayment penalty after three years. However, loans originated before that date would not be affected by the restriction (see www.freddiemac.com/singlefamily/ppmqanda.html downloaded on 2/14/05). If a subprime loan stipulates a prepayment penalty, Fannie Mae will consider the loan for purchase only if (i) the borrower receives a reduced interest rate or reduced fees, (ii) the borrower is provided an alternative mortgage choice, (iii) the nature of the penalty is disclosed to the borrower, and (iv) the penalty cannot be charged if the borrower defaults on the loan and the note is accelerated (www.fanniemae.com/newsreleases/2000/0710.jhtml).[11] Therefore, we may expect to see a decline in the use of prepayment penalties starting in 2000 and 2002, at least in part due to changes in the demand for subprime securities.

Despite these concerns, prepayment penalties have become a very important part of the subprime market. When interest rates are declining or steady, subprime loans tend to be prepaid at elevated rates compared with prime loans (Pennington-Cross, 2003, and UBS Warburg, 2002). In addition, subprime loans tend to default at elevated rates. As a result, the expected life of an average subprime loan is much shorter than that

[11] When a borrower defaults, the lender typically will send an acceleration note informing the borrower that the mortgage contract has been violated and all of the remaining balance and fees on the loan are due immediately.

Figure 20

Share of Loans with a Prepayment Penalty



SOURCE: LoanPerformance ABS securities data base of subprime loans.

of a prime loan. Therefore, there are fewer good (nonterminated) loans to generate income for an investor to compensate for terminated (defaulted and prepayed) loans. One mechanism to reduce the break-even price on these fast-terminating loans is to use prepayment penalties (Fortowsky and LaCour-Little, 2002). Although this same mechanism is used in the prime market, it is not as prevalent.

Figure 20 shows that, prior to 2000, the use of prepayment penalties grew quickly. Substantially more ARMs than FRMs face a prepayment penalty. For loans originated in 2000-02, approximately 80 percent of ARMs were subject to a prepayment penalty compared with approximately 45 percent of FRMs. Equally important, the share of ARMs and FRMs subject to a prepayment penalty rose dramatically from 1995 to 2000. In fact, at the end of the five-year period, ARMs were five times more likely and FRMs twice as likely to have prepayment penalties.

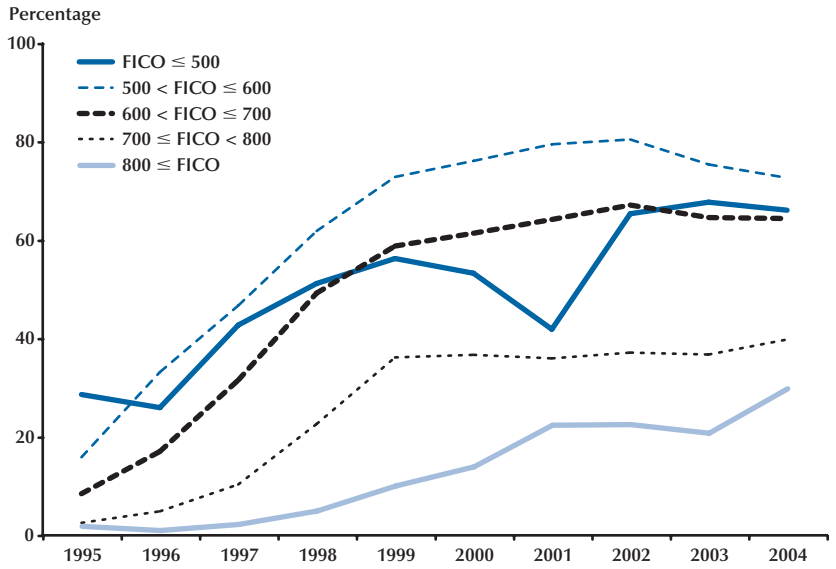
This rapid increase can at least partially be attributable to regulatory changes in the interpretation of the 1982 AMTPA by the Office of Thrift and Supervision (OTS). Before 1996, the OTS interpreted AMTPA as allowing states to restrict finance companies (which make many of the subprime loans) from using prepayment penalties, but the OTS exempted regulated federal depository institutions from these restrictions. In 1996, the OTS also allowed finance companies the same exemption. However, this position was short lived and the OTS returned to its prior interpretation in 2002.

In 2003 and 2004, prepayment penalties declined for ARMs and held steady for FRMs. This was likely caused by (i) the introduction of predatory lending laws in many states and cities (typically these include ceilings on interest rates and upfront fees, restrictions on prepayment penalties, and other factors)[12]; (ii) the evolving position of Fannie Mae and Freddie Mac on pre-

[12] For more details on predatory lending laws that are both pending and in force, the MBAA has a "Predatory Lending Law Resource Center" available at www.mbaa.org/resources/predlend/ and the Law Offices of Herman Thordsen also provide detailed summaries of predatory laws at www.lendinglaw.com/predlendlaw.htm.

Figure 21

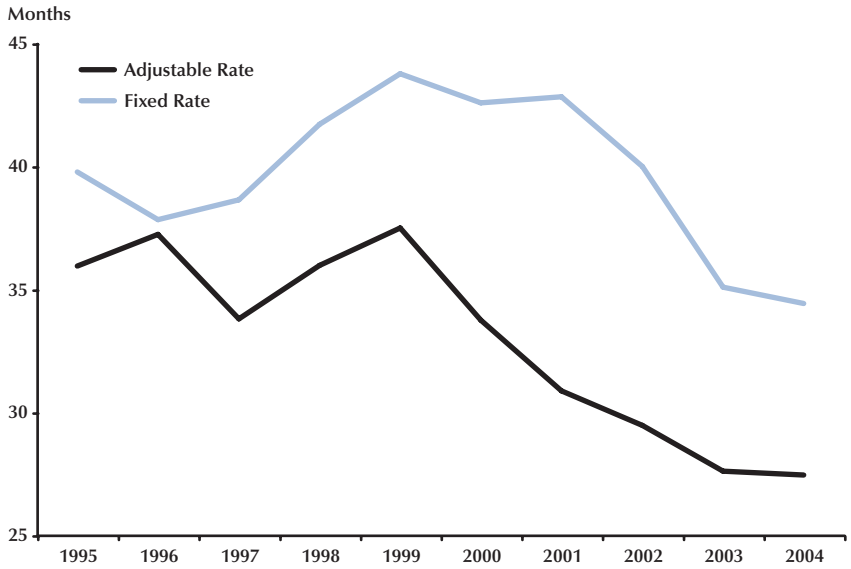
Share of Loans with a Prepayment Penalty by Credit Score



SOURCE: LoanPerformance ABS securities data base of subprime loans.

Figure 22

Length of Prepayment Penalty



SOURCE: LoanPerformance ABS securities data base of subprime loans.

payment penalties; and (iii) the reversed OTS interpretation of AMTPA in 2002 (see 67 Federal Register 60542, September 26, 2002), which again made state laws apply to finance companies just as they had prior to 1996.

The share of loans containing a prepayment penalty is lowest among borrowers with the highest, or best, FICO scores (see Figure 21). In 2003, for instance, about 20 percent of borrowers with a FICO score above 800 were subject to a prepayment penalty, whereas over 60 percent of borrowers with a FICO score below 700 faced such a penalty.

To understand the prevalence of these penalties, one must know how long prepayment penalties last. Figure 22 shows that the length of the penalty has generally been declining since 2000. Again, the introduction and threat of predatory lending laws and Freddie Mac purchase requirements (that the term of a prepayment penalty be no more than three years) is likely playing a role in this trend. In addition, FRMs tend to have much longer prepayment penalties. For example, in 2003, the average penalty lasted for almost three years for FRMs and a little over two years for ARMs, both of which meet current Freddie Mac guidelines.

CONCLUSION

As the subprime market has evolved over the past decade, it has experienced two distinct periods. The first period, from the mid-1990s through 1998-99, is characterized by rapid growth, with much of the growth in the most-risky segments of the market (B and lower grades). In the second period, 2000 through 2004, volume again grew rapidly as the market became increasingly dominated by the least-risky loan classification (A– grade loans). In particular, the subprime market has shifted its focus since 2000 by providing loans to borrowers with higher credit scores, allowing larger loan amounts, and lowering the down payments for FRMs. Furthermore, the subprime market had reduced its risk exposure by limiting the loan amount of higher-risk loans and imposing prepayment penalties on the majority of ARMs and low credit-score loans. The use of prepayment penalties has declined in the past few years because the securities market has adjusted to public concern about predatory lending and the regulation of finance companies has changed.

The evidence also shows that the subprime market has provided a substantial amount of risk-based pricing in the mortgage market by varying the interest rate of a loan based on the borrower's credit history and down payment. In general, we find that lenders of subprime loans typically require larger down payments to compensate for the higher risk of lower-grade loans. However, even with these compensating factors, borrowers with low credit scores still pay the largest premiums.

REFERENCES

Calem, Paul; Gillen, Kevin and Wachter, Susan. "The Neighborhood Distribution of Subprime Mortgage Lending." *Journal of Real Estate Finance and Economics*, 2004, *29*(4), pp. 393-410.

Capozza, Dennis R. and Thomson, Thomas A. "Subprime Transitions: Long Journey into Foreclosure." Presented at the American Real Estate and Urban Economics Annual Meeting, Philadelphia, PA, January 2005.

Fortowsky, Elaine B. and LaCour-Little, Michael. "An Analytical Approach to Explaining the Subprime-Prime Mortgage Spread." Presented at the Georgetown University Credit Research Center Symposium *Subprime Lending*, 2002.

Goldstein, Debbie and Son, Stacey Strohauer. "Why Prepayment Penalties are Abusive in Subprime Home Loans." Center for Responsible Lending Policy Paper No. 4, April 2, 2003.

Hillier, Amy E. "Spatial Analysis of Historical Redlining: A Methodological Exploration." *Journal of Housing Research*, November 2003, *14*(1), pp. 137-67.

Immergluck, Daniel and Wiles, Marti. *Two Steps Back: The Dual Mortgage Market, Predatory Lending, and the Undoing of Community Development*. Chicago: The Woodstock Institute, 1999.

Inside Mortgage Finance. *The 2004 Mortgage Market Statistical Annual*. Washington, DC: 2004.

Nichols, Joseph; Pennington-Cross, Anthony and Yezer, Anthony. "Borrower Self-Selection, Underwriting Costs, and Subprime Mortgage Credit Supply." *Journal of Real Estate Finance and Economics*, March 2005, *30*(2), pp. 197-219.

Pennington-Cross, Anthony. "Subprime Lending in the Primary and Secondary Markets." *Journal of Housing Research*, 2002, *13*(1), pp. 31-50.

Pennington-Cross, Anthony. "Credit History and the Performance of Prime and Nonprime Mortgages." *Journal of Real Estate Finance and Economics*, November 2003, *27*(3), pp. 279-301.

Pennington-Cross, Anthony. "The Value of Foreclosed Property." *Journal of Real Estate Research* (forthcoming).

Temkin, Kenneth; Johnson, Jennifer E.H. and Levy, Diane. *Subprime Markets, the Role of GSEs, and Risk-Based Pricing*. Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research, March 2002.

Tracy, Joseph; Schneider, Henry and Chan, Sewin. "Are Stocks Over-Taking Real Estate in Household Portfolios?" *Current Issues in Economics and Finance*, Federal Reserve Bank of New York, April 1999, *5*(5).

UBS Warburg. "Credit Refis, Credit Curing, and the Spectrum of Mortgage Rates." *UBS Warburg Mortgage Strategist*, May 21, 2002, pp. 15-27.

U.S. Department of Housing and Urban Development and U.S. Department of Treasury, National Predatory Lending Task Force. *Curbing Predatory Home Mortgage Lending*. Washington, DC: 2002.

Are the Causes of Bank Distress Changing? Can Researchers Keep Up?

Thomas B. King, Daniel A. Nuxoll, and Timothy J. Yeager

Since 1990, the banking sector has experienced enormous legislative, technological, and financial changes, yet research into the causes of bank distress has slowed. One consequence is that traditional supervisory surveillance models may not capture important risks inherent in the current banking environment. After reviewing the history of these models, the authors provide empirical evidence that the characteristics of failing banks have changed in the past ten years and argue that the time is right for new research that employs new empirical techniques. In particular, dynamic models that use forward-looking variables and address various types of bank risk individually are promising lines of inquiry. Supervisory agencies have begun to move in these directions, and the authors describe several examples of this new generation of early-warning models that are not yet widely known among academic banking economists.

Federal Reserve Bank of St. Louis *Review*, January/February 2006, *88*(1), pp. 57-80.

Understanding the causes of insolvency at financial institutions is important for both academic and regulatory reasons, and the effort to model bank deterioration was once a vibrant area of study in empirical finance. Significant advances were made between the late 1960s and late 1980s. Since then, research has slowed considerably on the characteristics of banks headed for trouble, reflecting a sense among researchers that the causes of banking problems are unchanging and well understood. In this article, we argue that this complacency may be unwarranted.[1] The rapid pace of technological and institutional change in the banking sector in recent years suggests that the dominant models may no longer accurately represent the nature of bank deterioration. Indeed, the few observations that we have of recent bank failures provide evidence consistent with this hypothesis.

The changes in the banking environment call for renewed research into the causes of bank distress. The federal supervisory agencies have established research programs pursuing this goal, but—because regulatory banking economists often work on projects with confidential data and because many ongoing projects are not formally disclosed to the public—it can be difficult for outside economists to benefit from this work. By describing some efforts that are currently underway to develop new early-warning models at the Federal Reserve and Federal Deposit Insurance Corporation (FDIC), we attempt to bridge that gap in the hope of stimulating more research in this area beyond that done by government agencies. One strand of the new monitoring devices attempts

[1] Note that we are not claiming that bank regulators have grown complacent, only that the academic community has focused its attention away from this issue.

Thomas B. King is an economist at the Federal Reserve Bank of St. Louis; Daniel A. Nuxoll is an economist at the Division of Insurance and Research, Federal Deposit Insurance Corporation; and Timothy J. Yeager is the Arkansas Bankers' Association Chair of Banking at the University of Arkansas. (Yeager was an assistant vice president and economist at the Federal Reserve Bank of St. Louis at the time this article was written.)
The authors thank Alton Gilbert, Hui Guo, Andy Meyer, Greg Sierra, and David Wheelock for helpful comments. The views expressed are those of the authors and are not necessarily official positions of the Federal Reserve Bank of St. Louis, the Board of Governors of the Federal Reserve System, or the FDIC.

© 2006, The Federal Reserve Bank of St. Louis. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

to complement traditional early-warning models by adopting a more theoretical approach using forward-looking variables. Another strand isolates and models unique banking risks to facilitate the risk-focused approach to bank supervision. A common objective of these models is an increased flexibility that will allow off-site surveillance to better keep pace with the dynamic banking environment going forward.

SURVEILLANCE MODELS IN HISTORICAL CONTEXT

Federal bank supervisors primarily use limited-dependent-variable regression models for off-site monitoring. Although we argue later that these models (like all models) have shortcomings, they reflect years of advancement in academic research, econometric modeling, and computer technology. In this section, we describe the evolution of off-site surveillance models, paying particular attention to the link between academic research and supervisory applications. Table 1 summarizes the evolution of various off-site surveillance systems at the Federal supervisory agencies from the mid-1970s to the present. The systems transitioned from simple screens, to hybrid models, to the econometric models used today.

Discriminant Analysis and Supervisory Screens

During the 1960s, several studies attempted to determine the usefulness of various financial ratios in predicting bankruptcy in non-bank firms. In his seminal article, Altman (1968) used discriminant analysis over five variables to determine the characteristics of manufacturing firms headed for bankruptcy. His paper ushered in a wave of research applying similar methodology specifically to depository institutions, including Stuhr and van Wicklen (1974), Sinkey (1975, 1978), Altman (1977), and Rose and Scott (1978).

Much of this early research on bank distress was conducted by economists within supervisory agencies, and some of it was specifically directed toward the establishment of an off-site early-warning model for use in everyday supervision. Because discrete-response-regression techniques were still relatively new and too computationally intensive to be practical, the initial screen-based systems adopted by all three federal agencies relied on a variant of discriminant analysis, comparing selected ratios to predetermined cutoff points and classifying banks accordingly.

The Office of the Comptroller of the Currency (OCC) adopted the first formal screen-based system called the National Bank Surveillance System (NBSS) in 1975. Previously, off-site monitoring had consisted largely of informal rules of thumb based on individual financial ratios. According to White (1992), the impetus for the shift toward a more systematic approach was the OCC's failure to detect the financial difficulties at two large institutions—United States National Bank and Franklin National Bank—that became insolvent in the early 1970s. The OCC's response to these shortcomings in off-site surveillance was, in part, to avail itself of new computing technology to condense the call-report data into key financial ratios for each bank under its supervision. One component of the NBSS, the Anomaly Severity Ranking System, ranked selected bank ratios within peer groups to detect outliers.

The FDIC and the Federal Reserve quickly followed the OCC with similar screen-based models of their own. In 1977, the FDIC introduced the Integrated Monitoring System. One component of this system was the humbly titled "Just A Warning System," which consisted of 12 financial ratios. The system compared each ratio with a benchmark ratio determined by examiner judgment. Banks with ratios that "failed" various screens were flagged for additional follow-up. The Federal Reserve adopted the Minimum Bank Surveillance System (later, the Uniform Bank Surveillance Screen), which examined seven bank ratios. These ratios were weighted by their Z-scores, which were then summed to yield a composite score for each bank. MBSS, which resulted from the research program described in Korobow, Stuhr, and Martin (1977), was the first surveillance model adopted by a supervisory body to employ formal statistical techniques.

Discrete-Response Models and Hybrid Systems

The development of discrete-response regression techniques, together with the increased availability of the computing power necessary to apply them to large datasets, aided the advancement of bank-distress models beginning in the late 1970s (Hanweck, 1977, Korobow, Stuhr, and Martin, 1977, and Martin, 1977). Because of its analytical simplicity, the logistic specification has been the favorite model of this type, although arctangent and probit models have also appeared occasionally.[2] As pointed out by Martin (1977), discriminant analysis can be viewed as a special case of logistic regression in that the existence of a unique linear discriminant function implies the existence of a unique logit equation, whereas the converse is not true. However, the existence of a linear discriminant function is commonly rejected when the number of observations of one class is substantially smaller than that in the other class. For this reason, early discriminant studies typically used subsamples of the population of safe banks (which have always far outnumbered risky banks by any measure), either matching them according to certain non-risk characteristics or randomly selecting the control sample. The use of a logit model obviates the need for these restrictive sampling methods.

Martin's (1977) study set the standard for discrete-response models of bank-failure prediction. Whereas most previous research had focused on a small sample of banks over two or three years, Martin used all Fed-supervised institutions during a seven-year period in the 1970s, yielding over 33,000 observations. In what would become a standard approach, he confronted the data agnostically with 25 financial ratios and ran several different specifications in search of the best fit. He found that capital ratios, liquidity measures, and profitability were the most significant determinants of failure over his sample period. Although Martin did not employ direct measures of asset quality, his indirect measures—provision expense and loan concentration—also turned out to be significant.

A host of other studies around the same time, using both logit and discriminant analysis, confirmed these basic results. Table 2 summarizes a selection of these papers. Poor asset quality and low capital ratios are the two characteristics of banks that have most consistently been associated with banking problems over time (Sinkey, 1978). Indeed, as described in Putnam (1983), early-warning research in the 1970s and 1980s displayed a remarkable consistency in the variables that emerged as important predictors of banking problems: profitability, capital, asset quality, and liquidity appeared as statistically significant in almost every study, even though they were often measured using different ratios.[3]

Motivated in part by the consistency of the pattern of bank deterioration, the federal banking agencies adopted the Uniform Financial Rating System in November 1979.[4] Under this system—which is still the primary rating mechanism for U.S. bank supervision—capital adequacy (C), asset quality (A), management competence (M), earnings performance (E), and liquidity risk (L) are each explicitly evaluated by examiners and rated on a 1 (best) to 5 (worst) scale. (Beginning in 1997, sensitivity to market risk (S) was adopted as a sixth component.) Examiners also assign a composite rating (CAMELS) on the same scale, reflecting the overall safety and soundness of the institution.

From a supervisory perspective, modeling CAMELS ratings allows examiners to observe estimates of current supervisory ratings on a quarterly basis, rather than only during an on-site exam. The availability of consistent supervisory-rating data beginning in 1979 allowed researchers to employ ordered logit techniques to estimate bank ratings. (See West, 1985; and Whalen and

[2] Linear regression analysis was explored early on by Meyer and Pifer (1970).

[3] More recently, some research has investigated the potential for local and regional economic data to add information about future banking conditions. However, the results have largely rejected this idea (e.g., Meyer and Yeager, 2001; Nuxoll, O'Keefe, and Samolyk, 2003; and Yeager, 2004). On the other hand, Neely and Wheelock (1997) show that bank *earnings* are highly correlated with state-level personal-income growth.

[4] Prior to 1979, the three federal regulatory agencies assigned banks scores for capital (1 to 4), asset quality (A to D), and management (S, F, or P), as well as a composite score (1 to 4).

Table 1

Evolution of Key Off-Site Surveillance Systems

Screen-Based Systems	Agency	Period used
National Bank Surveillance System (NBSS)	**OCC**	**1975 to ?**
Condensed the call-report data into key financial ratios and compared them to peer ratios. One output of the NBSS, the Anomaly Severity Ranking System, ranked bank ratios by peer group to detect outliers. Another output was the Bank Performance Report. In cooperation with the Fed and FDIC, the OCC transformed the Bank Performance Report into the Uniform Bank Performance Report (UBPR). Although the OCC no longer uses the NBSS, the UBPR is used presently by all federal and state supervisory agencies for both on-site and off-site analysis.		
Minimum Bank Surveillance Screen (MBSS)	**Federal Reserve**	**Late 1970s to mid-80s**
Employed a set of ratios as off-site screens and added institutions that lay outside a critical range to an "exception list" that received extra scrutiny. A composite score was also constructed by summing the normalized values of seven of these ratios.		
Integrated Monitoring System (IMS)	**FDIC**	**1977 to 1985**
A screening device within the IMS, called the "Just A Warning System" (JAWS), compared 12 key financial ratios to critical values as determined by examiner expertise. JAWS did not compute composite scores or make direct comparisons to peer levels.		
Uniform Bank Surveillance Screen (UBSS)	**Federal Reserve**	**Mid-1980s to 1993**
Improvement upon the MBSS. Computed peer-group percentiles of six financial ratios and summed them to derive the composite score. Banks in the highest percentiles of the composite score were placed on a watch list.		

Hybrid Systems	Agency	Period used
CAEL	**FDIC**	**1985 to late 1998**
Replaced IMS. An "expert system," designed to replicate the financial analysis that an examiner would perform to assign an examination rating. Ratios were chosen to evaluate capital (C), asset quality (A), earnings (E), and liquidity (L). Analysts subjectively determined the weights for each of the ratios that fed into the four CAEL components. The CAEL components were multiplied by their respective weights and summed to yield a composite CAEL score.		
Canary	**OCC**	**2000 to present**
Canary consists of a package of tools organized into four components: Benchmarks, Credit Scope, Market Barometers, and Predictive Models. Benchmarks are screen-based ratios that indicate risky thresholds. The Peer Group Risk Model is a predictive model that projects a bank's return on assets over the next three years under various economic scenarios.		

Table 1, cont'd

Limited-Dependent Variable Systems	Agency	Period used
System to Estimate Examination Ratings (SEER)	**Federal Reserve**	**1993 to present**
Replaced the UBSS. First named the Financial Institutions Monitoring System, SEER is a logit model that consists of two components, a "risk-rank" model that forecasts bank-failure probabilities and a "rating" model that estimates current CAMELS scores.		
Statistical CAMELS Off-site Rating (SCOR)	**FDIC**	**1998 to present**
Replaced CAEL. Like SEER, the model consists of two components: a CAMELS downgrade forecast and a rating forecast. The downgrade forecast computes the probability that a 1- or 2-rated bank will receive a 3, 4, or 5 rating at the next examination. The OCC also uses output from the SCOR model in off-site surveillance.		
CAMELS Downgrade Probability (CDP)	**Federal Reserve Bank of St. Louis**	**1999 to present**
Similar to the downgrade forecast of SCOR, the CDP estimates the probability that a 1- or 2-rated bank will be downgraded to a 3, 4, or 5 rating over the next two years.		

Forward-Looking Early-Warning Systems	Agency	Period used
Growth Monitoring Sytem (GMS)	**FDIC**	**2000 to present**
Although GMS was initially developed as an expert system and implemented in the 1980s, it was revised significantly in the late 1990s to employ explicit statistical techniques. GMS is a logit model of downgrades that estimates which institutions that are currently rated satisfactory are most likely to be classified as problem banks at the end of three years. Rather than using credit quality measures as independent variables, GMS includes forward-looking variables such as loan growth and noncore funding that can be precursors of problems that have yet to manifest themselves.		
Liquidity and Asset Growth Screen (LAGS)	**Federal Reserve Bank of St. Louis**	**2002 to present**
LAGS is conceptually similar to GMS, but it uses a dynamic vector autoregression approach to forecast the set of banks most likely to exploit moral hazard-incentives. Such banks exhibit rapid loan growth, increasing dependence on funding sources with no market discipline, and declining capital ratios. Like GMS, the model uses forward-looking variables.		

Risk-Focused Systems	Agency	Period used
Real Estate Stress Test (REST)	**FDIC**	**2000 to present**
REST attempts to identify those banks and thrifts that are most vulnerable to problems in real estate markets by subjecting them to the same stress as the New England real estate crisis of the early 1990s. Forecast measures of bank performance are translated to CAMELS ratings using the SCOR model. The result is a REST rating that ranges from 1 to 5.		
Economic Value Model (EVM)	**Federal Reserve**	**1998 to present**
The EVM is a duration-based economic value of equity model that estimates the loss in a bank's market value of equity given an instantaneous 200-basis-point interest rate increase. The model is useful to assess the bank's long-run sensitivity to interest rate risk.		

Table 2

Comparison of Selected Early Studies Predicting Bank Condition

Model number	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable	Failure	Rating	Failure	Failure	Failure	Rating	Failure	Rating
Technique	OLS	Discriminant analysis	Logit	Probit	Probit	Factor + Logit	Logit	Factor + Logit
No. of observations	60	214	33,627	221	820	~5,700	339	70
Sample period	1948-65	1967-68	1969-76	1971-76	1980-83	1980-82	1983-84	1983-86
Loans vs. securities mix		X	X	X	X	X	X	
Efficiency, net operating expense, or overhead	X	X	X	X	X	X		X
ROA or ROE	X	X	X			X	X	X
Capital/assets		X	X	X		X	X	X
Classified loans	X	X			X	X		
Loan mix	X		X			X	X	
Size		X		X		X		
Charge-offs			X		X			
Deposit mix	X					X		
Past-due or nonperforming loans					X			X
Liquid assets			X			X		
Volatile liabilities or jumbo CDs						X		X
Dividend payout ratio		X						
Interest income, expense, or margin						X		
Interest-rate sensitivity					X			
Provision expense			X					X
Insider activity	X							
Income volatility	X							
Balance sheet volatility	X							
Asset or loan growth	X							
Income growth				X				
Loan-loss reserves								X
Other	X						X	X

NOTE: Variables listed in the table are those included in each study. In most cases, variables were selected because of their significance, and so the table also largely reflects variables that were significant in predicting bank problems. In some studies, some additional variables were considered but they do not receive an "X" in the table because they were found to be statistically insignificant. The studies referenced are (1) Meyer and Pifer, 1970; (2) Stuhr and van Wicklen, 1974; (3) Martin, 1977; (4) Hanweck, 1977; (5) Bovenzi, Marino, and McFadden, 1983; (6) West, 1985; (7) Pantalone and Platt, 1987; and (8) Whalen and Thomson, 1988.

Thomson, 1988.)[5] Cole and Gunther (1998) demonstrate that actual supervisory ratings can become obsolete within as little as six months after being assigned. Similarly, Hirtle and Lopez (1999) find that the private supervisory information contained in these ratings decays as they age. These studies suggest that early-warning models that estimate current supervisory ratings are useful tools for supervisors to keep up with bank fundamentals without incurring the cost of an examination.[6]

The FDIC's CAEL model, introduced in 1985, represented a significant breakthrough in off-site monitoring devices. This "hybrid" system—a discrete-response framework coupled with examiner input—estimated ratings for four of the five CAMEL components based on quarterly call-report data. ('M' was not estimated.) For each CAEL component, experienced examiners subjectively weighted the relevant bank ratios; a rating table then mapped the model output to a rating ranging from 1 to 5. The rating table was updated each quarter to mirror the actual distribution of component CAMEL ratings in the previous year. CAEL then weighted the four estimated components themselves to yield a composite rating. In essence, the model was a calibrated limited-dependent-variable model, with examiner guidance replacing the computationally intensive econometric procedure.

The Current Surveillance Regime

A wealth of data on bank failures and CAMELS ratings throughout the 1980s and the rapid pace of computer technology in the 1980s and early 1990s allowed supervisory agencies to "catch up" with the banking and econometric research and develop off-site monitoring devices employing limited-dependent-variable econometric techniques. Table 3 compares the explanatory variables used in select previous and current early-warning systems. Two systems—SEER and SCOR—are the primary surveillance tools used today by the Fed and the FDIC, respectively.

In 1993, the Federal Reserve adopted as its in-house early-warning model the Financial Institutions Monitoring System, which was modified slightly and renamed the System to Estimate Examination Ratings (SEER). This model consists of two components: a "risk-rank" or failure model that estimates bank-failure probabilities and a "rating" model that estimates current CAMELS scores. The SEER failure model is designed to detect deficiencies in balance sheet and income statement ratios that are severe enough to cause an outright failure or a critical shortfall in capital. Because these events have been rare since the inception of SEER, the variables and coefficient estimates have remained frozen since they were first estimated on late-1980s and early-1990s failures. The SEER rating model, in contrast, is reestimated on a quarterly basis, allowing for different coefficient estimates—and indeed different independent variables—in each quarter. This model has the advantage of allowing for new sources of bank risk, but it can be difficult to interpret changes in risk when the main driver of the change is the inclusion of a variable that was not present in the model in the previous quarter. The two models are used together to achieve a balance between flexibility and consistency. As Cole, Cornyn, and Gunther (1995), Cole and Gunther (1998), and Gilbert, Meyer, Vaughan (1999) demonstrate, SEER's performance is superior to a variety of other early-warning systems, including actual CAMELS scores assigned by examiners, in terms of the trade-off between its type-I and type-II error.[7]

[5] West (1985) and Wang and Sauerhaft (1989) model supervisory ratings in a factor-analytic framework. Supervisory ratings had previously been used to measure composite risk in a discriminant-analysis study by Stuhr and van Wicklen (1974). Two other, related, lines of research begun in this period involve modeling time to failure (rather than failure probability) and regulatory closure-decision rules. Examples of the time-to-failure models, which typically involve Cox (1972) proportional-hazard specifications, can be found in Lane, Looney, and Wansley (1986), Whalen (1991), Helwege (1996), and Wheelock and Wilson (1995, 2000, 2005). For models of supervisory closure behavior, see Barth et al. (1989), Demirgüç-Kunt (1989), Thomson (1992), and Cole (1993).

[6] It is important to recognize that these models are intended as complements to, rather than substitutes for, on-site examination. Although CAMELS ratings do become stale rather quickly, Nuxoll, O'Keefe, and Samolyk (2003) and Wheelock and Wilson (2005) show that they still retain marginal predictive power for failures, beyond that contained in the call-report data. Thus, on-site examination appears to recover some information that is not available in bank financial statements.

[7] In this case, a type-I error occurs when a bank is not predicted to fail but does. A type-II error occurs when a bank is predicted to fail but does not. For obvious reasons, regulators are more concerned with type-I errors.

Table 3

Comparison of Early-Warning Systems

	JAWS	UBSS	CAEL	SEER	SCOR	Downgrade	GMS	LAGS
Agency	FDIC	FRB	FDIC	FRB	FDIC	FRB	FDIC	FRB
Model type	Screens	Screens	Hybrid	Logit	Logit	Logit	Logit	VAR
Tier-1 or tangible capital	X	X	X	X	X	X		X
Total or risk-weighted assets				X		X		X
Past due 30			X	X	X	X		
Past due 90		X	X	X	X	X		
Nonaccruals			X	X	X	X		
OREO			X	X	X	X		
Residential real estate loans	X			X		X		
C&I loans				X		X		
Securities	X			X		X		
Jumbo CDs	X			X		X		
Net Income (ROA)	X	X	X	X	X	X		
Charge-offs			X		X			
Provision expense	X		X		X			
Liquid assets	X	X			X			
Loan growth	X						X	
Total or risk-weighted asset growth		X					X	X
Volatile liability expense		X						
Volatile liabilities			X		X		X	
Loan-loss reserves			X		X			
Loan/deposit ratio	X		X					
Interest expense	X							
Loans and long-term securities					X		X	
NCNRP funding								X
Operating expenses or revenues	X		X					
Change in capital	X		X				X	
Change in deposits	X							
Dividends	X							
Region								
Prior composite supervisory rating							X	X
Prior supervisory management rating								

NOTE: For purposes of comparison, some liberties have been taken with variable definitions, e.g., such categories as liquid assets and tangible capital have been defined in slightly different ways in the various models and the construction of certain ratios differs slightly.

In 1998, the FDIC developed a model similar to SEER, known as the Statistical CAMELS Offsite Rating (SCOR). The SCOR model, which replaced CAEL, also consists of two components: a rating forecast and a CAMELS-downgrade forecast.[8] The rating component of the FDIC's SCOR model is similar to the SEER rating model. SCOR uses a multinomial logit model to estimate a composite CAMELS rating as well as ratings for all six of the CAMELS components, in keeping with the formulation of the preceding CAEL system. SCOR's downgrade component estimates probabilities that safe banks (those with ratings of 1 or 2) will receive ratings of 3, 4, or 5 at the next examination.

The Federal Reserve has recently undertaken a similar effort in modeling downgrades. Gilbert, Meyer, and Vaughan (2002) use a logistic model to estimate downgrade probabilities for CAMELS composites. The authors concluded that the variables included in SEER were also the most appropriate for their purposes; but one advantage of the CAMELS downgrade model relative to the SEER failure model is the ability to update the coefficients on a periodic basis.

In sum, researchers and practitioners have made considerable progress in developing models to predict bank distress. However, as we discuss below, these models must be complemented with newer models to account for evolution in the banking industry and nontraditional sources of bank risk.

THE NEED FOR NEW WORK

The sophistication of off-site early-warning systems since 1970 has certainly improved; but, given the dramatic changes in the banking sector over the past decade, we may expect that the current systems—like the screen-based mechanisms that preceded them—have already fallen behind the pace of financial evolution.[9] The main criticism of prevailing early-warning techniques is the implicit assumption that future episodes of bank distress will look similar to past episodes of distress. However, significant changes in the banking environment since 1990 combined with empirical evidence that bank-distress patterns may be changing suggest that new early-warning research is needed.

Recent Changes in the Banking Environment

Shifts in the banking environment erode confidence in early-warning models because the future is less likely to reflect the past. Since 1990, banks have faced significant legislative, financial, and technological innovations.

The post-1990 legislation, summarized in Table 4, intended to impose more market discipline on banks and remove anti-competitive barriers. The Federal Deposit Insurance Corporation Improvement Act of 1991 and the National Depositor Preference Act of 1993 shifted more of the burden of bank failure from taxpayers to uninsured creditors. Several studies have documented the changes in market discipline that appear to have been caused by this legislation (Flannery and Sorescu, 1996; Cornett, Mehran, and Tehranian, 1998; Marino and Bennett, 1999; Hall et al., 2002; Goldberg and Hudgins, 2002; Flannery and Rangan, 2003; and King, 2005). In addition, legislation removed geographic branching restrictions (Riegle-Neal Act of 1994) and product restrictions (Financial Services Modernization Act of 1999). Many banks have expanded into investment banking, insurance, and other financial services, and a small but increasing fraction of bank revenue derives from fee income generated by these operations (Yeager, Yeager, and Harshman, 2005). A likely outcome of these legislative changes is a more competitive banking industry that has the ability to assume different kinds of credit risk than it assumed in the past.

In addition to the legislative changes, financial markets have widened and deepened, presenting banks with new asset and liability management opportunities and challenges. Previously illiquid assets have become more liquid as secondary markets have developed and government-sponsored enterprises such as Fannie Mae and Freddie Mac have facilitated the growth of the mortgage market. Many of these products, however, contain embed-

[8] See Cole, Cornyn, and Gunther (1995) and Collier et al. (2003a).

[9] Hooks (1995) and Helwege (1996) provide evidence on the parameter instability of traditional early-warning models over time.

Table 4

Key Legislative Changes in the 1990s

Financial Institutions Reform, Recovery, and Enforcement Act of 1989 (FIRREA)
Opened FHLB membership to commercial banks. Previously membership had been available only to thrifts and certain insurance companies. Advances from the FHLB are a ready source of non-risk-priced funding. Over two-thirds of all banks are now FHLB members, and over half of them routinely utilize advances. As Stojanovic, Vaughan, and Yeager (2001) show, risky banks are more likely to rely on advances than safer banks.

Federal Deposit Insurance Corporation Improvement Act of 1991 (FDICIA)
Restricted regulatory forbearance and creditor protection through prompt corrective action and least-cost-resolution provisions. This legislation may have induced greater discipline in uninsured credit markets (see Goldberg and Hudgins, 2002 and Hall et al., 2002), resulting in higher funding costs and different liability structures for troubled institutions. Mandatory closure rules potentially increased the mean and reduced the variance of the capital levels of failing banks.

National Depositor Preference (1993)
Enacted as part of the Omnibus Budget Reconciliation Act of 1993, this legislation changed the failure-resolution hierarchy to make domestic depositors more senior claimants than foreign depositors. Like FDICIA, this legislation may have changed funding costs for risky banks and caused them to rearrange their liability structures. See Marino and Bennet (1999).

Reigle-Neal Interstate Banking and Branching Efficiency Act of 1994
Allowed bank branching across state lines. Although this Act allowed for greater geographic diversification, it also exposed banks to increased competition.

The Gramm-Leach-Bliley Act of 1999 (Financial Services Modernization Act)
Repealed the Glass-Steagal Act and allowed financial holding companies to engage in insurance, securities underwriting and brokerage services, and merchant banking. This Act introduced new potential sources of risk in banking, although it facilitated the diversification of some traditional sources of risk.

ded options that could increase exposure to interest rate risk. Liabilities have also evolved since 1990. Banks are relying increasingly on noncore funding such as brokered deposits and jumbo CDs (over $100,000) as traditional checking and savings accounts and local CDs are shrinking. In addition, the Federal Home Loan Bank (FHLB) opened its doors to commercial banks in 1989, quickly becoming an important nondeposit source of funding. (See Stojanovic, Vaughan, and Yeager, 2001; Bennett et al., 2005; and Craig and Thomson, 2003.) These changes potentially alter both interest rate and liquidity risks. Derivatives usage at commercial banks has also exploded—the notional amount of derivatives at commercial banks increased tenfold to more than $70 trillion between 1991 and 2003. Derivatives can be used to hedge risk, but they can also be used to speculate on market movements.[10] In addition, over-the-counter derivatives potentially expose banks to counterparty risk.

Finally, as in many other industries, technological innovations revolutionized the business of banking in the 1990s. Electronic payments, online banking, and credit scoring are now common and quickly growing activities. As Claessens, Glaessner, and Klingebiel (2002) argue, these developments have the potential to change the competitive landscape dramatically. They also allow for increased operational risk, including data theft from security vulnerabilities and the facilitation of money laundering.

Overall, the new products and markets that have become available to banks in the past decade provide opportunities to diversify and hedge risk in new ways. Yet they also carry dangers—if they are not fully understood or properly managed, new business lines may end up increasing risks

[10] The literature on the risk effects of derivative use is large. Recent contributions include Instefjord (2005), Duffee and Zhou (2001), and Sinkey and Carter (2000).

Table 5

Trends at Failed Banks, Before and After 1995

		Comparison of ratios at failed banks			Comparison of ratios at failed banks less peer values		
Variable	Quarters prior to failure	1995-2003 (%)	1984-94 (%)	Difference of means (%) (*t*-statistic)	1995-2003 (%)	1984-94 (%)	Difference of means (%) (*t*-statistic)
Jumbo CDs	1	14.70	18.80	–4.10** (–2.06)	4.10	9.30	–5.2*** (–2.65)
	6	13.40	21.30	–7.90*** (–5.04)	3.60	12.10	–8.5*** (–5.46)
Federal funds purchased	1	0.37	0.99	–0.62*** (–3.30)	–1.15	–0.20	–0.95*** (–5.07)
	6	0.77	1.29	–0.52 (–1.64)	–0.69	0.17	–0.86*** (–2.72)
Demand deposits	1	12.70	14.90	–2.20 (–1.23)	0.70	–4.80	5.5*** (–3.12)
	6	11.70	15.20	–3.5** (–1.99)	–0.40	–6.30	5.8*** (–3.29)
Loan-loss reserves/loans	1	4.04	3.14	0.90** (–2.15)	2.51	1.90	0.61 (–1.46)
	6	2.63	1.87	0.76*** (–3.7)	1.06	0.67	0.38* (–1.86)
Cash & due	1	7.11	8.20	–1.08 (–1.28)	1.81	–0.45	2.26*** (–2.67)
	6	6.14	9.03	–2.89*** (–3.7)	0.85	0.17	0.68 (–0.87)
Commercial real estate loans	1	15.80	11.60	4.1** (–2.16)	–0.10	3.10	–3.3* (–1.70)
	6	15.80	11.60	4.2** (–2.36)	1.10	3.50	–2.50 (–1.40)
Fee income	1	2.57	1.11	1.46** (–2.44)	1.58	0.34	1.24** (–2.07)
	6	2.87	1.00	1.86 (–1.59)	1.91	0.29	1.62 (–1.38)
OREO	1	1.70	3.48	–1.78*** (–3.78)	1.54	3.11	–1.57*** (–3.33)
	6	1.49	1.70	–0.22 (–0.55)	1.30	1.40	–0.10 (–0.25)
Total assets	1	$133M	$161M	–28M (–0.57)	–$88M	$47M	–$135M*** (–2.72)
	6	$137M	$192M	–$55M (–0.88)	–$51M	$88M	–$139M** (–2.23)

NOTE: This table shows differences in means for selected risk variables between failing banks in the period 1995-2003 compared with those in 1984-94. Both the differences in levels and the differences in levels less peer values for the corresponding periods are given, at both 1- and 6-quarter horizons. *,**, and *** indicate statistical significance at the 10, 5, and 1 percent levels, respectively. All of the nine variables reported here have exhibited significant changes since 1995 (by at least one of these difference-of-means tests) in their patterns as failure approaches.

for banks that move into them too hastily. With the increasing intensity of competition, many institutions have likely been tempted to do exactly that. The net effect on banks' risk positions is an empirical question.

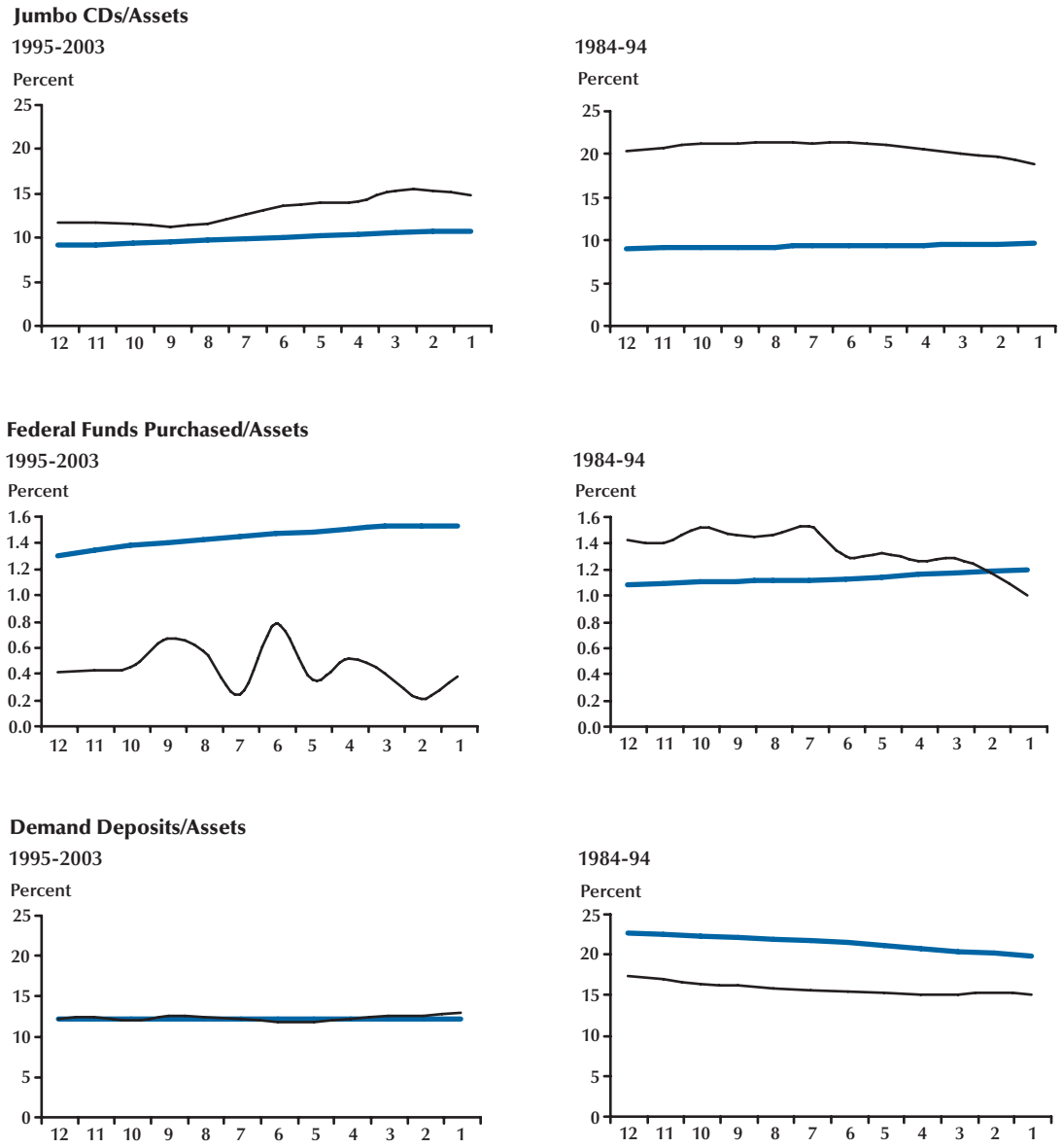
Evidence of Changes in the Nature of Bank Distress

Although none of the institutional changes mentioned above necessarily implies any fundamental change in the process through which banks deteriorate, together they constitute a prima facie case that, at the very least, the previous results should be reaffirmed. Simple empirical analysis indicates that some of the above changes may indeed have had an impact on the typical pattern of bank distress. Figure 1 plots nine key ratio averages for failing banks in the 12 quarters leading to failure between 1984 and 1994 and between 1995 and 2003 against the contemporaneous averages for banks that did not fail.[11] Of course, the number of failures in the earlier period was much larger—1,371 compared with 44—yet the patterns that emerge suggest that many characteristics of banks in the quarters before failure may have changed between the two time periods. Table 5, which reports difference-of-means tests

[11] The December 1994 cutoff was chosen to exclude the failures of the early-1990s banking crisis from the more recent sample. Other break dates around the same time yield similar results.

Figure 1

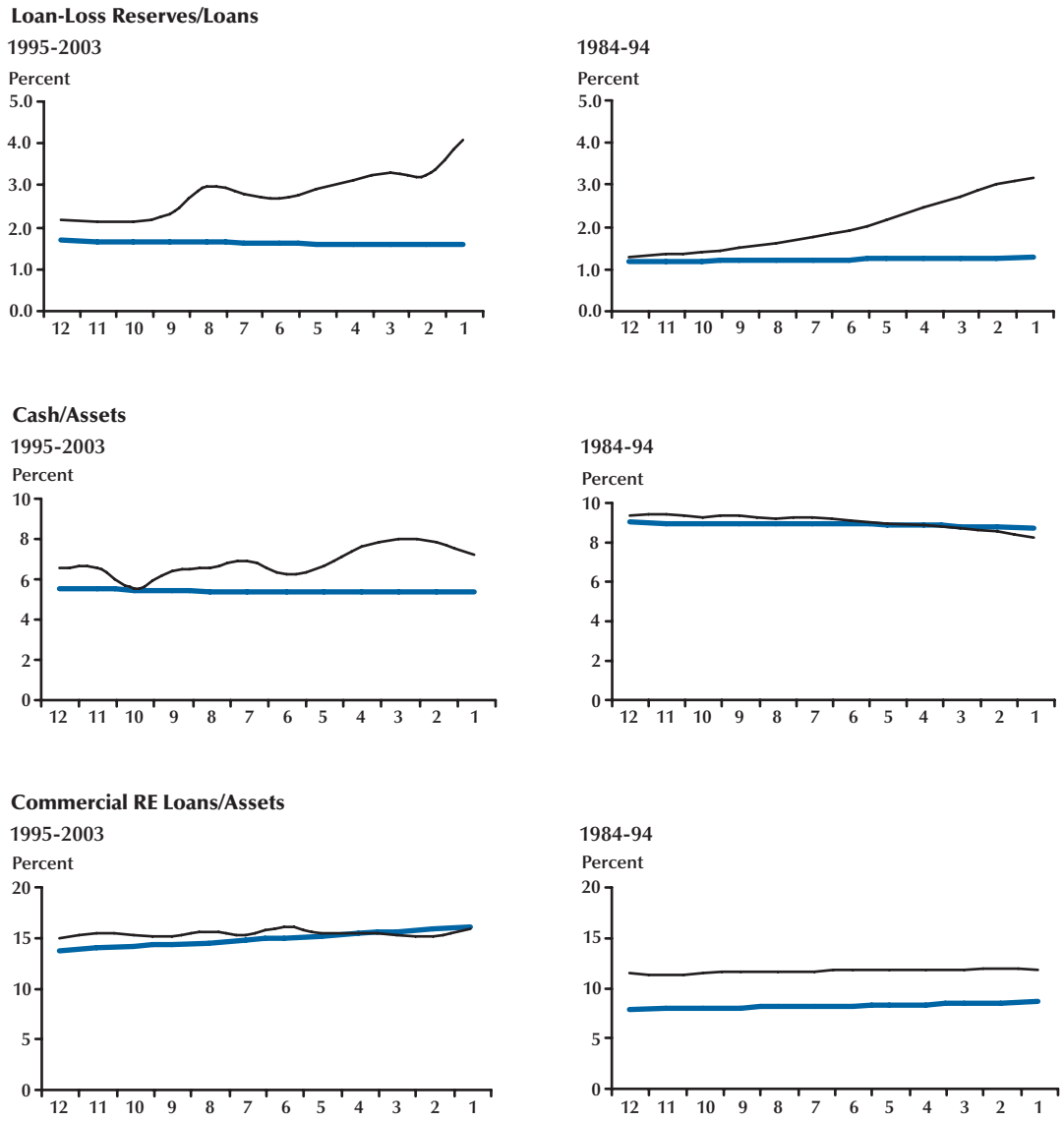
Trends at Failed Banks, Before and After 1995



NOTE: This figure presents the information in Table 5 in graphical form. In each case, the thin black line indicates the path of a failing bank as the failure date approaches and the thick blue line indicates the average values for non-failing banks. Values on the horizontal axis indicate the number of quarters prior to failure. For every variable reported here, there is an obvious change in the pattern between the two periods.

Figure 1, cont'd

Trends at Failed Banks, Before and After 1995

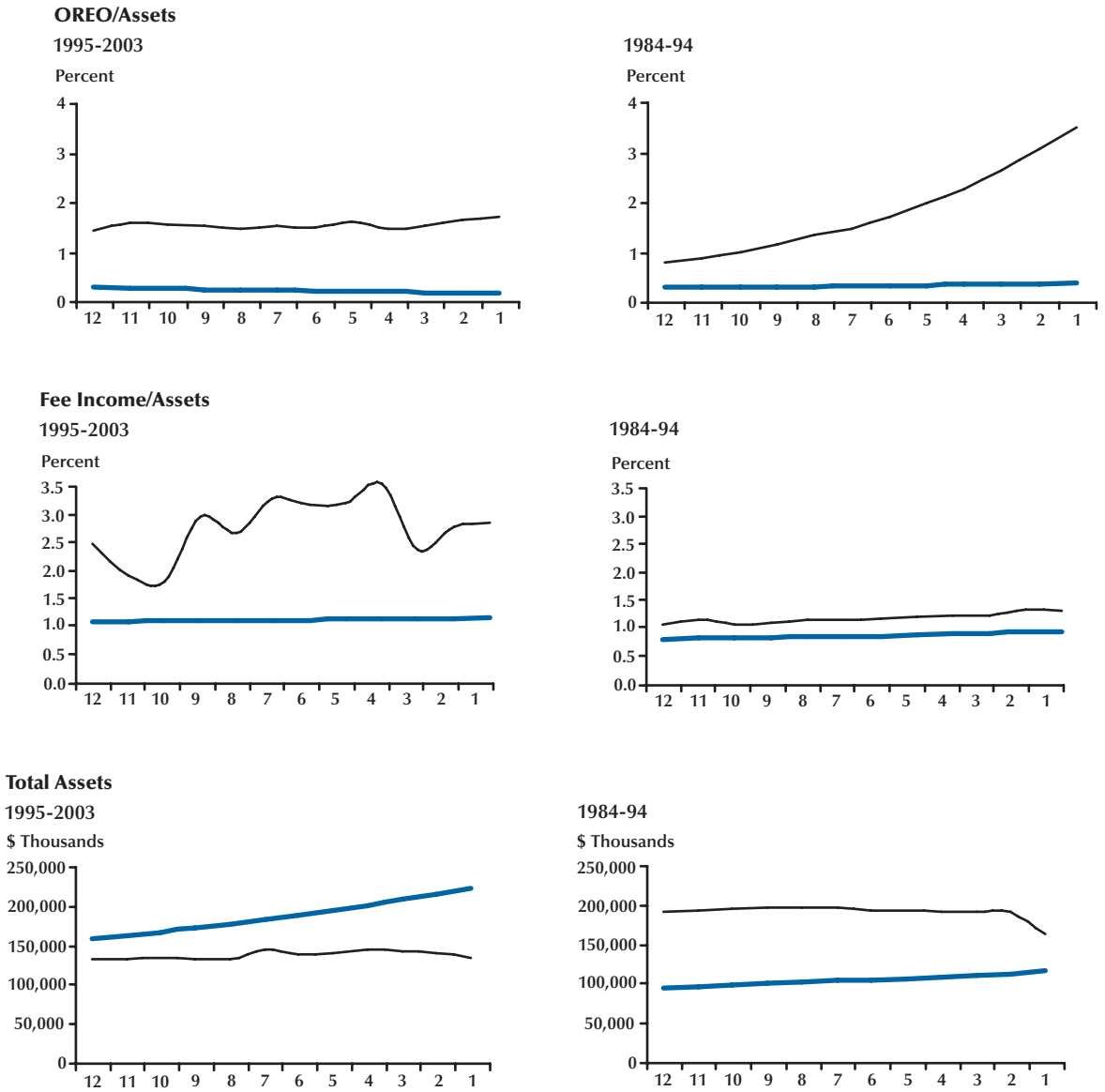


for the same series, shows that, despite the low number of failure observations in the second period, many of these changes are statistically significant. (The table reports the tests for one and six quarters prior to failure. The choice of the six-quarter horizon reflects the average time between bank exams.)

Failing banks in the 1995-2003 period had lower relative levels of liquidity risk than banks in the 1984-94 period. Specifically, between 1995 and 2003, failing banks relied substantially less on jumbo CDs and the purchase of federal funds, both in absolute terms and relative to safe banks. Although the ratio of demand deposits to total assets was lower for all banks in the later period, failing banks between 1995 and 2003 had ratios nearly identical to those at non-failing banks. In contrast, failing banks on average had significantly

Figure 1, cont'd

Trends at Failed Banks, Before and After 1995



fewer demand deposits as a percentage of assets than non-failing banks in the 1984-94 period. Finally, the cash-to-assets ratio increased at failing banks in the quarters leading up to failure in the 1995-2003 period, whereas that ratio displayed little pre-failure trend in the earlier period. These interperiod differences in liquidity risk could reflect the increased depositor discipline imposed by the legislative changes of the 1990s, as risky banks in the 1995-2003 period may have had a more difficult time attracting uninsured funds.

Credit-risk ratios also reflect significant differences between the two periods. Commercial real estate lending was significantly higher (about 4 percentage points, scaled by assets) at failing banks relative to non-failing banks in the earlier

period. In the later period the ratio was about the same at both failing and non-failing banks. Other real estate owned (OREO) as a percent of assets, previously one of the best predictors of failure, did not change substantially in the 1995-2003 period during the quarters leading up to failure. Although this ratio continues to be somewhat higher at failing banks relative to non-failing banks, the gap has shrunk, and the upward trend has nearly vanished. The loan loss reserves–to–total loans ratio was higher for failure banks in the later period than in the earlier period, although the ratio increased prior to failure in both time periods. The diminished importance of credit-risk ratios could reflect the improved risk-management processes at banks facilitated by the deepening of financial markets. Indeed, Schuermann (2004) argues that most banks came through the 2001 recession in excellent shape in part because of more effective risk management. Advances in credit scoring allowed banks to better risk-price their syndicated, retail, and small-business loans.

Two other ratios demonstrate the increased importance of diversification and nontraditional lines of business in recent years. Fee income as a percentage of assets, which was previously about the same at safe and failing banks, is now substantially higher for failing banks. Finally, failing banks were larger on average than non-failing banks in the earlier period but smaller in the later period, potentially reflecting the diversification benefits that banks receive from expanding in size and product offerings.

Despite the differences, we should be cautious about drawing strong conclusions from these graphs. The 1995-2003 sample contains only 44 bank-failure observations, so that, although most of our statistical tests yield statistically significant differences, the sample may not be entirely representative. In addition, some series that we have not emphasized have remained fairly constant. For example, failing banks continue to hold fewer mortgages and securities, and the pattern of capital deterioration has changed little. However, the fact remains that fundamental shifts in the banking environment make it possible that the path to bank distress has changed, and the recent data that are available are at least consistent with this possibility. Moreover, the shifts in the data—in variables associated with liquidity, credit, and operational risk—line up well with the types of institutional changes we know occurred during this period.

Because much of the academic research and most of the prevailing early-warning systems are based on data from the 1984-94 period, the above comparison gives us cause for concern. Indeed, these models tend to emphasize the variables that our evidence indicates have been *most* affected by the recent institutional changes. For example, eight of the eleven variables in SEER and 10 of the 12 variables in SCOR reflect either asset quality or liquidity. Recognition of the recent fundamental shifts in the nature of banking has motivated supervisors to consider new approaches to off-site monitoring.

NEW DIRECTIONS IN BANK-DISTRESS MODELS

In this section we describe some recent attempts by supervisory economists to build bank-distress models that (i) are less vulnerable than traditional models to the changing banking environment and (ii) are designed to assess risks that current models potentially overlook. We group the new models into two types: forward-looking models and risk-focused models. Forward-looking early-warning models may prove more robust to the changing bank environment because they rely on theory rather than past financial ratios to detect the circumstances that can lead banks to increase risk-taking. Risk-focused models reflect the shift to risk-focused supervision as explained in the Board of Governor's Supervision and Regulation Letter 97-25 titled "Risk-Focused Framework for the Supervision of Community Banks." The document, dated October 1, 1997, states the following:

> The objective of a risk-focused examination is to effectively evaluate the safety and soundness of the bank...focusing resources on the bank's highest risks. The exercise of examiner judgment to determine the scope of the examination during the planning process is crucial to the implementation of the risk-focused supervision framework, which provides obvious benefits

> such as higher quality examinations, increased efficiency, and reduced on-site examiner time... [E]ach Reserve Bank maintains various surveillance reports that identify outliers when a bank is compared to its peer group. The review of this information assists examiners in identifying both the strengths and vulnerabilities of the bank and provides a foundation from which to determine the examination activities to be conducted.

Rather than identifying banks with high levels of overall risk, risk-focused monitoring devices attempt to assess the particular risks of banking organizations, allowing examiners to allocate resources to upcoming exams more efficiently. Risk-focused models have the added advantage that they scrutinize risks that traditional models may overlook because those risks were not systematically important in historical episodes of bank distress. We emphasize, however, that the new models should be viewed as complements to rather than substitutes for the more comprehensive and time-tested systems.

Forward-Looking Models

Forward-looking models tend to focus on asset growth and liquidity as key risk indicators. Adverse selection and moral hazard incentives provide complementary stories for why banks pursuing rapid asset-growth strategies may be ramping up risk.

The adverse selection story views banks as having well-established relationships with a core set of customers. On the liability side of the balance sheet, these customers provide stable low-cost funding, while on the asset side the bank has information about the creditworthiness of these customers that generally is not available to other lenders. Banks that pursue a rapid growth strategy must move into new markets or offer new products, finding both a new set of borrowers and the funds to finance the growth. Although growth is not a problem per se, the bank will suffer from adverse selection if its pool of prospective new borrowers is composed disproportionately of those who have been rejected by other banks. The question is whether the bank has sufficient expertise and devotes sufficient resources to address the credit problems inherent in rapid growth. These problems are not observable immediately because it takes time for loans to become delinquent.

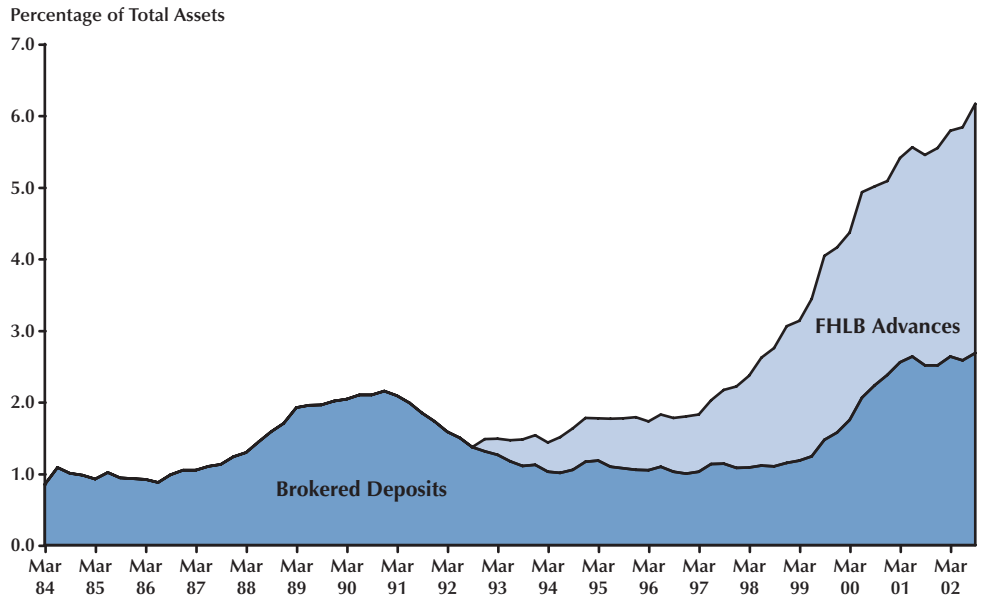
The moral hazard story views deposit insurance and other sources of collateralized funding as vehicles for bank risk-taking. Banks keep the profits if the risks pay off, but leave the losses to the FDIC in the event of failure. Banks with relatively high capital ratios have incentives to manage their banks prudently because the owners of the bank have their own funds at stake. If capital ratios begin to slip, however, those incentives may erode (Keeley, 1990). When bank performance begins to deteriorate for whatever reason, managers and owners increasingly face the prospect of losing their wealth and jobs should regulators close the bank. Rather than watch the bank fail, management might prefer to gamble for resurrection by booking high-risk loans funded with insured or collateralized funding. Indeed, this type of behavior is often blamed, in part, for the magnitude of the 1980s' thrift crisis (White, 1991).

Banks traditionally have tried to avoid market discipline by relying on core deposits, and some evidence suggests that riskier banks shift to core funding for exactly this reason.[12] Managers adopting this strategy, however, run up against two constraints. First, banks that deliberately try to sidestep market discipline with FDIC-insured deposits may invite greater regulatory scrutiny. Second, the limited supply of core funding imposes a natural ceiling on asset growth. Since the early 1990s, competition for insured deposits has intensified. Faced with less insured funding and greater demand for bank assets, managers have sought new funding sources. Banks that want to grow quickly but are unwilling to pay the risk premia demanded by uninsured liability holders may turn to noncore, non-risk-priced (NCNRP) sources of funding such as insured brokered deposits and FHLB advances. Brokered deposits funded much of the risky growth at thrifts during the 1980s. FHLB advances, which were historically available only to thrifts but became available to commercial banks in 1989, have many of

[12] Billet, Garfinkle, and O'Neal (1998).

Figure 2

Noncore, Non-Risk-Priced Funding at U.S. Banks



the same properties as brokered deposits.[13] Both types of funding are easily accessible in large quantities, and neither is priced according to the failure risk of the borrower. Brokered deposits are insured by the FDIC, while FHLB advances are fully collateralized. The lenders, therefore, have little incentive to monitor a borrowing bank's condition.

As Figure 2 illustrates, bank reliance on brokered deposits and FHLB advances is at an historically high level, both in absolute terms and as a percentage of total bank assets. Advances in particular have grown from essentially 0 to 3.5 percent of banks' balance sheets since the early 1990s. Furthermore, rapid loan growth has accompanied the growth in noncore funding at many institutions. Between 1994 and 2004, bank lending increased 39 percent faster than total national income. Although aggregate capital levels and overall bank condition remained relatively sound over this period, the rapid growth could be an indication of imprudent lending. The FDIC and the Federal Reserve Bank of St. Louis have independently developed alternative early-warning models called the Growth Monitoring System (GMS) and the Liquidity and Asset Growth Screen (LAGS), respectively, to address the adverse selection and moral hazard concerns. We briefly describe each in turn.

Growth Monitoring System. The FDIC has used the GMS as part of its off-site review process since the mid-1980s. The original model was an "expert system" in that its parameter values were assigned based on professional judgment, rather than statistical analysis. Weights were assigned to a number of growth-related variables in an attempt to identify those institutions most in danger of a rating downgrade. In the late 1990s, the FDIC developed a new version of this model using statistical techniques. This newer version of GMS, implemented in 2000, uses a logit model of downgrades, much like more traditional models, estimating which institutions that are

[13] Stojanovic, Vaughan, Yeager (2001) provide further discussion of why the FHLB might create incentives for abnormal risk-taking and evidence in support of this hypothesis. Wang and Sauerhaft (1989) show that thrift reliance on FHLB advances and brokered deposits were associated with worse supervisory ratings in the 1980s.

currently rated satisfactory are most likely to be classified as problem banks at the end of three years. Rather than using credit-quality measures as independent variables, GMS includes forward-looking variables that can be precursors of problems that have yet to become manifest. The key variables in the model are indicated in Table 3.[14] Two variables have the most effect on the results: loan growth and noncore funding. Although the coefficient magnitudes vary somewhat over time, they are both statistically and economically significant. More rapid loan growth and heavy dependence on noncore funding generally lead to higher estimated default probabilities.

Back-testing of GMS shows that the model has significant forecasting power.[15] Between 1996 and 2000, approximately 30 percent of the banks with GMS rankings at or above the 98th percentile received a rating of 3 or worse over the next five years.[16] Among the banks with rankings at the 79th percentile or lower, just 8 percent were downgraded, so banks in the top two percentiles were approximately two and a half times more likely to receive a rating of 3 or worse.

The performance of GMS is even better when flagging more severe problems. Banks with GMS rankings at or above the 98th percentile were downgraded to a CAMELS 4 or 5 or failed 9.5 percent of the time; in contrast, banks with GMS ratings in the lower 79th percentile were downgraded to a rating of 4 or 5 or failed only 1.3 percent of the time. Finally, banks with GMS rankings at or above the 98th percentile were over eight times more likely to fail (0.76 percent) than those banks with rankings in the 79th percentile or lower (0.09 percent). It should be noted that while the GMS model has notable success in identifying risky institutions, many banks with high GMS rankings are never downgraded. In other words, the type-II error rate is high.

Liquidity and Asset Growth Screen. Like GMS, LAGS attempts to flag banks that use particular funding vehicles to fuel rapid asset growth. The central idea is that a bank that experiences a combination of falling capital ratios, rapid asset growth, and a surge in noncore, non-risk-priced funding exhibits the classic characteristics of moral hazard.

The LAGS model consists of ten separate panel vector autoregressions (VARs), identical in their variables but estimated on banks of different inflation-adjusted asset classes. The four dependent variables in the VARs are the quarterly growth rate of risk-weighted assets; the ratio of brokered deposits and FHLB advances to total assets; the CAMELS composite score; and the ratio of equity to total assets.[17] The equations are estimated on rolling samples of quarterly data, updated every three months to include the most recent figures available. The key variable in the model is the CAMELS score. Banks that have higher forecasted CAMELS ratings over a three-year horizon are interpreted as being in greater danger of moral-hazard-induced risk.

The charts in Figure 3 show how LAGS works for a hypothetical bank as of June 2004. In each of the four panels, the data to the left of the vertical black lines represent the bank's behavior over the previous two years. To the right of the black lines, the graphs show the LAGS forecasts. LAGS predicts that the sample bank's CAMELS score will rise from its present level of 1 to 1.78 over the next three years. LAGS ranks banking institutions by the predicted rise in total risk.

A closer look at the sample bank's recent history gives us an idea of why the model predicts

[14] Noncore funding, loans to total assets, and assets per employee are adjusted for size peers. The growth variables and the change in loan mix are not adjusted because there is no evidence that the size peers differ. All growth rates are measured year over year to avoid problems of seasonal adjustment. The growth rates of loans and assets are adjusted for mergers, but the growth rates in noncore funding and equity are not. This adjustment means that the model ignores acquisitions unless the acquisitions have eroded equity or made the bank more dependent on noncore funding.

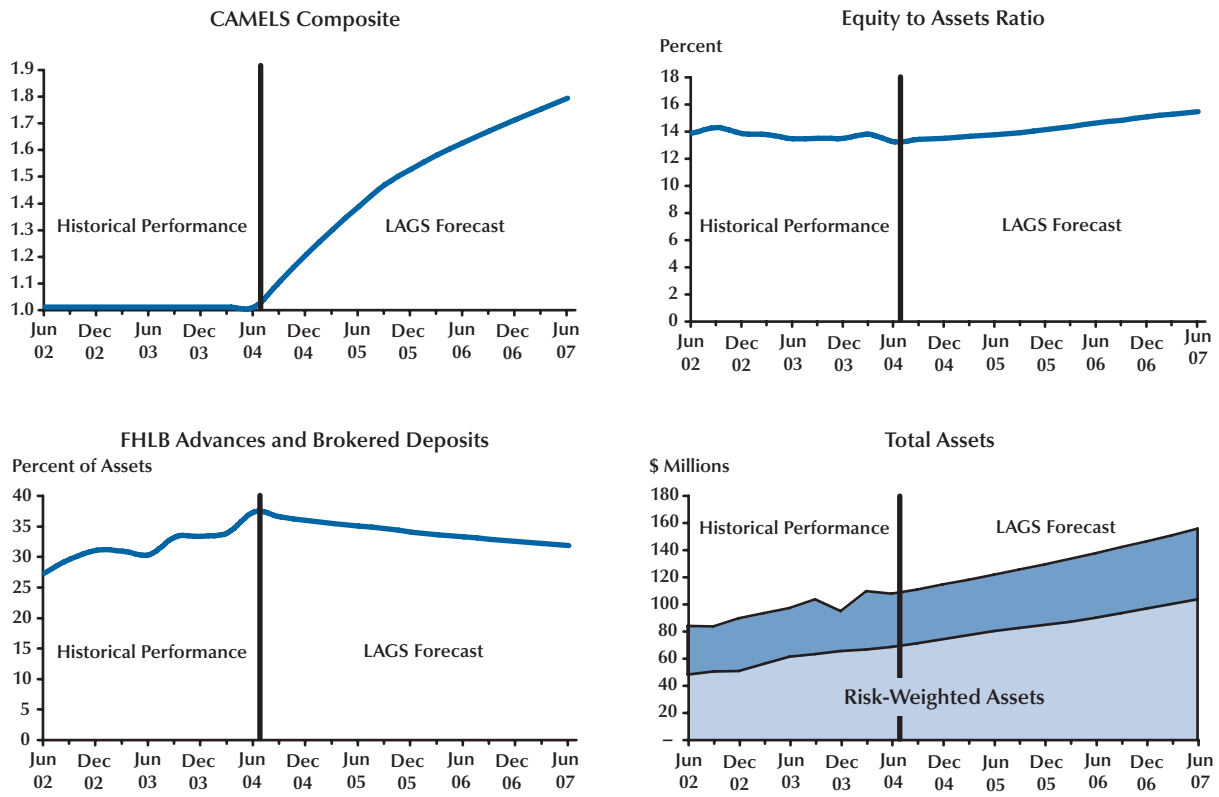
[15] The GMS system has also had particular success identifying recent failures due to fraud, although the exact reasons for this success require further investigation.

[16] Of course, the full five years has not passed for ratings assigned in the year 2000. The results are for those banks that survived five years or that filed a September 2003 call report.

[17] Eight quarterly lags of each of these four variables are included as regressors in each of the four equations. The equations also include intercept terms. In total, then, LAGS consists of 40 linear regression equations each containing 36 variables. Banks are excluded from the sample if they are less than eight quarters old or have merged with another institution within the previous eight quarters. As of June 30, 2004, the dataset included approximately 175,000 observations.

Figure 3

LAGS Forecasts for an Anonymous Bank as of June 2004



such a dramatic rise in risk. The bank grew rapidly between June 2002 and June 2004, increasing its assets by half and ratcheting up its risk-weighted asset ratio. The bank funded a substantial portion of this growth with FHLB advances and brokered deposits. As of June 2004, these liabilities supported over 35 percent of the bank's total assets, a ratio that rose more than 10 percentage points during the previous two years. Meanwhile, capital declined by about 100 basis points. The bank, therefore, displays key moral hazard characteristics.

Given the narrow focus of the LAGS model, we would not expect its performance to be as impressive as that of a more comprehensive model such as SEER, yet LAGS does display significant discriminatory ability. Between March 1998 and June 2001, 21.7 percent of banks with a CAMELS rating of 2 and a LAGS score at the 90th percentile or above were downgraded to a CAMELS score of 3, 4, or 5 or failed within the following three years. In addition, 47.1 percent of the 2-rated banks with a LAGS score at the 99th percentile or above were downgraded or failed within three years. By contrast, only 12.7 percent of banks below the 90th percentile were subsequently downgraded or failed.[18]

Risk-Focused Models

In addition to becoming more forward-looking, bank-distress models are also evolving to accommodate the risk-focused framework. Several off-site–monitoring devices have already been

[18] As noted, the LAGS coefficients are reestimated every quarter. The numbers reported in this paragraph reflect the estimates actually used in each quarter (rather than, say, the most recent set). In other words, they reflect out-of-sample forecasting ability.

developed by the FDIC and the Federal Reserve, and more are in development. We describe two of these models here.

Real Estate Stress Test. Real estate crises have been perennial causes of bank failures.[19] In 2000, the FDIC implemented a Real Estate Stress Test (REST) that attempts to identify those banks and thrifts that are most vulnerable to problems in real estate markets.[20]

The REST model incorporates the experience of the New England real estate crisis of the early 1990s. Conceptually, the model subjects banks to the same stress as that crisis and forecasts the resulting CAMELS ratings. REST was developed by regressing performance data for New England banks in December 1990 on performance and portfolio data for the same banks in December 1987. These regressions identify the factors that were observable in 1987 that later were associated with safety and soundness concerns. A concentration in construction and development loans is the primary risk factor, but there are a host of secondary factors, such as concentrations in commercial mortgages, commercial and industrial loans, mortgages on multifamily housing, reliance on noncore funding, and rapid growth. These regressions are used to forecast measures of bank performance which are then translated to CAMELS ratings using the SCOR model. The result is a REST rating that ranges from 1 to 5. The output from the model is distributed to FDIC examiners as well as examiners from other federal and state banking agencies. The model has been validated using data from other real estate downturns; it can identify banks that are vulnerable from real estate exposure three to seven years in advance.

Because of the long horizon, banks with poor REST ratings are not an immediate concern. More importantly, the model does not consider the underwriting standards and other aspects of risk management that the bank uses to control its exposure to real estate downturns. Consequently, examiners use the output from the REST model for examination planning. The model produces a set of "weights" indicating which variables are the most responsible for the poor rating, giving examiners a sense of the aspects of a bank's operations that deserve the most attention.

Interest Rate Risk. The savings and loan crisis of the 1980s focused increased attention in the banking industry on interest rate risk. Economists at the Board of Governors responded by developing a duration-based measure of interest rate risk that could be used for surveillance and risk-scoping purposes.[21] The model, titled the Economic Value Model (EVM), was launched in the first quarter of 1998 by producing a confidential quarterly surveillance report (called the Focus report) for each commercial bank.

The EVM aggregates balance sheet items into various buckets based upon maturity and optionality. The model then uses the duration from a proxy financial instrument for each bucket to calculate the "risk weight," or the change in economic value of those items that would result from a 200-basis-point instantaneous rise in rates. For example, the EVM places all residential mortgages that reprice or mature within 5 to 15 years in the same bucket. If the risk weight for the 5- to 15-year mortgages were 7.0, the value of the 5- to 15-year mortgages would be estimated to decline by 7.0 percent following an immediate 200-basis-point rate hike. The change in economic value is repeated for each balance sheet bucket. The predicted change in economic value of the bank's equity, then, is the difference between the predicted change in assets and the predicted change in liabilities.

Recent research by Sierra and Yeager (2004) shows that the EVM effectively ranks banks by their exposure to rising interest rates. That is, banks that the model predicts to be the most vulnerable to rising interest rates suffer the largest declines in income and equity following an interest rate hike. These banks also show the largest gains in income and equity following interest rate declines. Bank supervisors can use the model's output to rank banks by interest rate risk. If a bank is found to be an outlier, the examiner in charge will emphasize that risk in the next exam.

[19] See Herring and Wachter (1999).

[20] See Collier et al. (2003b).

[21] See Embersit and Houpt (1991) and Houpt and Wright (1996) for details.

CONCLUSION

After their initial introduction in the 1970s, studies on the causes of bank distress made rapid progress, fueled by considerable academic interest. In recent years, this interest has waned outside the regulatory community, possibly reflecting a belief that the causes of bank distress are well understood. However, significant legislative, financial, and technological innovations may make it necessary to supplement the prevailing academic and regulatory models with a new generation of forward-looking and risk-focused monitoring systems.

Newer forward-looking models at the FDIC and the Federal Reserve include the Growth Monitoring System and the Liquidity and Asset-Growth Screen. Risk-focused models include the Real Estate Stress Test and the Economic Value Model. Additional monitoring devices such as those analyzing liquidity risk, operational risk, and counterparty risk seem promising lines of inquiry.

REFERENCES

Altman, Edward I. "Financial Ratios, Discriminant Analysis, and the Prediction of Corporate Bankruptcy." *Journal of Finance*, September 1968, *23*(4), pp. 589-609.

Altman, Edward I. "Predicting Performance in the Savings and Loan Association Industry." *Journal of Monetary Economics*, October 1977, *3*(4), pp. 443-66.

Barth, James; Brumbaugh, Dan Jr.; Sauerhaft, Daniel and Wang, George H.K. "Thrift Institution Failures: Estimating the Regulator's Closure Rule," in George Kaufman, ed., *Research in Financial Services*. Greenwich, CT: JAI Press, 1989, pp. 1-25.

Bennett, Rosalind L.; Vaughan, Mark D. and Yeager, Timothy J. "Should the FDIC Worry about the FHLB? The Impact of Federal Home Loan Bank Advances on the Bank Insurance Fund." Supervisory Policy Analysis Working Paper 2005-1, Federal Reserve Bank of St. Louis, July 2005.

Billet, Matthew T.; Garfinkle, Jon A. and O'Neal, Edward S. "The Cost of Market vs. Regulatory Discipline in Banking." *Journal of Financial Economics*, 1998, *48*(3), pp. 333-58.

Bovenzi, John F.; Marino, James A. and McFadden, Frank E. "Commercial Bank Failure Prediction Models." Federal Reserve Bank of Atlanta *Economic Review*, November 1983, *68*, pp. 14-26.

Claessens, Stijn; Glaessner, Thomas and Klingebiel, Daniela. "Electronic Finance: Reshaping the Financial Landscape around the World." *Journal of Financial Services Research*, August-October 2002, *22*(1-2), pp. 29-61.

Cole, Rebel A. "When Are Thrift Institutions Closed? An Agency-Theoretic Model." *Journal of Financial Services Research*, December 1993, *7*(4), pp. 283-307.

Cole, Rebel A.; Cornyn, Barbara G. and Gunther, Jeffery W. "FIMS: A New Monitoring System for Banking Institutions." *Federal Reserve Bulletin*, January 1995, *8*(1), pp. 1-15.

Cole, Rebel A. and Gunther, Jeffery W. "Predicting Bank Failures: A Comparison of On- and Off-Site Monitoring Systems." *Journal of Financial Services Research*, April 1998, *13*(2), pp. 103-17.

Collier, Charles; Forbush, Sean; Nuxoll, Daniel A. and O'Keefe, John. "The SCOR System of Off-Site Monitoring." *FDIC Banking Review*, Third Quarter 2003a, *15*(3), pp. 17-32.

Collier, Charles; Forbush, Sean and Nuxoll, Daniel A. "The Vulnerability of Banks and Thrifts to a Real Estate Crises." *FDIC Banking Review*, Fourth Quarter 2003b, *15*(4), pp. 19-36.

Cornett, Marcia M.; Mehran, Hamid and Tehranian, Hassan. "The Impact of Risk-Based Premiums on FDIC-Insured Institutions." *Journal of Financial Services Research*, April 1998, *13*(2), pp. 153-69.

Cox, D.R. "Regression Models and Life Tables." *Journal of the Royal Statistical Society*, 1972, Series B (34), pp. 187-220.

Craig, Ben R. and Thomson, James B. "Federal Home Loan Bank Lending to Community Banks: Are

Targeted Subsidies Desirable?" *Journal of Financial Services Research*, February 2003, *23*(1), pp. 5-28.

Demirgüç-Kunt, Asli. "Modeling Large Commercial-Bank Failures: A Simultaneous-Equations Analysis." Working Paper 8905, Federal Reserve Bank of Cleveland, March 1989.

Duffee, Gregory R. and Zhou, Chunsheng. "Credit Derivatives in Banking: Useful Tools for Managing Risk?" *Journal of Monetary Economics*, August 2001, *48*(1), pp. 25-54.

Embersit, James A. and Haupt, James V. "A Method for Evaluating Interest Rate Risk in U.S. Commercial Banks." *Federal Reserve Bulletin*, August 1991, *77*(8), pp. 625–37.

Flannery, Mark J. and Rangan, K. "Market Forces at Work in the Banking Industry: Evidence from the Capital Buildup of the 1990s." Working paper, University of Florida–Gainesville, 2003.

Flannery, Mark J. and Sorescu, Sorin M. "Evidence of Bank Market Discipline in Subordinated Debenture Yields: 1983–1991." *Journal of Finance*, September 1996, *51*(4), pp. 1347-77.

Gilbert, R. Alton; Meyer, Andrew P. and Vaughan, Mark D. "The Role of Supervisory Screens and Econometric Models in Off-Site Surveillance." Federal Reserve Bank of St. Louis *Review*, November/December 1999, *81*(6), pp. 2-27.

Gilbert, R. Alton; Meyer, Andrew P. and Vaughan, Mark D. "Could a CAMELS Downgrade Model Improve Off-Site Surveillance?" Federal Reserve Bank of St. Louis *Review*, January/February 2002, *84*(1), pp. 47-63.

Goldberg, Lawrence G. and Hudgins, Sylvia C. "Depositor Discipline and Changing Strategies for Regulating Thrift Institutions." *Journal of Financial Economics*, February 2002, *63*(2), pp. 263-74.

Hall, John R.; King, Thomas B.; Meyer, Andrew P. and Vaughan, Mark D. "Did FDICIA Enhance Market Discipline at Community Banks?" in George G. Kaufman, ed., *Research in Financial Services: Private and Public Policy*. Volume 14. Boston: Elsevier, 2002, pp. 63-94.

Hanweck, Gerald A. "Predicting Bank Failure." Board of Governors of the Federal Reserve System, *Research Papers in Banking and Financial Economics*, November 1977, *19*.

Helwege, Jean. "Determinants of Savings and Loan Failures: Estimates of a Time-Varying Proportional Hazard Function." *Journal of Financial Services Research*, December 1996, *10*(4), pp. 373-92.

Herring, Richard J. and Wachter, Susan M. "Real Estate Booms and Banking Busts—An International Perspective." Occasional Paper No. 58. Group of Thirty, 1999.

Hirtle, Beverly J. and Lopez, Jose A. "Supervisory Information and the Frequency of Bank Examinations," Federal Reserve Bank of New York *Economic Policy Review*, April 1999, *5*(1), pp. 1-19.

Hooks, Linda M. "Bank Asset Risk: Evidence from Early-Warning Models." *Contemporary Economic Policy*, October 1995, *13*(4), pp. 36-50.

Houpt, James V. and Wright, David M. "An Analysis of Commercial Bank Exposure to Interest Rate Risk." *Federal Reserve Bulletin*, February 1996, *82*(2), pp. 115-28.

Instefjord, Norvald. "Risk and Hedging: Do Credit Derivatives Increase Bank Risk?", *Journal of Banking and Finance*, February 2005, *29*(2), pp. 333-45.

Keeley, Michael C. "Deposit Insurance, Risk, and Market Power in Banking." *American Economic Review*, December 1990, *80*(5), pp. 1183-200.

King, Thomas B. "Discipline and Liquidity in the Market for Federal Funds." Supervisory Policy Analysis Working Paper 2003-2, Federal Reserve Bank of St. Louis, October 2005.

Korobow, Leon; Stuhr, David P. and Martin, Daniel. "A Nationwide Test of Early Warning Research in Banking." Federal Reserve Bank of New York *Quarterly Review*, Autumn 1977, *2*(2), pp. 37-52.

Lane, William R.; Looney, Stephen W. and Wansley. James W. "An Application of the Cox Proportional Hazards Model to Bank Failure." *Journal of Banking and Finance*, December 1986, *10*(4), pp. 511-31.

Marino, James A. and Bennett, Rosalind L. "The Consequences of National Depositor Preference." *FDIC Banking Review*, October 1999, *12*(2), pp. 19-38.

Martin, Daniel. "Early Warning of Bank Failure: A Logit Regression Approach." *Journal of Banking and Finance*, November 1977, *1*, pp. 249-76.

Meyer, Paul A. and Pifer, Howard W. "Prediction of Bank Failures." *Journal of Finance*, September 1970, pp. 853-68.

Meyer, Andrew P. and Yeager, Timothy J. "Are Small Rural Banks Vulnerable to Local Economic Downturns?" Federal Reserve Bank of St. Louis *Review*, March/April 2001, *83*(2), pp. 25-38.

Neely, Michelle C. and Wheelock, David C. "Why Does Bank Performance Vary Across States?" Federal Reserve Bank of St. Louis *Review*, March/April 1997, *79*(2), pp. 27-40.

Nuxoll, Daniel; O'Keefe, John and Samolyk, Katherin. "Do Local Economic Data Improve Off-Site Bank Warning Models?" *FDIC Banking Review*, 2003, *15*(2), pp. 39-53.

Pantalone, Coleen C. and Platt, Marjorie B. "Predicting Commercial Bank Failure since Deregulation." Federal Reserve Bank of Boston *New England Economic Review*, July/August 1987, pp. 37-47.

Putnam, Barron H. "Early-Warning Systems and Financial Analysis in Bank Monitoring." Federal Reserve Bank of Atlanta *Economic Review*, November 1983, *68*, pp. 6-14.

Rose, Peter S. and Scott, William L. "Risk in Commercial Banking: Evidence from Postwar Failures." *Southern Economic Journal*, July 1978, *45*(1), pp. 90-106.

Schuermann, Til. "Why Were Banks Better Off in the 2001 Recession?" Federal Reserve Bank of New York *Current Issues in Economics and Finance*, January 2004, *10*(1), pp. 1-7.

Sierra, Gregory E. and Yeager, Timothy J. "What Does the Federal Reserve's Economic Value Model Tell Us about Interest Rate Risk at U.S. Community Banks?" Federal Reserve Bank of St. Louis *Review*, November/December 2004, *86*(6), pp. 45-60.

Sinkey, Joseph F. Jr. "A Multivariate Statistical Analysis of the Characteristics of Problem Banks." *Journal of Finance*, March 1975, *30*(1), pp. 21-36.

Sinkey, Joseph F. Jr. "Identifying 'Problem' Banks: How Do the Banking Authorities Measure a Bank's Risk Exposure?" *Journal of Money, Credit, and Banking*, May 1978, *10*(2), pp. 184-93.

Sinkey, Joseph F. Jr. and Carter, David A. "Evidence on the Financial Characteristics of Banks That Do and Do Not Use Derivatives." *Quarterly Review of Economics and Finance*, Winter 2000, *40*(4), pp. 431-49.

Stojanovic, Dusan; Vaughan, Mark D. and Yeager, Timothy J. "Do Federal Home Loan Bank Advances and Membership Lead to More Bank Risk?" in Federal Reserve Bank of Chicago, ed., *The Financial Safety Net: Costs, Benefits, and Implications for Regulations—Proceedings of the 37th Annual Conference on Bank Structure and Competition*, May 2001, pp. 165-96.

Stuhr, David P. and van Wicklen, Robert. "Rating the Financial Condition of Banks: A Statistical Approach to Aid Bank Supervision." Federal Reserve Bank of New York *Monthly Review*, September 1974, pp. 233-8.

Thomson, James B. "Modeling the Bank Regulator's Closure Option: A Two-Step Logit Regression Approach." *Journal of Financial Services Research*, May 1992, *6*(1), pp. 5-23.

Wang, George H.K. and Sauerhaft, Daniel. "Examination Ratings and the Identification of Problem/Non-Problem Thrift Institutions." *Journal of Financial Services Research*, October 1989, *2*(4), pp. 319-42.

West, Robert Craig. "A Factor-Analytic Approach to Bank Condition." *Journal of Banking and Finance*, June 1985, *9*(2), pp. 253-66.

Whalen, Gary. "A Proportional Hazards Model of Bank Failure: An Examination of Its Usefulness as an Early Warning Tool." Federal Reserve Bank of Cleveland *Economic Review*, First Quarter 1991, pp. 20-31.

Whalen, Gary and Thomson, James B. "Using Financial Data to Identify Changes in Bank Condition." Federal Reserve Bank of Cleveland *Economic Review*, Second Quarter 1988, pp. 17-26.

Wheelock, David C. and Wilson, Paul W. "Explaining Bank Failures: Deposit Insurance, Regulation, and Efficiency." *Review of Economics and Statistics*, 1995, *77*(4), pp. 689-700.

Wheelock, David C. and Wilson, Paul W. "Why Do Banks Disappear? The Determinants of U.S. Bank Failures and Acquisitions." *Review of Economics and Statistics*, February 2000, *82*(1), pp. 127-38.

Wheelock, David C. and Wilson, Paul W. "The Contribution of On-Site Examination Ratings to an Empirical Model of Bank Failures." *Review of Accounting and Finance*, 2005 (forthcoming).

White, Eugene N. *The Comptroller and the Transformation of American Banking*. Washington, DC: Office of the Comptroller of the Currency, 1992, pp. 1960-90.

White, Lawrence J. *The S&L Debacle: Public Policy Lessons for Bank and Thrift Regulation*. New York: Oxford University Press, 1991.

Yeager, Timothy J. "The Demise of Community Banks? Local Economic Shocks Are Not to Blame." *Journal of Banking and Finance*, September 2004, *28*(9), pp. 2135-53.

Yeager, Timothy J.; Yeager, Fred C. and Harshman, Ellen. "The Financial Services Modernization Act: Evolution or Revolution?" Federal Reserve Bank of St. Louis Working Paper No. 2004-05, 2005.

Replicability, Real-Time Data, and the Science of Economic Research: FRED, ALFRED, and VDC

Richard G. Anderson

This article discusses the linkages between two recent themes in economic research: "real time" data and replication. These two themes share many of the same ideas, specifically, that scientific research itself has a time dimension. In research using real-time data, this time dimension is the date on which particular observations, or pieces of data, became available. In work with replication, it is the date on which a study (and its results) became available to other researchers and/or was published. Recognition of both dimensions of scientific research is important. A project at the Federal Reserve Bank of St. Louis to place large amounts of historical data on the Internet holds promise to unify these two themes.

Federal Reserve Bank of St. Louis *Review*, January/February 2006, *88*(1), pp. 81-93.

REPLICATION AND REAL-TIME ECONOMETRICS

During the past 25 years, two themes have flowed steadily, albeit often quietly, through economic research: "real time" data and replication. In *replication* studies, the issue is determining which data were used and whether the author performed the calculations as described; in *real-time* data studies, the issue is determining the robustness of the study's findings to data revisions. These themes share the same core idea: that scientific research has an inherent time dimension. In both real-time data and replication studies, the time dimension is the date on which particular observations, or pieces of data, became available to researchers. Projects at Harvard University and at the Federal Reserve Bank of St. Louis promise to improve the quality of empirical economic research by unifying these themes.[1]

Although replication studies focus on the *correctness* of results and real-time studies on their *robustness*, economic theory suggests that these are related—the likelihood that an author's error will become visible to other researchers is an inverse function of the cost of conducting tests for replicability and robustness. Yet, for the profession, excessive emphasis on the criminal-detection aspects of replication (Did the author fake the results? Or did the author cease experimenting prematurely when a favorable result appeared?) has tended to increase the reluctance of researchers to share data and program code. That is, to the extent that the profession overemphasizes the manhunt of David Dodge's 1952 *To Catch a Thief*, it risks foregoing the benefits of Sir Isaac Newton's 1676 dictum, "If I have seen further it is by standing on the shoulders of giants."

The incentives and disincentives for a researcher to share data have been discussed by numerous authors (e.g., Fienberg, Martin, and Starf,

[1] The fallacy that neither real-time data nor replication studies are needed because "important" results always will sift to the top through repeated studies is addressed at length in Anderson et al. (2005).

Richard G. Anderson is a vice president and economist at the Federal Reserve Bank of St. Louis. This is a revised version of paper prepared for the American Economic Association meeting, Philadelphia, PA, January 2005. The author thanks Bruce McCullough of Drexel University, Gary King of Harvard University, and Dan Newlon of the National Science Foundation for their comments. Giang Ho provided research assistance.

© 2006, The Federal Reserve Bank of St. Louis. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

1985; Boruch and Cordray, 1985; Dewald, Thursby, and Anderson, 1986; Feigenbaum and Levy, 1993; Anderson and Dewald, 1994; Bornstein, 1991; Bailar, 2003).[2] Researchers receive a stream of rewards for the new knowledge contained in a published article, which begins with publication and eventually tapers to near zero. Furnishing the data to other researchers invites the risk that a replication will demonstrate the article's results to be false, an event which immediately ends the reward stream. If the replication further uncovers malicious or unprofessional behavior (such as fraud or other unethical conduct), "negative rewards" flow to the researcher.

Creating original research manuscripts for professional journals is craft work. Although often referred to as "knowledge workers," researchers might equally well be regarded as artisans, with creative tasks that include collecting data, writing code for statistical analysis or model simulation, and authoring the final manuscript.[3] Similar to the work of other craftsmen, researchers' output contains intellectual property—not only the final manuscript, but also the data and programs developed during its creation. Yet, for academic-type researchers, some of the intellectual property must be relinquished so the work can be published in peer-reviewed journals. This conflict creates a strategic game in which the researcher feels compelled to reveal a sufficient amount of his material to elicit publication, while simultaneously seeking to retain for himself as much of the intellectual property as possible. There are few, if any, models of this process in the economics literature. One such analysis is presented by Anderson et al. (2005), based on the Crawford and Sobel (1982) model of strategic information withholding. A complete presentation of their theoretical analysis is beyond the scope of this paper. The results buttress, however, the commonsense intuition that so long as withholding data and program code does not reduce the post-publication stream of rewards (and disclosure of data and program code does not increase it), researchers will rationally choose not to disclose data and programs.[4] Such models largely explain the well-known proclivity of academic researchers in many disciplines, including economics, to keep secret their data and programs. For the progress of scientific economic research, such an equilibrium is suboptimal.

One solution to suboptimal equilibria is collective action. One collective action is for professional journals to archive data and program code.[5] Such archives—which permit low-cost, anonymous, ad hoc replication—can improve the quality of published research by way of an effect reminiscent of Baumol-like credible threats of market entry. This process was well-described by the University of Chicago's John Bailar (2003) at a recent National Research Council conference:

> Of all the public myths about how science is done, one of the broadest and most persistent is that scientific method rests on replication of critical observations [i.e., results]. Straight replication is in fact uncommon, largely, I believe, because no scientist gets much professional credit for straightforward replication unless the findings are critical, there is suspicion of fraud, or there is some other unusual condition such that slavish replication of the methods reported might have some meaning not attached to the first round. Here I exclude replication by an independent investigator for the sole purpose of assuring himself or herself that the original results are correct and that the methods are working properly, as a preliminary to going further in some way.
>
> Overall, replication...seems to be one of those ideals that get a fair amount of discussion

[2] Some data cannot be shared. Examples include confidential banking data held by the Federal Reserve; micro data held by the Bureau of the Census; and various financial data, including that licensed by the University of Chicago's Center for Research in Security Prices (CRSP). In some cases, the owners/licensors of such data have archived the datasets built and used by individual researchers and made the datasets available to subscribers.

[3] Indeed, economists and other scientists often refer to "polishing" a final manuscript, in the spirit of woodworkers or stone masons polishing their work.

[4] The model of Feigenbaum and Levy (1993), in which rewards to researchers are driven by citations, also suggests that the divergence between the search for truth and rational individual choice will be largest for younger researchers (such as those without academic tenure), who will be less inclined to search for errors than older researchers and less inclined to devote scarce time to documenting their work.

[5] Historical data, cataloged and indexed by the day on which the data became available to the public, often are referred to as "vintage" data.

> but have little influence on behavior. *Perhaps what is most important is that the original investigators publish background and methods with enough detail and precision for a knowledgeable reader to replicate the study if he had the resources and inclination to do so.* [emphasis added]

In a recent article, Pesaran and Timmermann (2005, p. 221) offer a formal statement of the correspondence between the universe of all possible datasets and an article's specific dataset:

> Let χ denote the time-invariant universe of all possible prediction variables that could be considered in the econometric model, while N_{xt} is the number of regressors available at time $\boldsymbol{t}$ so $X^t = (x_{1t},\ldots,x_{N_{xt}}) \subseteq \chi$. N_{xt} is likely to grow at a faster rate than the sample size, T. At some point there will therefore be more regressors than time-series observations. However, most new variables will represent different measurements of a finite number of underlying economic factors such as output/activity, inflation and interest rates.

Below, we use the notation X^t to denote the set of all observations [values, measurements], on a fixed list of economic variables, that have been published as of (up to and including) date t. Assuming that an author has not falsified or erroneously transcribed data values, the true dataset for a published article will be contained within the universe of all such datasets X^t, where t is no greater than the date on which the original author completed his research. Unfortunately, such datasets often are too large to be compiled by individual researchers.

Historical data, cataloged and indexed by the day on which the data became available to the public, are referred to as "vintage" data. Collections of such data—X^t in the notation above—are referred to as "real-time" datasets and are indexed by the date of the most recent data included, t. The first large-scale project to collect and make available to the public vintage macroeconomic data was started in 1991 by the Federal Reserve Bank of Philadelphia to assess the accuracy of forecasts collected in the Survey of Professional Forecasters (Croushore and Stark, 2001). That project, and its data, is referred to as the Real Time Dataset for Macroeconomists (RTDSM). The design of the Philadelphia RTDSM project seeks to provide snapshots of the values of certain macroeconomic variables as they were available to the public at the end of the 15th day of the center month of each quarter. Hence, although both monthly and quarterly data are included, the dataset's primary value is macroeconomic modeling and forecasting at quarterly frequencies.[6]

Data Projects

This essay discusses two on-going projects to support the collection and dissemination of vintage data. The first, ALFRED™, is the Federal Reserve Bank of St. Louis's Archival Federal Reserve Economic Data. (Some features of ALFRED are still under development and may not be available at the time this article is printed; see alfred.stlouisfed.org for an introduction, instructions, and updates.) The second, VDC, is Harvard University's Virtual Data Center project. The projects differ significantly from each other, and from the Philadelphia project, in several aspects:

- Data frequency: Both the RTDSM and ALFRED projects focus on macroeconomic data. The RTDSM dataset is designed for a quarterly observational data frequency. The ALFRED project is designed for data at a daily frequency or lower (e.g., weekly, biweekly, monthly, etc)—that is, any data frequency currently supported in the Federal Reserve Bank of St. Louis's FRED® (Federal Reserve Economic Data) database. The VDC project, discussed further below, focuses on archiving and sharing complete datasets from specific research studies and articles. As such, it is data-frequency independent.
- Data vintages: The RTDSM project provides snapshots of the values of certain macroeconomic variables as they were known by the public at the close of business on the 15th day of each quarter's center month.

[6] See the Federal Reserve Bank of Philadelphia's web site for details and documentation, available at www.phil.frb.org/econ/forecast/reaindex.html, as of October 18, 2005.

The ALFRED project, operating at a daily frequency, provides daily (end-of-day) snapshots of the values of all variables in the FRED database. The VDC project, because it stores complete datasets as provided by researchers, has no explicit vintage component. The lack of a vintage component restricts the value of its design, for economists, to replication alone—an important function, but distinctly different from studies of robustness that require vintage-indexed data such as RTDSM and ALFRED.

- Data updating: Neither the RTDSM nor VDC projects have a mechanism to automatically update their data vintages. Data are added to RTDSM as Philadelphia staff determine which figures were available to the public on the specified days. Datasets are added to the VDC project as researchers place them on the Internet and the VDC servers index their location. The architecture of ALFRED differs. "Under the hood," ALFRED and FRED share the same database architecture. In this shared design, data values on FRED that are revised—that is, replaced with newly released numbers—are automatically added to ALFRED as vintage data. Combined with a history of release dates for major economic indicators such as GDP and employment, ALFRED uniquely provides a day-by-day vintage snapshot of the evolution of macroeconomic variables. This architecture, as discussed further below, uniquely allows ALFRED to be used for both replication and robustness studies.

Further discussion of ALFRED's time-indexed architecture is contained in a subsequent section.

The Literature

Studies that explore the sensitivity of empirical results to data vintage and how policymaking might incorporate the data revision process have a long history. Early studies are reviewed by Croushore and Stark (2001). Selected recent studies include Neely, Roy, and Whiteman (2001), Orphanides (2001), Stark and Croushore (2002), Christoffersen, Ghysels, and Swanson (2002), Orphanides and van Norden (2002, 2003), Bernanke and Boivin (2003), Faust, Rogers and Wright (2003), Koenig, Dolmas and Piger (2003), Svensson and Woodford (2003, 2004), Clark and Kozicki (2004), and Kishor and Koenig (2005). The analysis of vintage, or "real-time," data also is a popular conference topic—examples include the Federal Reserve Bank of Philadelphia's 2001 "Conference on Real Time Data Analysis,"[7] the Bundesbank's 2004 conference "Real-Time Data and Monetary Policy,"[8] and the CIRANO/Bank of Canada's October 2005 "Workshop on Macroeconomic Forecasting, Analysis and Policy with Data Revision."[9]

One of the earlier experiments to demonstrate the dependence of empirical results on data vintage is reported in Dewald, Thursby, and Anderson (1986). During their project at the *Journal of Money, Credit, and Banking* from 1982 to 1984, a large number of authors, when asked to submit datasets and programs, replied that they did not save publicly available macroeconomic data because the data could easily be collected from published sources and their empirical results were nearly invariant to the vintage of the data.[10] To test this assertion, Dewald, Thursby, and Anderson examined in detail one article, Goldberg and Saunders (1981), which contains a model of the growth of foreign banks in the United States. The article's authors furnished their banking data (which had required considerable effort to collect) but not their macroeconomic data, which they said had been collected from various issues of the *Survey of Current Business* and *Federal Reserve Bulletin* with no record made of which numbers were obtained from which issues. Dewald, Thursby, and Anderson collected from the *Survey of Current Business* all published values on three macroeconomic variables used in the article

[7] Program and papers were available at www.phil.frb.org/econ/conf/rtdaconf.html, as of October 18, 2005.

[8] Program and papers were available at www.bundesbank.de/vfz/vfz_daten.en.php, as of October 18, 2005.

[9] Program and papers were available at www.cirano.qc.ca/financegroup/Real-timeData/program.php, as of October 18, 2005.

[10] Even authors who submitted data seldom noted the dates on which they collected the data or the dates on which the data had been published.

(imports, investment, and GNP) during the period 1972:Q4 through 1982:Q3.[11] From these data, they estimated 500 variants of the Goldberg-Saunders model, summarizing the results in a set of histograms (Dewald, Thursby, and Anderson, 1986, p. 599). Overall, the coefficient estimates obtained varied widely and the modal values often were far from the coefficients in the Goldberg and Saunders article.

DATA SHARING

The arguments above suggest that the quality of empirical economic science is positively correlated with the extent to which researchers preserve and share datasets and program code. This theme is commonplace in science. In 1979, the Committee on National Statistics of the National Research Council (of the National Academy of Sciences) sponsored a conference on the role of data sharing in social science research. A subsequent subcommittee on sharing research data stated the issues clearly (Fienberg, Martin, and Starf, 1985, pp. 3-4):

> Data are the building blocks of empirical research, whether in the behavioral, social, biological, or physical sciences. To understand fully and extend the work of others, researchers often require access to data on which that work is based. Yet many members of the scientific community are reluctant or unwilling to share their data even after publication of analyses of them. Sometimes this unwillingness results from the conditions under which data were gathered; sometimes it results from a desire to carry out further analyses before others do; and sometimes it results from the anticipated costs, in time or money, or both.
>
> The Committee on National Statistics believes that sharing scientific data with colleagues reinforces the practice of open scientific inquiry. Cognizant of the often substantial costs to the original investigator for sharing data, the committee seeks to foster attitudes and practices within the scientific community that encourage researchers to share data with others as much as feasible.

The subcommittee offered 16 recommendations (see the appendix) for improving the quality of social science research through data sharing. The recommendations are so straightforward as to seem self-evident: Sharing data should be standard practice; researchers should retain data for a reasonable period after publication; researchers requesting data should bear the costs of providing data; funding organizations should encourage data sharing by requesting a data-sharing plan in requests for funding; and journals should encourage authors to share data. Yet, two decades later, most of these are not yet standard operating procedures in economic research.

At the time of the National Research Council's 1979 conference, the National Science Foundation's (NSF) policies embodied many of the Council's later recommendations. The NSF Grant Policy Manual NSF-77-47, as revised October 1979, states the following in paragraph 754.2:

> Data banks and software, produced with the assistance of NSF grants, having utility to others in addition to the grantee, shall be made available to users, at no cost to the grantee, by publication or, on request, by duplication or loan for reproduction by others...Any out of pocket expenses incurred by the grantee in providing information to third parties may be charged to the third party.

Subsequent to publication of Dewald, Thursby, and Anderson (1986), the NSF's social science program adopted a policy of requiring that investigators place data and software in a public archive after their award expired.[12] The NSF also began asking researchers, in applications for subsequent funding, what data and software from previous awards had been disseminated. Today, the NSF policy is clear. The NSF's current *Grant Proposal Guide* (NSF 04-23, effective September 2004), section VI, paragraph I, states that

> NSF advocates and encourages open scientific communication...It expects PIs [principal

[11] The published Goldberg-Saunders article used data through 1980:Q1; Dewald, Thursby, and Anderson collected data through the December 1982 issue of the *Survey*.

[12] I am indebted to Dan Newlon, head of the economics program at the National Science Foundation, for the information contained in this paragraph.

> investigators] to share with other researchers, at no more than incremental cost and within a reasonable time, the data, samples, physical collections and other supporting materials created or gathered in the course of the work. It also encourages grantees to share software and inventions, once appropriate protection for them has been secured, and otherwise act to make the innovations they embody widely useful and usable. NSF program management will implement these policies, in ways appropriate to field and circumstances, through the proposal review process; through award negotiations and conditions; and through appropriate support and incentives for data cleanup, documentation, dissemination, storage and the like. Adjustments and, where essential, exceptions may be allowed to safeguard the rights of individuals and subjects, the validity of results and the integrity of collections, or to accommodate legitimate interests of investigators.

The NSF's *Grant Policy Manual* (NSF 02-151, effective August 2, 2002), paragraph 734, "Dissemination and Sharing of Research Results," contains similar statements:

> b. Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing. Privileged or confidential information should be released only in a form that protects the privacy of individuals and subjects involved. General adjustments and, where essential, exceptions to this sharing expectation may be specified by the funding NSF Program or Division for a particular field or discipline to safeguard the rights of individuals and subjects, the validity of results, or the integrity of collections or to accommodate the legitimate interest of investigators. A grantee or investigator also may request a particular adjustment or exception from the cognizant NSF Program Officer.
>
> c. Investigators and grantees are encouraged to share software and inventions created under the grant or otherwise make them or their products widely available and usable.
>
> d. NSF normally allows grantees to retain principal legal rights to intellectual property developed under NSF grants to provide incentives for development and dissemination of inventions, software and publications that can enhance their usefulness, accessibility and upkeep. Such incentives do not, however, reduce the responsibility that investigators and organizations have as members of the scientific and engineering community, to make results, data and collections available to other researchers.

With such a strong policy in place, data warehousing in economic research should be commonplace. In fact, it remains rare. As of this writing, 10 professional economics journals have data and/or program archives: *American Economic Review*; *Econometrica*; *Macroeconomic Dynamics*; *Journal of Money, Credit, and Banking*; Federal Reserve Bank of St. Louis *Review*; *Economic Journal*; *Journal of Applied Econometrics*; *Review of Economic Studies*; *Journal of Political Economy*; and *Journal of Business and Economic Statistics*. Some require both data and program files, others only data. In addition, a public archive for data and programs from published articles in any professional journal has been maintained since 1995 by the Inter-university Consortium for Political and Social Research at the University of Michigan; except for articles related to the Panel Study of Income Dynamics at Michigan, all of the economics-related articles' data and programs in the archive are from the Federal Reserve Bank of St. Louis *Review*.

The ALFRED project has the potential, for macroeconomic research, to eliminate the need for journals to store authors' datasets. ALFRED, as explained further below, will sharply reduce the costs that a researcher in macroeconomics would bear in documenting, storing, and distributing the data used in a research project. In this aspect, data archives at journals are a "complementary technology" to the vintage archive structure of ALFRED, and, both in concept and execution, more similar to the dataset-archiving-and-indexing design of the VDC project.

DATA WAREHOUSING: THE FRASER AND ALFRED PROJECTS

The collection and distribution of data has classic public-good characteristics, including economies of scale, network effects, and first-mover advantages. Yet, large-scale systems for archiving and distributing economic data are rare.

As noted above, data warehousing for "real time" economic research began with the Federal Reserve Bank of Philadelphia (Croushore and Stark, 2001, p.112):

> In creating our real-time data set, our goal is to provide a basic foundation for research on issues related to data revision by allowing researchers to use a standard data set, rather than collecting real-time data themselves for every different study.

Two current data-warehousing projects of the Federal Reserve Bank of St. Louis are in the same spirit: FRASER® (Federal Reserve Archival System for Economic Research) and ALFRED. Together, these projects seek to provide a comprehensive archive of economic statistical publications and data. Initially, the projects will focus on government macroeconomic data but eventually will be expanded to include less aggregate data.

FRASER

The FRASER project (fraser.stlouisfed.org) is an Internet archive of images of statistical publications. The long-term goal, essentially, is to include all the statistical documents ever published by the U.S. government, plus other selected documents from both private and public sources. FRASER is an "open standards" project—that is, any organization that wishes to submit images is encouraged to do so provided that the images satisfy the requirements suggested by the U.S. Government Printing Office's committee of experts on digital preservation.[13] To date, however, most contributions have been boxes of printed paper materials, rather than images.

ALFRED

The ALFRED project is an archive of machine-readable real-time data. Since 1989, the Federal Reserve Bank of St. Louis has provided data to the public on its FRED system.[14] Initially, ALFRED will be populated with archived FRED data beginning December 1996. Later, other historical data will be added, including data extracted from FRASER images. Links will be provided between ALFRED data and the historical FRASER publications. The purpose of the FRED data system is to distribute the most-recent value for each variable and date; the purpose of the ALFRED system is to distribute in a similar method all data values previously entered into FRED, plus additional data. (Again, ALFRED is a work under construction; some features outlined here are part of the ALFRED architecture, and some are not yet available.)

ALFRED is a relational database built on PostgreSQL. Users are able to request subsets of data by means of an automated interface. In ALFRED, every data point is tagged with the name of its source and its publication date. For real-time projects, a researcher need only submit a variable list and a desired range of vintages (that is, "as of" dates); ALFRED will return all values for those variables that were available during the specified date range, each tagged with its as-of date (that is, its publication date). For replication studies, in the unlikely circumstance that the original researcher used the most recently published values for all variables and dates, a researcher need submit only a list of variable names and the as-of date when the original researcher collected his data. In the more common circumstance, in which the original researcher is uncertain whether he collected the then-most-recent data, a putative range of collection (as-of) dates may be submitted; with luck, some mixture of the retrieved values perhaps will reproduce the original published results.

Combined, the FRASER and ALFRED projects are a "statistical time machine" that, on request, furnishes to researchers both universes of data, χ, and time-indexed "real-time" subsets, N_{xt}.

[13] U.S. Government Printing Office (2004); Federal Reserve Bank of St. Louis (2004).

[14] Initially, FRED operated as a dial-up computer bulletin board system. In 1995, shortly after the release of version 1.0 of the Mosaic web browser, FRED appeared as a web site on the Internet.

As of this writing (October 2005), portions of ALFRED remain in development and not all features are implemented. Essential, however, will be a scheme to uniquely identify the historiography of data retrieved from ALFRED. The current design proposal includes the concept of a *research dataset signature*, or RDS. The proposed RDS is a human-readable string of ASCII characters that uniquely identifies a data series extracted from FRED or ALFRED. A dataset containing multiple time series (variables) will have an RDS for each series. The proposed character encoding pattern is as follows:

- 1-20: the FRED/ALFRED variable name[15];
- 21-22: the frequency code, e.g., "A", "Q", "BW";
- 23-25: the seasonal adjustment code, either the string "SA" or "NSA"[16];
- 26-34: the 24-hour Greenwich Mean Time date on which the data were downloaded, in the form DDMMMYYYY (that is, the form "27JAN2001");
- 35-42: the 24-hour Greenwich Mean Time time-of-day at which the data were downloaded, in the form HH:MM:SS (HH=hour, MM=minute, SS=second).

Greenwich Mean Time is used because it does not vary with the geographic location of the researcher or the season of the year.[17] These date and time-of-day formats are sufficiently general to accommodate researchers located anywhere in the world. Within each RDS string, character fields will be left-justified and padded on the right with spaces (ASCII 32). The RDS is somewhat shorter than might be anticipated because it does not include a "start" and "end" date. In FRED and ALFRED, users are not permitted to select/download a subset of a time series. A time series must be downloaded in its entirety or not at all. (The user is free to discard any unwanted data after download.) This permits a shorter signature.[18] Finally, as plain text, RDS strings may easily be included in working papers and journal articles.

The proposed architecture for ALFRED follows, in part, the bitemporal SQL (structured query language) database structure of Snodgrass and Jensen (1999).[19] In this design, each datum (observation) for each time series will be stored with three 2-element date vectors. One vector demarcates the beginning and end of the *measurement* interval for the observation, the second the beginning and end of the *validity* interval, and the third the beginning and end of the *transaction* interval.

Measurement intervals are straightforward. The measurement interval for GDP during 2004:Q1, for example, would be {1Jan2004, 31March2004}; a daily interest rate might have an interval of the form {5Jan2004, 5Jan2004}; and a monthly average interest rate might have an interval of the form {1Jan2004, 31Jan2004}. This system encompasses, in a uniform way, all data frequencies.

Validity intervals demarcate the time periods during which a datum was the most recently published value. As an example, consider 2004:Q1. See Table 1. During 2004, the Bureau of Economic Analysis published four measurements on 2004:Q1 GDP: April 29 ("advance"), May 27 ("preliminary"), June 25 ("final"), and July 30 (this fourth value has no commonly used label). During 2005, the BEA published one measurement, on July 29, as part of the 2005 benchmark revisions. The validity intervals shown reflect these dates. Note that the *fifth* validity interval is open-ended and will remain so until the next revised value is published.

[15] Currently, all FRED/ALFRED variable names are 8 characters in length and composed only of the characters 0 to 9 and A to Z—that is, ASCII characters 48 to 57 and 65 to 90. The signature's name field, to allow future expansion, is 20 characters in length and allows the underscore, ASCII 95, as well as 0 to 9 and A to Z.

[16] In the current FRED nomenclature, the seasonally adjusted character of the series can be inferred from the variable name. This field is included to increase the human usability of the signature and for possible future expansion of the FRED/ALFRED nomenclature.

[17] Greenwich Mean Time is named for the Royal Observatory at Greenwich, England. A discussion of Greenwich Mean Time is available at www.greenwichmeantime.com, as are conversions to local time zones.

[18] Internally, the database software distinguishes between dates on which a variable is not defined (such as the Federal Reserve's M2 monetary aggregate prior to 1959) and dates on which the series is defined (that is, was visible to observers monitoring the series at that date) but for which values are missing because, for example, certain printed publications cannot be located.

[19] The database design is due to George Essig, senior web developer, Federal Reserve Bank of St. Louis.

Table 1

Revision History for 2004:Q1 GDP, April 29 through December 29, 2004

Value	Variable name	Measurement interval	Validity interval	Transaction interval
11447.8	GDP	{1Jan2004, 31Mar2004}	{29Apr2004, 26May2004}	{29Apr2004, 31Dec9999}
11459.6	GDP	{1Jan2004, 31Mar2004}	{27May2004, 24Jun2004}	{27May2004, 31Dec9999}
11451.2	GDP	{1Jan2004, 31Mar2004}	{25Jun2004, 29Jul2004}	{25Jun2004, 31Dec9999}
11472.6	GDP	{1Jan2004, 31Mar2004}	{30Jul2004, 28Jul2005}	{30Jul2004, 31Dec9999}
11457.1	GDP	{1Jan2004, 31Mar2004}	{29Jul2005, 31Dec9999}	{29Jul2005, 31Dec9999}

Transaction intervals show dates on which Federal Reserve Bank of St. Louis staff entered or changed data values. The first date is the date on which the datum was added to the database. The ending date of the interval is infinity (open-ended) and will remain so indefinitely unless the datum (in the first column) is erroneous. Erroneous values in the database never are changed or removed—doing so would destroy the database's historical integrity. Once a datum has been made visible to web site customers, integrity of the database requires that the row is never modified or deleted. Instead, erroneous values are corrected by adding an additional row to the database for the same measurement and validity intervals. When a new row is added to correct an error, the end date of the erroneous row's validity interval will be set to the day on which the new row is added, and the start date of the new row's transaction interval will be set to the same date.[20] A customer selecting a date interval that includes the correction date will receive both the original erroneous datum and the corrected datum, plus a message warning that the observation for that date was corrected. The customer is responsible for checking his empirical results using both values.

The initial version of ALFRED will not include transaction intervals. This omission matters not at all so long as data never are changed after being made visible to researchers via the Internet. Because initial data will be machine-loaded from archival files, data entry errors and corrections are unlikely. Programming ALFRED using the three-interval architecture is significantly more difficult than with a two-interval (measurement, validity) design and would significantly lengthen ALFRED's development.

The ALFRED project will make it unnecessary to archive datasets for studies based on data obtained from FRED so long as the author retains the RDS signatures for the dataset.[21] Yet, what of the careless or forgetful researcher who does not retain the RDS? For them, the current design includes automatic archiving and retrieval of RDS data if the researcher signs up for a user account. After so doing, they will be offered the opportunity to save, on FRED and ALFRED, the RDS strings for every series they download. Putative replicators need only ask the original researcher to retrieve the RDS strings and make them available.

The ALFRED system, when completed, promises to unify the concepts of real-time data and replication in economic research.

[20] Once created, the integrity of the database requires that every row be retained in the database; otherwise, entering the same data signature string on different dates would retrieve different data, which is unacceptable. Including a transaction interval is an essential design element in a database system that guarantees time-invariance of retrieved data when the data are subject both to revision by the publisher and to possible human data-entry error.

[21] An additional, related part of the FRASER/ALFRED project, nearing completion, is a catalog of available federal, state, and local data series. In its intent and structure, the catalog resembles the "Statistical Knowledge Network" discussed by Hert, Denn, and Haas (2004) except that the catalog does not attempt to explain or instruct in the ways that the data might be used to conduct economic analyses, as these authors suggest their metadata might be able to do. Also, to the extent that descriptive metadata is stored as XML tags, the design of Hert, Denn, and Haas is compatible with the VDC/DDI initiative to cross-index data from various servers on the Internet. Further, since the St. Louis economic data catalog will index data from all government agencies, it partially circumvents the barrier to cross-government IT collaboration discussed by Mullen (2003), although the *GPO Access* project (U.S. Government Printing Office, 2001) has been charged by the Congress to promote electronic dissemination of data and documents.

THE VDC PROJECT

The VDC project of Harvard University, similar to ALFRED, has as its goal increasing the replicability of research.[22] Unlike ALFRED, however, the VDC project itself does not include data collection. Rather, the heart of the VDC project is to provide a low-cost, integrated suite of software that will allow other researchers a forum for archiving and sharing data. More precisely, the VDC project furnishes "an (OSS) [open-source software] digital library system 'in a box' for numeric data" that "provides a complete system for the management, dissemination, exchange, and citation of virtual collections of quantitative data."[23]

An essential component of the VDC project's architecture are the tools to encourage researchers—including individuals, professional journals, and research institutions—to use a single set of formatting and labeling standards when they place datasets on the Internet. In turn, a loosely coupled web of VDC servers will locate, index, and catalog the datasets, making them available to other researchers. The VDC's proposed formatting and labeling standards are those of the University of Michigan's Data Documentation Initiative (DDI) project, an accepted standard in the document and knowledge management arena.[24] The formatting consists solely of inserting plain text XML tags within text data files, easily done within many programs or a simple text editor.

If successful, the VDC project promises the type of network effects, well-known to economists, that accompany (and drive) the adoption of standards. Because each VDC node maintains a catalog of materials held on other VDC nodes, as the VDC network expands, additional researchers will find it increasingly attractive to join, so as to make their work visible to the growing community.

[22] Altman et al. (2001, p. 464).

[23] See http://thedata.org/. The VDC and St. Louis projects share the same core open-source components: Linux, Apache, and PostgresSQL. The St. Louis middleware is coded as server-side PHP scripts, similar in spirit if not code to the Java servlets used in VDC.

[24] On DDI, see Blank and Rasmussen (2004) and www/icpsr.umich.edu/DDI/.

A COMPARISON OF PROJECTS: ALFRED AND VDC

Both the VDC and ALFRED projects provide tools that promise to improve the scientific quality of empirical economic research. But, their philosophies and architectures differ. Altman et al., for example, writes that "The basic object managed in the [VDC] system is the *study*" [emphasis added]. In the St. Louis ALFRED project, the basic objects managed are a published study's set of signature strings, which permit repeated extraction of the same dataset from an underlying, encompassing database.

Because the VDC project focuses on preserving specific datasets from specific studies, it is well-suited for archiving both experimental and non-experimental data. The ALFRED project's focus on archiving vintages of non-experimental data makes it better suited to macroeconomic research, both real-time and replication studies. In *replication* studies, the issue is determining which data were used in a study and whether the calculations were performed as described; in *real-time* data studies, the issue is determining the robustness of the study's findings to data revisions. At least for aggregate macroeconomic data, an archival system that can do two things—provide a later investigator with the previous researcher's original data *as well as* provide earlier and later published values of the same variables—has the promise of combining a "simple" replication study with a real-time, data-based robustness study. In Pesaran and Timmermann's notation, the archival system must be able to produce, on demand, both the universe of all observations on the variables of interest, χ, and all possible time-indexed "real time" subsets, N_{xt}. Although careful use of XML tags in a VDC/DDI system might permit support for such real-time econometrics, it likely would require attaching XML tags to each data point in each study.

CONCLUSION

Data archiving, data sharing, and replication are hallmarks of science, necessary to explore the correctness of published results. Real-time data

studies are important to address the robustness of published results. These two lines of inquiry are linked by means of their recognition that empirical economic research is inherently time-indexed. Although quite different, the VDC and ALFRED projects promise to assist and improve the quality of empirical economic research by reducing the cost of both lines of inquiry.

REFERENCES

Altman, Micah L.; Diggory, Andreev M.; King, G.; Sone, A.; Verba, S.; Kiskis, Daniel L. and Krot, M. "A Digital Library for the Dissemination and Replication of Quantitative Social Science Research: The Virtual Data Center." Social Science Computer Review, 2001, 19, pp. 458-70.

Anderson, Richard G.; Greene, William H.; McCullough, Bruce M. and Vinod, H.K. "The Role of Data & Program Code Archives in the Future of Economic Research." Working Paper 2005-14, Federal Reserve Bank of St. Louis, 2005.

Anderson, Richard G. and Dewald, William G. "Replication and Scientific Standards in Applied Economics a Decade After the Journal of Money, Credit and Banking Project." Federal Reserve Bank of St. Louis *Review*, November/December 1994, *76*(6), pp. 79-83.

Bailar, John C. "The Role of Data Access in Scientific Replication." Presented at the October 16-17, 2003 workshop "Confidential Data Access for Research Purposes," held by the Panel on Confidential Data Access for Research Purposes, Committee on National Statistics, National Research Council, 2003; www7.nationalacademies.org/cnstat/John_Bailar.pdf.

Bernanke, Ben S. and Boivin, Jean. "Monetary Policy in a Data-Rich Environment." *Journal of Monetary Economics*, 2003, *50*, pp. 525-46.

Blank, Grant and Rasmussen, Karsten Boye. "The Data Documentation Initiative." *Social Science Computer Review*, Fall 2004, *22*(3), pp. 307-18.

Bornstein, Robert F. "Publication Politics, Experimenter Bias and the Replication Process in Social Science Research," in James Neuliep, ed., *Replication Research in the Social Sciences*. Thousand Oaks, CA: Sage Publications, 1991, pp. 71-84.

Boruch, Robert F. and Cordray, David S. "Professional Codes and Guidelines in Data Sharing," in Stephen E. Fienberg, Margaret E. Martin, and Miron L. Starf, eds. *Sharing Research Data*. Washington DC: National Academy Press, 1985; www.nap.edu/openbook/030903499X/html/index.html.

Christoffersen, Peter; Ghysels, Eric and Swanson, Norman R. "Let's Get 'Real' About Using Economic Data." *Journal of Empirical Finance*, 2002, *9*, pp. 343-60.

Clark, Todd E. and Kozicki, Sharon. "Estimating Equilibrium Real Interest Rates in Real Time." Working Paper 04-08, Federal Reserve Bank of Kansas City, September 2004.

Crawford, Vincent P. and Sobel, Joel. "Strategic Information Transmission." *Econometrica*, November 1982, *50*(6).

Croushore, Dean and Stark, Tom. "A Real-Time Data Set for Macroeconomists." *Journal of Econometrics*, 2001, *105*, pp. 111-30.

Dewald, William G.; Thursby, Jerry G. and Anderson, Richard G. "Replication in Empirical Economics: The *Journal of Money, Credit and Banking* Project." *American Economic Review*, September 1986, *76*(4), pp. 587-603.

Faust, Jon; Rogers, John H. and Wright, Jonathan H.. "Exchange Rate Forecasting: The Errors We've Really Made." *Journal of International Economics*, May 2003, *60*(1), pp. 35-59.

Federal Reserve Bank of St. Louis. "As It Happened: Economic Data and Publications as Snapshots in Time," presentation by Robert Rasche, Katrina Stierholz, Robert Suriano, and Julie Knoll, at the Fall Federal Depository Librarians Conference, October 19, 2004, Washington, DC; fraser.stlouisfed.org/fdlp_final.pdf.

Feigenbaum, S. and Levy, D. "The Market for (Ir)Reproducible Econometrics" and "Response to the Commentaries." *Social Epistemology*, 1993, *7*(3), pp. 215-232 and pp. 286-92.

Fienberg, Stephen E.; Martin, Margaret E. and Starf, Miron L., eds. *Sharing Research Data*. Washington, DC: National Academy Press, 1985; www.nap.edu/openbook/030903499X/html/index.html.

Goldberg, Lawrence G. and Saunders, Anthony. "The Growth of Organizational Forms of Foreign Banks in the U.S." *Journal of Money, Credit, and Banking*, August 1981, pp. 365-74.

Hert, Carol A.; Denn, Sheila and Haas, Stephanie W. "The Role of Metadata in the Statistical Knowledge Network." *Social Science Computer Review*, Spring 2004, *22*(1), pp. 92-99.

Kishor, N. Kundan and Koenig, Evan F. "VAR Estimation and Forecasting When Data Are Subject to Revision." Working Paper 2005-01, Federal Reserve Bank of Dallas, February 2005.

Koenig, Evan F.; Dolmas, Sheila and Piger, Jeremy. "The Use and Abuse of Real-Time Data in Economic Forecasting." *Review of Economics and Statistics*, 2003, *85*, pp. 618-28.

Mullen, Patrick R. "The Need for Government-Wide Information Capacity." *Social Science Computer Review*, Winter 2003, *21*(4), pp. 456-63.

National Research Council. *Access to Research Data in the 21st Century: An Opening Dialogue Among Interested Parties*. Report of a workshop on the Shelby Amendment held by the Science, Technology and Law Panel of the National Research Council, March 12, 2001. Washington, DC: National Academy Press, 2002.

Neely, Christopher J.; Roy, Amlan and Whiteman, Charles H. "Risk Aversion versus Intertemporal Substitution: A Case Study of Identification Failure in the Intertemporal Consumption Capital Asset Pricing Model." *Journal of Business and Economic Statistics*, October 2001, *19*(4), pp. 395-403.

Orphanides, Athanasios. "Monetary Policy Rules Based on Real-Time Data." *American Economic Review*, September 2001, *91*(4), pp. 964-85.

Orphanides, Athanasios and van Norden, Simon. "The Unreliability of Output-Gap Estimates in Real Time." *Review of Economics and Statistics*, November 2002, pp. 569-83.

Orphanides, Athanasios and van Norden, Simon. "The Reliability of Inflation Forecasts Based on Output Gap Estimates in Real Time." Working Paper 2003s-01, CIRANO, HEC, Montreal, 2003.

Pesaran, Hashem and Timmermann, Allan. "Real-Time Econometrics." *Econometric Theory*, 2005, *21*(1), pp. 212-31.

Snodgrass, Richard T. and Jensen, Christian S. *Developing Time-Oriented Database Applications in SQL*. San Francisco CA: Morgan Kaufman, 1999.

Stark, Thomas and Croushore, Dean. "Forecasting with a Real-Time Dataset for Macroeconomists." *Journal of Macroeconomics*, 2002, *24*, pp. 507-31.

Svensson, Lars E.O. and Woodford, Michael. "Indicator Variables for Optimal Policy." *Journal of Monetary Economics*, 2004, *50*, pp. 691-720.

Svensson, Lars E.O. and Woodford, Michael. "Indicator Variables for Optimal Policy Under Asymmetric Information." *Journal of Economic Dynamics and Control*, 2003, *28*, pp. 661-90.

U.S. Government Printing Office. *Biennial Report to Congress on the Status of GPO Access*. Washington, DC: U.S. GPO, 2001; www.gpoaccess.gov/biennial/index.html.

U.S. Government Printing Office. *Report on the Meeting of Experts on Digital Preservation*. Washington, DC: U.S. GPO, March 12, 2004a; www.gpoaccess.gov/about/reports/preservation.pdf.

U.S. Government Printing Office. *Concept of Operations for the Future Digital System*. Washington, DC: U.S. GPO, October 1, 2004; www.gpo.gov/news/2004/ConOps_1004.pdf.

APPENDIX

Recommendations of the Committee on National Statistics of the National Research Council and the National Academy of Sciences

The Committee on National Statistics' final report (Fienberg, Martin, and Starf, 1985) offered 16 specific recommendations regarding sharing research data. These recommendations, little noticed during the past 20 years, are as relevant today as then. Taken together, they form a foundation, or body of knowledge, for best-practice in empirical scientific research. Their recommendations are reproduced here because, although they sound scientific and sensible, most have been ignored in economic science.

For all researchers:

1. Sharing data should be a regular practice.

For initial investigators:

2. Investigators should share their data by the time of publication of initial major results of analyses of the data except in compelling circumstances.
3. Data relevant to public policy should be shared as quickly and widely as possible.
4. Plans for data sharing should be an integral part of a research plan whenever data sharing is feasible.
5. Investigators should keep data available for a reasonable period after publication of results from analyses of the data.

For subsequent analysts:

6. Subsequent analysts who request data from others should bear the associated incremental costs.
7. Subsequent analysts should endeavor to keep the burdens of data sharing on initial investigators to a minimum and explicitly acknowledge the contribution of the initial investigators.

For institutions that fund research:

8. Funding organizations should encourage data sharing by careful consideration and review of plans to do so in applications for research funds.
9. Organizations funding large-scale, general-purpose data sets should be alert to the need for data archives and consider encouraging such archives where a significant need is not now being met.

For editors of scientific journals:

10. Journal editors should require authors to provide access to data during the peer review process.
11. Journals should give more emphasis to reports of secondary analyses and to replications.
12. Journals should require full credit and appropriate citations to original data collections in reports based on secondary analyses.
13. Journals should strongly encourage authors to make detailed data accessible to other researchers.

For other institutions:

14. Opportunities to provide training on data-sharing principles and practices should be pursued and expanded.
15. A comprehensive reference service for computer-readable social science data should be developed.
16. Institutions and organizations through which scientists are rewarded should recognize the contributions of appropriate data-sharing practices.

Federal Reserve Bank of St. Louis
P.O. Box 442
St. Louis, MO 63166-0442