Recent Developments in the Analysis of Monetary Policy Rules

Bennett T. McCallum

It is a great privilege for me to be giving this year's Homer Jones Memorial Lecture, in recognition of Homer Jones's outstanding role in the development of monetary policy analysis. I did not know him personally, but I have been very strongly influenced by economists who knew and admired him greatly—Karl Brunner, Milton Friedman, and Allan Meltzer come to mind immediately. My work has also been influenced by writings coming from the research department of the Federal Reserve Bank of St. Louis, which he directed, and by the availability of monetary data series developed there.

For this lecture I originally had planned a title of "The Evolution of Monetary Policy Analysis, 1973-1998." As it happens, I have decided to place more emphasis on today's situation and less on its evolution. But, a few words about history may be appropriate. I had chosen 1973 as the starting point for a review because there was a sharp break in both academic analysis and in real-world monetary institutions during the period around 1971-73. Regarding institutions, of course, I am referring to the breakdown of the Bretton Woods exchange-rate system, which was catalyzed by the U.S. government's decision in August 1971 not to supply gold to other nations' central banks at $35 per ounce. This abandonment of the system's nominal anchor naturally led other nations to be unwilling to continue to peg their currency values to the (overvalued) U.S. dollar, so the par-value arrangements disintegrated. New par values were painfully established during the December 1971 meeting at the Smithsonian Institution, but after a new crisis, the system crumbled in March 1973.

In terms of monetary analysis, the starting date of 1973 has the disadvantage of missing the publication in 1968 and 1970 of the Andersen-Jordan (1968) and Andersen-Carlson (1970) studies, which many of you will know were written at the St. Louis Fed under the directorship of Homer Jones. These studies were, to an extent, a follow-up to the Friedman-Meiselman (1963) paper, which had set off a period of intellectual warfare between economists of a then-standard Keynesian persuasion and those who were shortly (Brunner, 1968) to be termed "monetarists."[1] But my reason for beginning slightly later is that the years 1971-73 featured the publication of six papers that initiated the rational expectations revolution. The most celebrated of these is Lucas's (1972a) "Expectations and the Neutrality of Money," but his other papers (1972b) and (1973) also were extremely influential as were Sargent's (1971 and 1973). The sixth paper is Walters (1971), which had little influence but was, I believe, the first publication to use rational expectations (RE) in a macro-monetary analysis.

At first there was much resistance to the RE hypothesis, partly because it initially was associated with the policy-ineffectiveness proposition. But, it gradually swept the field in both macro and microeconomics, primarily because it seems extremely imprudent for policy analysis to be conducted under the assumption that any particular pattern of expectational errors will prevail in the future—and ruling out all such patterns implies RE.

There were other misconceptions regarding rational expectations, the most prominent of which was that Lucas's famous "critique" paper (1976) demonstrated that policy analysis with econometric models

Table 1

Conference Contributors

	NBER Conference Jan. 17-18, 1998	
	Papers	Discussants
Academic	5.5	8
Central Bank	3.5	1
Total	9	9

	Riksbank-IIES Conference, June 12-13, 1998		
	Papers	Discussants	Panelists
Academic	5	1	2
Central Bank	2	6	3
Total	7	7	5

was a fundamentally flawed undertaking. Actually, of course, Lucas and Sargent showed instead that certain techniques were flawed, if expectations are indeed rational, and that more sophisticated techniques are called for. But by 1979, John Taylor, last year's Homer Jones lecturer, had demonstrated that these techniques are entirely feasible. Nevertheless, this misunderstanding—and others concerning the role of money[2]—led to a long period during which there was a great falling off in the volume of sophisticated, yet practical, monetary policy analysis. One reason was the upsurge of the real-business-cycle (RBC) approach to macroeconomic analysis, which in its standard version assumes that price adjustments take place so quickly that, for practical purposes, there is continuous market clearing for all commodities, including labor. In this case, monetary policy actions will, in most models, have little or no effect on real macroeconomic variables at cyclical frequencies. Of course this has been a highly controversial hypothesis and I am on record as finding it quite dubious (McCallum, 1989). But my attitude is not altogether negative about RBC analysis because much of it has been devoted to the development of new theoretical and empirical tools, ones that can be employed

without any necessary acceptance of the RBC hypothesis about the source of cyclical fluctuations.

In recent years, in fact, these tools have been applied in a highly promising fashion. Thus a major movement has been underway to construct, estimate, and simulate monetary models in which the economic actors are depicted as solving dynamic optimization problems and then interacting on competitive markets,[3] as in the RBC literature, but with some form of nominal price and/or wage "stickiness" built into the structure. The match between these models and actual data is then investigated, often by standard RBC procedures, for both real and monetary variables and their interactions. The objective of this line of work is to combine the theoretical discipline of RBC analysis with the greater empirical validity made possible by the assumption that prices do not adjust instantaneously. Basically, the attempt is to develop a model that is truly structural, immune to the Lucas critique, and appropriate for policy analysis.

As a consequence of this movement, and some other activities to be mentioned shortly, the state of monetary policy analysis today (March 1999) is remarkably different than it was only a few years ago. Most of the changes are clearly welcome improvements, although some are of more debatable merit. Let me now describe central aspects of the current situation before turning to an evaluation and an application.

One striking feature of research on monetary policy today is the extent of interaction between central-bank and academic economists and the resulting similarity of the research conducted. This feature is illustrated nicely by the contributions to two recent conferences entitled "Monetary Policy Rules." The first of these was sponsored by the National Bureau of Economic Research (NBER), held January 17-18, 1998, in Islamorada, Florida. The second, held June 12-13, 1998, in Stockholm, was jointly sponsored by the Sveriges Riksbank (the Swedish central bank) and the Institute for International Economic Studies at Stockholm University.

[2] Here I have in mind the promotion of a class of overlapping-generations models in which the asset termed money plays no medium-of-exchange role.

[3] Actually, writings in this literature typically express their analysis as pertaining to economies featuring monopolistic competition. In typical cases, most of the results are independent of the extent of monopoly power, which then could be virtually zero.

In Table 1, the figures on contributors clearly indicate that both academic and central-bank participation was substantial, but they do not begin to tell the whole story. They do not show, for example, that four of the papers were authored jointly by one economist from each group. Nor do they reveal that two of the designated academics were central bankers until very recently; that three others had (like the St. Louis Fed's William Poole and Robert Rasche) moved in the opposite direction; or that currently one is both a leading professor and a member of the Bank of England's Monetary Policy Board. The fact that several academic participants are regular central-bank consultants is also not shown.

But to get the full flavor of the extent to which central-bank and academic monetary analysis has done away with distinctions that were important only recently, one needs to read the papers. It is my impression that if the authors' names were removed, one would find it extremely difficult to tell which group the author or authors came from. To me, this intense interaction seems to represent a very positive change, and is one toward which several regional Federal Reserve Banks (including St. Louis) have contributed greatly.

In the research presented at these two conferences there was not just a similarity of technique across groups, but also a considerable amount of agreement across authors about the outline of an appropriate framework for the analysis of monetary policy issues. Such agreement can be dangerous, of course, but it certainly facilitates communication. In fact, there remains room for quite a bit of substantive disagreement within the framework, so on balance I find this similarity somewhat encouraging. In any event, I would like to describe this framework and then take up some major issues that I hope you will find interesting.

The nearly standard framework at the NBER and Riksbank conferences is a quantitative macroeconomic model that includes three main components. These are:

- An IS-type relation (or set of relations) that specifies how interest-rate movements affect aggregate demand and output;

- A price adjustment equation (or set of equations) that specifies how inflation behaves in response to the output gap and to expectations regarding future inflation; and
- A monetary policy rule that specifies each period's settings of an interest-rate instrument.

These settings typically are made in response to recent or predicted values of the economy's inflation rate and its output gap. A leading example of such a rule will be considered at length shortly. Most of these are quarterly models and most incorporate rational expectations. They are estimated by various methods, including the approach called "calibration," but in all cases an attempt is made to produce a quantitative model in which parameter values are consistent with actual time-series data for the United States or some other economy. These models are intended to be structural (i.e., policy invariant) and in some cases this attempt is enhanced by a modeling strategy that features explicit optimization by individual agents acting in a dynamic and stochastic environment.

To study effects of policy behavior, stochastic simulations are conducted using the model at hand with alternative policy rules, with summary statistics being calculated to represent performance measured by average values of the variability of inflation, the output gap, and interest rates. A few of the models are constructed so that each simulation implies a utility level for the representative individual agent; in such cases, utility-based performance measures can be calculated. In several studies, effort is taken to make the policy rules *operational*, which, with an interest instrument, means a realistic specification of information available to the central bank when setting its instrument.

In discussing in more detail the components of this framework, it will be useful to have an algebraic representation of a simple special case. Here I will use y_t to denote the natural logarithm of real gross domestic product (GDP) during quarter t, with $\bar{y}_t$ being the capacity or potential or natural rate value of y_t. Then $\tilde{y}_t = y_t - \bar{y}_t$ is the output gap. Also p_t is the log of

the price level so Δp_t is the inflation rate while g_t represents real government purchases and R_t is the level of the short-term nominal interest rate used as the central bank's instrument.

$$(1) \quad y_t = \beta_0 + \beta_1 E_t y_{t+1} + \beta_2 (R_t - E_t \Delta p_{t+1}) + \beta_3 (g_t - E_t g_{t+1}) + v_t$$

$$(2) \quad \Delta p_t = \alpha_1 E_t \Delta p_{t+1} + (1-\alpha_1) \Delta p_{t-1} + \alpha_2 (y_t - \bar{y}_t) + u_t$$

$$(3) \quad R_t = \bar{r} + E_t \Delta p_{t+j} + \mu_1 (E_t \Delta p_{t+j} - \pi^*) + \mu_2 (y_t - \bar{y}_t) + e_t$$

Here $E_t z_{t+j}$ is the rationally formed expectation at time t of the value of z that will prevail in period $t+j$, so $E_t \Delta p_{t+1}$ is the expected inflation rate and $R_t - E_t \Delta p_{t+1}$ is the one-period real rate of interest. The terms v_t, u_t, and e_t represent random disturbance factors that impinge on the choices of individuals and the central bank; these are not observable to an econometrician. The parameters designated β, α, and μ do not change with time, unlike the variables that carry the subscript t. All parameters except β_2 are presumed to be positive.

Relation 1 is a so-called IS function in which β_2 is a negative number, reflecting the hypothesis that the real rate of interest has a negative effect on demand; higher real interest rates tend to depress spending by households and firms. If $\beta_1 = 0$, then the IS function would be one of the textbook Keynesian variety that is somewhat lacking in theoretical justification. With $\beta_1 = 1$, however, we have a forward-looking "expectational" or "intertemporal" IS relation of the type several authors have shown to be implied, under reasonable conditions, by optimizing dynamic behavior.[4] With this latter type of relationship, the proper appearance of government purchases is as shown in equation 1. This is of some interest, for it implies that if changes in g_t are approximately permanent, then an upward jump in g_t will be offset by an upward jump in $E_t g_{t+1}$, leaving demand unaffected. That type of phenomenon may be the reason that many investigators have obtained econometric results

suggesting that government purchases have insignificant explanatory power for aggregate demand.

The price-adjustment equation 2 is written so as to accommodate either the entirely forward-looking Calvo-Rotemberg model,[5] in which case $\alpha_1 = 1$, or a two-period version of the Fuhrer and Moore (1995) model (with $\alpha_1 = 0.5$). Neither of these, I would point out, satisfies the strict version of the natural rate hypothesis (NRH) due to Lucas (1972b), which postulates that monetary policy cannot keep $y_t > \bar{y}_t$ permanently by *any* sustained scheme of behavior. (More precisely, the NRH implies that $E(y_t - \bar{y}_t) = 0$ for any policy rule.[6]) I personally consider this violation to be a weakness, an indication that specification 2 is faulty.[7] But both the Calvo-Rotemberg and Fuhrer-Moore models are more attractive (and plausible) in that regard than the NAIRU class,[8] which gets more attention from the press and practical commentators, for the latter class implies that an increasing inflation rate will keep output high forever (in contrast to either of the mentioned versions of 2). That the press—and even some professional publications[9]—fails to distinguish between the NRH and the NAIRU concept is, in my opinion, slightly disgraceful, especially since the very term NAIRU suggests an incompatibility with the NRH.[10]

The third component of this simple system is the monetary policy rule that is shown in equation 3. It suggests that with μ_1 and μ_2 positive the central bank will raise R_t, thereby tightening policy, when inflation exceeds its target value π^* and/or when output is high relative to capacity. Thus equation 3 has been written in approximately the form suggested by Taylor (1993), which has come to be known as "the Taylor rule." I will have quite a bit to say about that rule below, but for the moment I wish to take up the point that the system (equations 1-3) does not include a money demand-function. Indeed, it does not refer to any monetary quantity measure in any way whatsoever. To anyone steeped in the tradition of Homer Jones, this strikes a rather dissonant note. So let's take a minute to consider whether this is sensible.

---

[4] These authors include Kerr and King (1996), McCallum and Nelson (1999), and Woodford (1995).

[5] The references are Calvo (1983) and Rotemberg (1982).

[6] It can be verified easily that equation 2 implies that if policy generates inflation such that $E(\Delta p_t - \Delta p_{t-1}) \neq 0$, then $E(y_t - \bar{y}_t) \neq 0$.

[7] One of the few relations with price stickiness that satisfies the NRH is my own favorite, the P-bar model used by McCallum and Nelson (1999). Its weakness is that it does not yield as much persistence in inflation as appears in the data.

[8] Typified by $\Delta p_t = \alpha_1 \Delta p_{t-1} + (1-\alpha_1)\Delta p_{t-2} + \alpha_2(y_t - \bar{y}_t) + u_t$.

[9] See, e.g., the symposium in the winter 1997 issue of the *Journal of Economic Perspectives*.

[10] The term "non-accelerating-inflation rate of unemployment" suggests a relationship between $\Delta p_t - \Delta p_{t-1}$ and $y_t - \bar{y}_t$ in immediate contradiction to the NRH.

To do that, suppose that we add to the system a standard money demand function. Let m_t be the log of the money stock, either the monetary base or M1 depending on whether or not a banking sector behavior is included. Then we have

(4) $\qquad m_t - p_t = \gamma_0 + \gamma_1 y_t + \gamma_2 R_t + \varepsilon_t$

where ε_t is the random component of money demand. Here y_t is a proxy measure of the transactions that money facilitates and R_t is an (overly simple) measure of the opportunity cost of holding money rather than some other asset. In an actual application, some account might have to be taken of technical progress in the payments process, but for present purposes that complication is unnecessary. The first basic point to be made is that if we append equation 4 to the system (equations 1-3), it plays no essential role. It merely determines how much money has to be supplied by the central bank in order to implement its interest rate policy rule, equation 3. The system (equations 1-3) determines the same values for Δp_t, y_t, and R_t whether equation 4 is recognized or not, presuming that $\bar{y}_t$ and g_t are exogenously given. This is the basic point that has led many researchers to ignore money and, indeed, that has led the staff of the Fed's Board of Governors to construct a large, sophisticated, and expensive new macroeconometric model that does not recognize money in any capacity.[11] But is the point *valid*?

Evidently, there are at least two requirements for it to be valid. First, the central bank of the economy being modeled actually conducts policy by manipulating a real-world counterpart of R_t, while paying no decisive attention to current movements in m_t. It is widely agreed that this is the case for the United States and most other industrialized nations, including Germany.[12] Second, it must be the case that m_t does not appear in correctly specified versions of either equations 1 or 2. With respect to the latter, that condition would seem to be satisfied; but for the expectational IS function 1 it is more problematic. What is required in a mainstream theoretical analysis[13] is that the

transaction-cost function, which describes the way that money (the medium of exchange) facilitates transactions, must be *separable* in m_t and the spending variable such as y_t. But there is no theoretical reason for that to be the case and it clearly is *not* the case for my own preferred specification. So what is actually being assumed implicitly, by analyses that exclude m_t (i.e., $m_t - p_t$) from the relation 1, is that the effects of money holdings on spending are quantitatively small (indeed negligible). This is a belief with a long tradition, and I am inclined to think that it is probably justifiable, but the whole matter needs additional study.

One of the fortuitous events that led to today's era of cooperation between central-bank and academic economists was the publication of a 1993 paper by John Taylor—the one in which he explicitly proposed the now famous Taylor rule. By writing his rule in terms of the instrument actually used by central banks and expressing his formula with brilliant simplicity, Taylor made the concept of a monetary rule more palatable to central bankers—especially as he showed that recent U.S. experience had in fact conformed to his formula rather closely.[14] Simultaneously, the step was attractive to academics because it enabled them both to simplify their analysis, by discarding money demand functions, and also to be more realistic.

The precise rule proposed by Taylor (1993) for the U.S. economy is as follows:

(5) $R_t = \Delta p_t^a + 0.5\ (\Delta p_t^a - \pi^*) + 0.5\ \tilde{y}_t + \bar{r}.$

Here Δp_t^a is the average inflation rate over the past four quarters—a proxy for expected inflation—and $\tilde{y}_t$ is $y_t - \bar{y}_t$, the output gap. For $\bar{r}$, the average *real* rate of interest, Taylor assumed 2 percent (per year) and for the inflation target π^* he also assumed 2 percent. So he actually wrote the expression, with p denoting inflation, y denoting $\tilde{y}$, and r instead of R, as follows: $r = p + 0.5y + 0.5(p - 2) + 2$. In thinking about this rule, it is important to recognize that it does not involve the fallacy of using a nominal interest rate as an indicator of monetary tightness or ease.

[11] See Brayton, et. al. (1997).

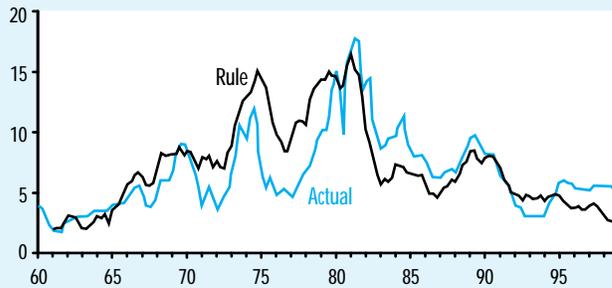[12] On this point, see Clarida and Gertler (1996).

[13] Such as that of Walsh (1998) or McCallum and Goodfriend (1987).

[14] It also helped, I am sure, that he emphasized that rule-like behavior does not require literal, strict adherence to a specified formula.

Figure 1

Taylor Rule and Actual Values for U.S., 1961-98

Federal Funds Rate

[Figure: line chart comparing "Rule" and "Actual" federal funds rate from 1960 to 1998, y-axis from 0 to 20, x-axis years 60, 65, 70, 75, 80, 85, 90, 95.]

Rather, it compares the real rate $R_t - \Delta p_t^a$ with its long-run equilibrium value $\bar{r}$ and adjusts the former upward if the current situation, represented by $0.5\ (\Delta p_t^a - \pi^*) + 0.5\ \tilde{y}_t$, calls for a tighter stance.

To illustrate the workings of the Taylor rule we can look at a diagram, similar to one recently constructed by Taylor (1999), that compares actual historical values of the U.S. federal funds rate with values that would have been dictated by the rule during the years 1960-98. In Figure 1 we see that the two curves agree very closely during the years 1987-94, but disagree sharply for the period from 1965-78, with the Taylor rule calling for much tighter policy through most of that period. Both of these comparisons are quite encouraging for the Taylor rule, for most analysts would now agree that U.S. policy was quite good during 1987-94 and considerably too loose during 1965-78.

If you find the Taylor rule interesting, you can always keep up to date on its advice by going to the web site of the St. Louis Fed. In the publication entitled *Monetary Trends*, the bank plots a different but related diagram that shows what the implicit inflation target of the Fed has been recently, according to the Taylor rule, and recent values of the federal funds rate. The diagram available in February 1999 shows that as of mid-1998 the implicit target was about 1 percent inflation. Thus, the Taylor rule indicates that the recent U.S. monetary stance has been slightly more restrictive than one that would yield 2 percent inflation, the value that most analysts consider to best represent the Fed's actual (although unstated) inflation target.

On the same page of *Monetary Trends* there is another chart that pertains to a different rule, one that I am happy to say is known as the McCallum rule. It is entirely appropriate that my rule appears after Taylor's, because his is much more popular with both central bankers and academics. A major reason is that mine is expressed in terms of settings for the growth rate of the adjusted monetary base—currency plus bank reserves—rather than any interest rate. Therefore, Taylor's is much more realistic in the sense of pertaining to the central bank's actual instrument variable. In fact, many central bankers view discussions of the monetary base with about the same enthusiasm as I would have for the prospect of being locked in a telephone booth with someone who had a bad cold, or some other infectious disease.

That does not necessarily mean, however, that a base-oriented rule will give poorer advice concerning monetary policy. Historically, my rule—which adjusts the base growth rate up or down when nominal GDP growth is below or above a chosen target value[15]—has agreed with Taylor's over many periods. But, they differed in the United Kingdom during the late 1980s when mine would have called for tighter policy and Taylor's for looser. Since that was a period during which U.K. inflation rose rather rapidly—after having been temporarily subdued by the onslaught of Margaret Thatcher—this episode is one that can be pointed out, when I want to argue the merits of my rule.

I also must say that it would be very wrong to interpret this contrast of rules as representing a dispute between Taylor and me. I believe that the two of us are striving for basically the same policy goals: a stable, rule-like monetary policy designed to keep inflation low and to do what little it can to stabilize real output fluctuations. Furthermore, I am confident that he shares this belief. And I certainly have no hesitation in saying that he has been the more effective spokesman for our cause.

[15] The target value Δx^* equals the desired average rate of inflation plus the expected long-run average rate of growth of real output—say, $2.0 + 2.5 = 4.5$ percent per year (or 0.01125 in quarterly fractional units). Then the rule is $\Delta b_t = \Delta x^* - \Delta v_t^a + 0.5(x_{t-1}^* - x_{t-1})$ where b_t and x_t are logs of the base and nominal GDP while Δv_t^a is the average rate of growth of base velocity over the previous four years. Also, x_t^* is the target value of x_t for period t, equal to $x_{t-1} + \Delta x^*$.

That said, in closing I would like to apply our two rules to the extremely important case of Japan during the 1990s. To do this with the Taylor rule requires us to adopt values for π^* and $\bar{r}$, the inflation target and the long-run average real interest rate. For the former, I again will take 2 percent in measured terms (which probably overstates the actual inflation rate in Japan by about 1 percent). For $\bar{r}$, Taylor's (1993) procedure was to use a number close to the long-run average rate of output growth. At present, this is hard to judge in Japan but I will use 3 percent since output grew at a rate of 4 percent over 1972-92. Estimating the output gap is even more difficult, but here my procedure is to fit a trend line for $\bar{y}_t$ over 1972:1-1992:4, and then to assume a growth rate of $\bar{y}_t$ equal to 2.5 percent since 1992:2.[16]

The results of this exercise are shown in Figure 2. That policy needed to be much tighter over 1972-78 shows up clearly, and that policy was on track or somewhat too tight over 1982-87 is suggested. But our main interest resides in more recent policy. Figure 2 indicates that it was about right over 1988-93, but, except for 1997, has been too tight since 1994. At the end of 1998, the call rate was slightly over 3 percent too high, the rule-indicated value being –3.0 percent. Of course this latter value is not feasible, but it indicates that the rule calls for much more stimulative policy than what actually prevailed in late 1998.[17]

Now let us see what the McCallum rule has to say. For this exercise I adopt the same value of π^* and use 3 percent as the long-run average growth rate of real output, yielding a nominal GDP growth target of 5 percent per year or $\Delta x^* = 0.0125$ in quarterly log units. The results of this exercise are shown in Figure 3, with the base growth rates expressed in per-annum percentage points. Here, when the solid rule-suggested values are greater than the dotted actual values for base growth, the indication is that policy should have been looser. Thus, we see that this rule agrees with Taylor's regarding 1972-78 and 1994-98. It suggests that policy was too loose on average over 1986-89 (when U.S. policymakers were encouraging a weaker yen).

Figure 2

Taylor Rule and Actual Values for Japan, 1972-98
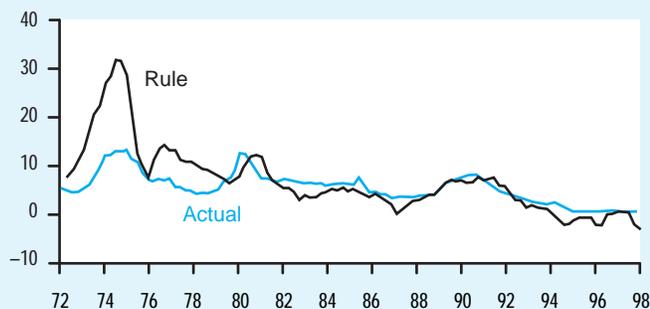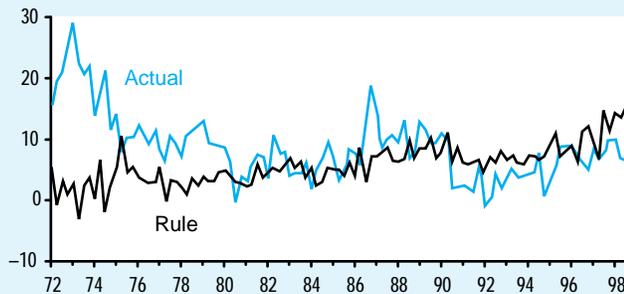
Overnight Call Rate

Figure 3

McCallum Rule and Actual Values for Japan, 1972-98

Growth Rate of Monetary Base

And regarding the more recent period, Figure 3 agrees that policy has been too tight during 1994-98 but suggests that this period of monetary stringency began several years earlier—around the middle of 1990.

I believe that most academic analysts quite recently have come to share the viewpoint indicated in this last picture, i.e., that Japanese monetary policy has been too tight since the early 1990s. It is extremely unfortunate for Japan and perhaps for the world that this view did not prevail sooner. In fact, it did prevail among economists of a monetarist or semi-monetarist persuasion. My own small contributions are mentioned in footnote 19. More prominently, the written contributions of Goodfriend (1997) and

[16] This is in my opinion a weakness of the Taylor rule; knowledge of the level of $\bar{y}_t$ is not needed for mine.

[17] Most commentators simply assert that negative nominal interest rates are impossible. I believe that statement is too strong, partly for reasons indicated by Thornton (1999). But rates well below zero do seem implausible.

Taylor (1997) called for greater monetary stimulus by Japan, including, if necessary, purchases of foreign exchange or non-traditional assets.[18] Milton Friedman's *Wall Street Journal* article of December 1997 put forth a similar position quite strongly, as did Allan Meltzer's piece in the *Financial Times* (1998).

During the years 1995-98, however, it was orthodox opinion in the financial press—including the *Financial Times* and *The Economist*—that monetary policy could provide no more stimulus in Japan "because interest rates were already as low as they could go." This view was not challenged by most academics. Figure 3, however, indicates that a policy rule that uses the monetary base as an essential variable would have been giving signals indicative of overly tight policy for years, if anyone had bothered to look.[19] The Taylor rule concurs, but it did not begin to give these signals until later—and also does not agree regarding the period 1986-89. My conclusion is that one does not have to be an opponent of the Taylor rule or the analytical framework shown in equations 1-3—which I am not—to believe that there remains an extremely important role to be played by measures of the monetary base and other monetary aggregates. I would like to believe that Homer Jones would have approved of this conclusion.

REFERENCES

Andersen, Leonall C., and Jerry L. Jordan. "Monetary and Fiscal Actions: A Test of Their Relative Importance in Economic Stabilization," this *Review* (November 1968), pp. 11-24.

Andersen, Leonall C., and Keith M. Carlson. "A Monetarist Model for Economic Stabilization," this *Review* (April 1970), pp. 7-25.

Brayton, Flint, Andrew Levin, Ralph Tryon, and John C. Williams. "The Evolution of Macro Models at the Federal Reserve Board," *Carnegie-Rochester Conference Series on Public Policy* (December 1997), pp. 43-81.

Brunner, Karl. "The Role of Money and Monetary Policy," this *Review* (July 1968), pp. 9-24.

Calvo, Guillermo A. "Staggered Prices in a Utility Maximizing Framework," *Journal of Monetary Economics* (September 1983), pp. 383-98.

Clarida, Richard, and Mark Gertler, "How the Bundesbank Conducts Monetary Policy," in *Reducing Inflation: Motivation and Strategy,*

Christina D. Romer and David H. Romer, eds. University of Chicago Press, 1996, pp. 363-406.

Friedman, Milton, and David Meiselman. "The Relative Stability of Monetary Velocity and the Investment Multiplier in the United States, 1897-1958," in *The Commission on Money and Credit, Stabilization Policies,* Prentice Hall, 1963, pp. 165-268.

Friedman, Milton. "Rx for Japan: Back to the Future," *Wall Street Journal* (Dec. 17, 1997), p. 22.

Fuhrer, Jeffrey C., and George R. Moore. "Inflation Persistence," *Quarterly Journal of Economics* (February 1995), pp. 127-59.

Goodfriend, Marvin. "Comments," in *Towards More Effective Monetary Policy,* Iwao Kuroda, ed., St. Martin's Press, 1997, pp. 289-95.

Kerr, William, and Robert G. King. "Limits on Interest Rate Rules in the IS Model," Federal Reserve Bank of Richmond *Economic Quarterly* (Spring 1996), pp. 47-75.

Lucas, Robert E., Jr. "Expectations and the Neutrality of Money," *Journal of Economic Theory* (April 1972a), pp. 103-24.

_____. "Econometric Testing of the Natural-Rate Hypothesis," *The Econometrics of Price Determination,* Otto Eckstein, ed., Board of Governors of the Federal Reserve System, 1972b, pp. 50-9.

_____. "Some International Evidence on Output-Inflation Tradeoffs," *American Economic Review* (June 1973), pp. 326-34.

_____. "Econometric Policy Evaluation: A Critique," *Carnegie-Rochester Conference Series on Public Policy* (1976), pp. 19-46.

McCallum, Bennett T. "Monetary versus Fiscal Policy Effects: A Review of the Debate," in *The Monetary Versus Fiscal Policy Debate: Lessons from Two Decades,* Hafer, R.W., ed., Rowman and Allanheld, 1986, pp. 9-29.

_____. "Real Business Cycle Models," in *Modern Business Cycle Theory,* Robert J. Barro, ed., Harvard University Press, 1989, pp. 16-50.

_____. "Specification and Analysis of a Monetary Policy Rule for Japan," Bank of Japan, *Monetary and Economic Studies* (November 1993), pp. 1-45.

_____ and Marvin Goodfriend. "Demand for Money: Theoretical Studies," in *The New Palgrave: A Dictionary of Economics.* John Eatwell, Murray Milgate, and Peter Newman, eds., Stockton Press, 1987.

_____ and Monica Hargraves. "A Monetary Impulse Measure for Medium-Term Policy Analysis," *Staff Studies for the World Economic Outlook,* (September 1995), International Monetary Fund, pp. 52-69.

_____ and Edward Nelson. "An Optimizing IS-LM Specification for Monetary Policy and Business Cycle Analysis," *Journal of Money, Credit, and Banking* (August 1999a), pp. 296-316.

_____ and _____. "Performance of Operational Policy Rules in an Estimated Semi-Classical Structural Model," in *Monetary Policy Rules,* John B. Taylor, ed., University of Chicago Press, 1999b, pp. 15-45.

_____ and _____. "Nominal Income Targeting in an Open-Economy Optimizing Model," *Journal of Monetary Economics* (June 1999c), pp. 553-78.

[18] These contributions were delivered at the Seventh International Conference of the Bank of Japan, held in Tokyo in October 1995.

[19] In fact I did look, although in a less effective way, in McCallum (1993) and McCallum and Hargraves (1995). The story was similar to that as shown in Figure 3.

Meltzer, Allan H. "Time to Print Money," *Financial Times* (July 17, 1998), p. 14.

Rotemberg, Julio J. "Monopolistic Price Adjustment and Aggregate Output," *Review of Economic Studies* (October 1982), pp. 517-31.

Sargent, Thomas J. "A Note on the Accelerationist Controversy," *Journal of Money, Credit, and Banking* (August 1971), pp. 721-25.

_____. "Rational Expectations, the Real Rate of Interest, and the Natural Rate of Unemployment," *Brookings Papers on Economic Activity* (No. 2, 1973), pp. 429-72.

Taylor, John B. "Estimation and Control of a Macroeconomic Model with Rational Expectations," *Econometrica* (September 1979), pp. 1267-86.

_____. "Discretion versus Policy Rules in Practice," *Carnegie-Rochester Conference Series on Public Policy* (December 1993), pp. 195-214.

_____. "Policy Rules as a Means to a More Effective Monetary Policy," in *Towards More Effective Monetary Policy*. Iwao Kuroda, ed., St. Martin's Press,1997, pp. 28-39.

_____. "An Historical Analysis of Monetary Policy Rules," in *Monetary Policy Rules*, John B. Taylor, ed., University of Chicago Press, 1999, pp. 319-41.

Thornton, Daniel L. "Nominal Interest Rates: Less than Zero?" Federal Reserve Bank of St. Louis *Monetary Trends* (January 1999), p. 1.

Walsh, Carl E. *Monetary Theory and Policy*, MIT Press, 1998.

Walters, Alan A. "Consistent Expectations, Distributed Lags, and the Quantity Theory," *Economic Journal* (June 1971), pp. 273-81.

Woodford, Michael. "Price Level Determinacy Without Control of a Monetary Aggregate," *Carnegie-Rochester Conference Series on Public Policy* (December 1995), pp. 1-46.

Christopher J. Neely is a senior economist at the Federal Reserve Bank of St. Louis. Kent Koch provided research assistance.

An Introduction to Capital Controls

Christopher J. Neely

Moreover, it may well be asked whether we can take it for granted that a return to freedom of exchanges is really a question of time. Even if the reply were in the affirmative, it is safe to assume that after a period of freedom the regime of control will be restored as a result of the next economic crisis.
—Paul Einzig, *Exchange Control*, MacMillan and Company, 1934.

Currency controls are a risky, stopgap measure, but some gaps desperately need to be stopped.
—Paul Krugman, "Free Advice: A Letter to Malaysia's Prime Minister," *Fortune*, September 28, 1998.

Unlike many topics in international economics, *capital controls*—taxes or restrictions on international transactions in assets like stocks or bonds—have received cursory treatment in textbooks and scant attention from researchers. The consensus among economists has been that capital controls—like tariffs on goods—are obviously detrimental to economic efficiency because they prevent productive resources from being used where they are most needed. As a result, capital controls gradually had been phased out in developed countries during the 1970s and 1980s, and by the 1990s there was substantial pressure on less-developed countries to remove their restrictions, too (*New York Times*, 1999). The topic almost had been relegated to a curiosity.

Several recent developments, however, have rekindled interest in the use and study of capital controls. First, the resumption of large *capital flows*—trade in assets— to developing countries during the late 1980s and early 1990s created new problems for policymakers. Second, a string of exchange rate/financial crises during the 1990s—the European Monetary System crises of 1992-93, the Mexican crisis of 1994 and the Asian financial crisis of 1997-98—focused attention on the asset transactions that precipitated them. In particular, Malaysia's adoption of capital controls on September 1, 1998, has prompted increased media attention and has renewed debate on the topic.

Modern capital controls were developed by the belligerents in World War I to maintain a tax base to finance wartime expenditures. Controls began to disappear after the war, only to return during the Great Depression of the 1930s. At that time, their purpose was to permit countries greater ability to reflate their economies without the danger of capital flight. In fact, the International Monetary Fund (IMF) Articles of Agreement (Article VI, section 3) signed at the Bretton-Woods conference in 1944 explicitly permitted capital controls.[1] One of the architects of those articles, John Maynard Keynes, was a strong proponent of capital controls and the IMF often was seen as such during its early years. During the Bretton-Woods era of fixed-exchange rates, many countries limited asset transactions to cope with balance-of-payments difficulties. But, recognition of the costs and distortions created by these restrictions led to their gradual removal in developed countries over the last 30 years. The United States, for example, removed its most prominent capital controls in 1974 (Congressional Quarterly Service, 1977). During the last 10 years even less-developed countries began to liberalize trade in assets.

The purpose of this article is to introduce *Review* readers to the debate on capital controls, to explain the purposes

[1] "Article VI. Section 3. Controls of capital transfers: Members may exercise such controls as are necessary to regulate international capital movements, but no member may exercise these controls in a manner which will restrict payments for current transactions or which will unduly delay transfers of funds in settlement of commitments, except as provided in Article VII, Section 3(b) and in Article XIV, Section 2."

and costs of controls and why some advocate their reintroduction. To lay the groundwork for understanding restrictions on capital flows, the next section of the article describes capital flows and their benefits. The third section characterizes the most common objectives of capital controls with an emphasis on the recent debate about using controls to foster macroeconomic stability. Then the many types of capital controls are distinguished from each other and their effectiveness and costs are considered. In addition, accompanying shaded inserts outline specific case studies in capital controls: the U.S. Interest Equalization Tax of 1963, the Chilean *encaje* of the 1990s, and the restrictions imposed by Malaysia in September 1998.

CAPITAL FLOWS

To understand what capital controls do, it is useful to examine capital flows—trade in real and financial assets. International purchases and sales of existing real and financial assets are recorded in the *capital account* of *the balance of payments*.[2] Real assets include production facilities and real estate while financial assets include stocks, bonds, loans, and claims to bank deposits.[3] Capital account transactions often are classified into portfolio investment and direct investment. Portfolio investment encompasses trade in securities like stocks, bonds, bank loans, derivatives, and various forms of credit (commercial, financial, guarantees). Direct investment involves the purchase of real estate, production facilities, or substantial equity investment.[4] When a German corporation, BMW, for example, builds an automobile factory in South Carolina, that is direct investment. On the other hand, when U.S. investors buy Mexican government bonds, that is portfolio investment.

A country is said to have a *deficit* in the capital account if it is accumulating net claims on the rest of the world by purchasing more assets and/or making more loans to the rest of the world than it is receiving. A country, like Japan, with a capital account deficit is also said to experience a *capital outflow*. Accumulating claims on the rest of the world is a form of national saving. Conversely, a country is said to have a *surplus* in the capital account—or a *capital inflow*—if the rest of the world is accumulating net claims on it, as is the case with the United States.[5] Just as individuals must avoid borrowing excessively, policymakers must make sure that the rest of the world does not accumulate too many net claims on their countries—in other words, that their countries do not sell assets/borrow at an unsustainable rate.[6]

Benefits of Capital Flows

Economists have long argued that trade in assets (capital flows) provides substantial economic benefits by enabling residents of different countries to capitalize on their differences. Fundamentally, capital flows permit nations to trade consumption today for consumption in the future—to engage in intertemporal trade (Eichengreen, et al. 1999). Because Japan has a population that is aging more rapidly than that of the United States, it makes sense for Japanese residents to purchase more U.S. assets than they sell to us. This allows the Japanese to save for their retirement by building up claims on future income in the United States while permitting residents of the United States to borrow at lower interest rates than they could otherwise pay.

A closely related concept is that capital flows permit countries to avoid large falls in national consumption from economic downturn or natural disaster by selling assets to and/or borrowing from the rest of the world. For example, after an earthquake devastated southern Italy on November 23, 1980, leaving 4,800 people dead, Italians borrowed from abroad (ran a capital account surplus) to help repair the damage. Figure 1 illustrates the time series of the Italian capital account from 1975 through 1985.

A third benefit is that capital flows permit countries as a whole to borrow in order to improve their ability to produce goods and services in the future—like individuals borrowing to finance an educa-

tion. To cite just one example, between 1960 and 1980 Koreans borrowed funds from the rest of the world equal to about 4.3 percent of gross domestic product (GDP) annually to finance investment during Korea's period of very strong growth (see Figure 2).

These arguments for free capital mobility are similar to those that are used to support free trade. Countries with different age structures, saving rates, opportunities for investment, or risk profiles can benefit from trade in assets. More recently, economists have emphasized other benefits of capital flows such as the technology transfer that often accompanies foreign investment, or the greater competition in domestic markets that results from permitting foreign firms to invest locally (Eichengreen, et al. 1999). The benefits of capital flows do not come without a price, however. Because capital flows can complicate economic policy or even be a source of instability themselves, governments have used capital controls to limit their effects (Johnston and Tamirisa, 1998).

PURPOSES OF CAPITAL CONTROLS

A capital control is any policy designed to limit or redirect capital account transactions. This broad definition suggests that it will be difficult to generalize about capital controls because they can take many forms and may be applied for various purposes (Bakker, 1996). Controls may take the form of taxes, price or quantity controls, or outright prohibitions on international trade in assets.[7]

Revenue Generation and Credit Allocation

The first widespread capital controls were adopted in WWI as a method to finance the war effort. At the start of the war, all the major powers suspended their participation in the gold standard for the duration of the conflict but maintained fixed-exchange rates. All the belligerents restricted *capital outflows*, the purchase of foreign assets or loans abroad. These

Figure 1

Italian Capital Account Surplus as a Percentage of GDP
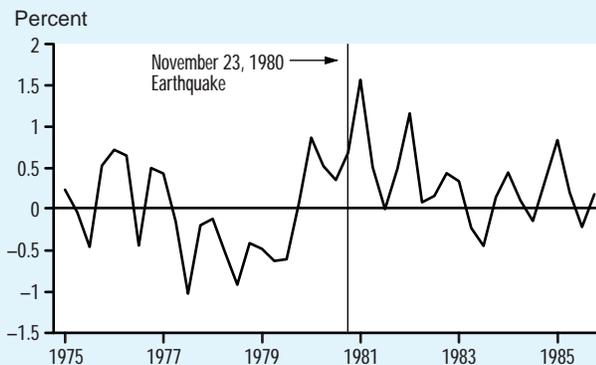
SOURCE: International Financial Statistics

Figure 2

South Korean Growth and Capital Surplus
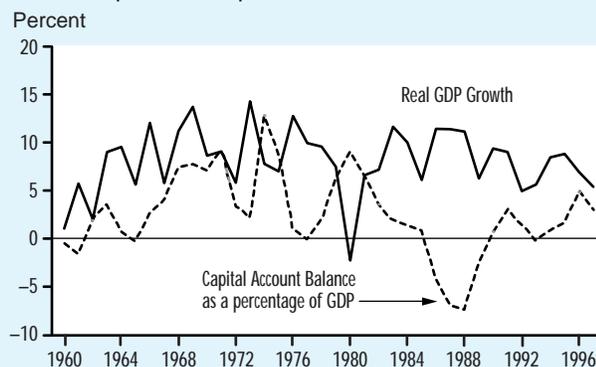
SOURCE: International Financial Statistics and Mitchell (1998)

restrictions raised revenues in two ways. First, by keeping capital in the domestic economy, it facilitated the taxation of wealth and interest income (Bakker, 1996). Second, it permitted a higher inflation rate, which generated more revenue. Capital controls also reduced interest rates and therefore the government's borrowing costs on its own debt (Johnston and Tamirisa, 1998). Since WWI, controls on capital outflows have been used similarly in other —mostly developing—economies to generate revenue for governments or to permit them to allocate credit domestically without risking capital flight (Johnston and

[7] Alesina, Grilli, and Milesi-Ferretti (1994) and Grilli and Milesi-Ferretti (1995) empirically examine factors associated with capital controls.

Table 1

Purposes of Capital Controls

Purpose of Control	Method	Direction of Control	Example
Generate Revenue/ Finance War Effort	Controls on capital outflows permit a country to run higher inflation with a given fixed-exchange rate and also hold down domestic interest rates.	Outflows	Most belligerents during WWI and WWII
Financial Repression/ Credit Allocation	Governments that use the financial system to reward favored industries or to raise revenue, may use capital controls to prevent capital from going abroad to seek higher returns.	Outflows	Common in developing countries
Correct a Balance of Payments Deficit	Controls on outflows reduce demand for foreign assets without contractionary monetary policy or devaluation. This allows a higher rate of inflation than otherwise would be possible.	Outflows	U.S. interest equalization tax, 1963-74
Correct a Balance of Payments Surplus	Controls on inflows reduce foreign demand for domestic assets without expansionary monetary policy or revaluation. This allows a lower rate of inflation than would otherwise be possible.	Inflows	German Bardepot scheme, 1972-74
Prevent Potentially Volatile Inflows	Restricting inflows enhances macroeconomic stability by reducing the pool of capital that can leave a country during a crisis.	Inflows	Chilean *encaje*, 1991-98
Prevent Financial Destabilization	Capital controls can restrict or change the composition of international capital flows that can exacerbate distorted incentives in the domestic financial system.	Inflows	Chilean *encaje*, 1991-98
Prevent Real Appreciation	Restricting inflows prevents the necessity of monetary expansion and greater domestic inflation that would cause a real appreciation of the currency.	Inflows	Chilean *encaje*, 1991-98
Restrict Foreign Ownership of Domestic Assets	Foreign ownership of certain domestic assets—especially natural resources—can generate resentment.	Inflows	Article 27 of the Mexican constitution
Preserve Savings for Domestic Use	The benefits of investing in the domestic economy may not fully accrue to savers so the economy, as a whole, can be made better off by restricting the outflow of capital.	Outflows	
Protect Domestic Financial Firms	Controls that temporarily segregate domestic financial sectors from the rest of the world may permit domestic firms to attain economies of scale to compete in world markets.	Inflows and Outflows	

Tamirisa). Table 1 summarizes the purposes of capital controls.

Balance of Payments Crises

During the Great Depression, controls simultaneously were used to achieve greater freedom for monetary policy and exchange rate stability—goals that have remained popular. To understand why controls have been used in this way, it is necessary to understand balance of payments problems and their solutions (Johnston and Tamirisa, 1998). At a given exchange rate, a country often will want to collectively purchase more goods, services and assets than the rest of the world will buy from it. Such an imbalance is called a *balance of payments deficit* and may come about for any one of a number of reasons: 1) The domestic business cycle may be out of sync with that of the rest of the world; 2) There may have been a rapid change in the world price of key commodities like oil; 3) Expansionary domestic policy may have increased demand for the rest of the world's goods; 4) Large foreign debt interest obligations may surpass the value of the domestic economy's exports; 5) Or, a perception of deteriorating economic policy may have reduced international demand for domestic assets.[8] In the absence of some combination of exchange rate and monetary policy by the deficit country, excess demand for foreign goods and assets would bid up their prices—typically through a fall in the foreign exchange value (a *devaluation* or *depreciation*) of the domestic currency—until the deficit was eliminated.[9]

There are four policy alternatives to correct an imbalance in international payments: 1) Permit the exchange rate to change, as described above; 2) Use monetary policy—unsterilized foreign exchange intervention—to correct the imbalance through domestic demand; 3) Attempt to *sterilize* the monetary changes to isolate the domestic economy from the capital flows; and 4) Restrict capital flows.[10] Each alternative has disadvantages.

When domestic residents purchase more goods and assets from foreigners

than foreigners purchase from domestic residents, the exchange rate (the price of foreign currency) tends to rise. If the exchange rate is flexible, the foreign currency tends to *appreciate* and the domestic currency tends to *depreciate*. The depreciation of the domestic currency raises prices of imported goods and assets to domestic residents and lowers the prices of domestic goods and assets on world markets, reducing the relative demand for foreign goods and assets until the imbalance in the balance of payments is eliminated. A country with a fixed exchange rate similarly may correct a balance of payments deficit by changing the exchange rate peg—devaluing the currency—but this option foregoes the benefits of exchange rate stability for international trade and policy discipline. In addition, it may reduce the public's confidence in the monetary authorities' anti-inflation program.

If a government is committed to maintaining a particular fixed exchange rate, on the other hand, its central bank can prevent the depreciation of its currency with contractionary monetary policy—by selling domestic bonds.[11] Alternatively, the central bank might sell foreign exchange to affect the monetary base, in which case the action is known as *unsterilized foreign exchange intervention.* In either case, such a sale lowers the domestic money supply and raises domestic interest rates—lowering domestic demand for imports—while reducing the prices of domestic goods, services, and assets relative to their foreign counterparts. The reduced demand and higher prices for foreign goods, services, and assets would eliminate a balance of payments deficit. However, this defense of the exchange rate requires that monetary policy be devoted solely to maintaining the exchange rate; it cannot be used to achieve independent domestic inflation or employment goals. In this case, for example, the contraction temporarily will reduce domestic demand and employment, which may be undesirable. A country that uses monetary policy to defend the exchange rate in the face of imbalances in international payments is said to *subordinate*

[8] Often, the term "balance of payments deficit" describes an imbalance in the current account (goods, services, factor payments, and unilateral transfers). Here, it describes the sum of the current account and the capital account. Countries also may demand fewer goods and assets from the rest of the world—balance of payments surpluses—but this article concentrates on balance of payments deficits because most countries find balance of payments surpluses easier to manage.

[9] When a flexible exchange rate currency gains or loses value, it is said to *appreciate* or *depreciate*, respectively. Fixed-rate currencies are said to be *revalued* or *devalued* when their price rises or falls.

[10] There are other policies—fiscal and regulatory—that may be used to manage the effects of capital flows but they will be ignored to simplify the discussion.

[11] Foreign exchange operations that do not affect the domestic monetary base are called "sterilized," while those that do affect the monetary base are called "unsterilized." Sales of any asset would tend to lower the domestic money supply and raise interest rates because when the monetary authority receives payment for the asset, the payment ceases to be part of the money supply. Fiscal policy also may have an effect on exchange rates but taxing and spending decisions usually are more constrained than monetary decisions.

domestic monetary policy to exchange rate concerns.

Rather than subordinate monetary policy to maintaining the exchange rate, some central banks have attempted to recapture some monetary independence by *sterilizing*—or reversing—the effect of foreign exchange operations on the domestic money supply. Sterilization of sales of foreign exchange (foreign bonds), for example, would require the central bank to buy an equal amount of domestic bonds, leaving domestic interest rates unchanged after the inflow. It generally is believed that sterilized intervention does not affect the exchange rate and so it is not very effective in recapturing monetary independence (Edwards, 1998b).

If international investors don't believe that the monetary authorities will defend the exchange rate with tighter monetary policy, they will expect devaluation—a fall in the relative price of domestic goods and assets—and will sell domestic assets to avoid a loss. Such a sale increases relative demand for foreign assets, exacerbating the balance of payments deficit, and speeds the devaluation.[12]

Capital flows play a crucial role in balance of payments crises in two ways. Swings in international capital flows can create both a balance-of-payments problem and—if the exchange rate is not defended—expedite devaluation under fixed-exchange rates. Thus, in the presence of free capital flows, a country wishing to maintain a fixed-exchange rate must use monetary policy solely for that purpose. As McKinnon and Oates (1966) argued, no government can maintain fixed-exchange rates, free capital mobility, and have an independent monetary policy; one of the three options must give. This is known as the "incompatible trinity" or the trilemma (Obstfeld and Taylor, 1998). Policymakers wishing to avoid exchange-rate fluctuation and retain scope for independent monetary policy must choose the fourth option, restrict capital flows.

By directly reducing demand for foreign assets and the potential for speculation against the fixed-exchange rate,

controls on capital outflows allow a country to maintain fixed-exchange rates and an independent domestic monetary policy while alleviating a balance-of-payments deficit.[13] The monetary authorities can meet both their internal goals (employment and inflation) and their external goals (balance of payments).[14] Thus, capital controls are sometimes described in terms of the choices they avoid: to *prevent capital outflows* that, through their effect on the *balance of payments*, might either endanger *fixed-exchange rates* or *independence of monetary policy*.

Real Appreciation of the Exchange Rate

While capital outflows can create balance-of-payments deficits, capital inflows can cause *real appreciation* of the exchange rate.[15] During the 1980s and 1990s a number of developing countries completed important policy reforms that made them much more attractive investment environments. Eichengreen, et al. (1999) report that net capital flows to developing countries tripled from $50 billion in 1987-89 to more than $150 billion in 1995-97. These large capital inflows to the reforming countries tended to drive up the prices of domestic assets. For countries with flexible exchange rates, the exchange rates appreciated, raising the relative prices of the domestic countries' goods. For countries with fixed-exchange rates, the increased demand for domestic assets led the monetary authorities to buy foreign exchange (sell domestic currency), increasing the domestic money supply and ultimately the prices of domestic goods and assets. In either case, the prices of domestic goods and assets rose relative to those in the rest of the world—a real appreciation—making domestic exported goods less competitive on world markets and hurting exporting and import-competing industries.[16] Because of these effects, the problem of real exchange-rate appreciation from capital inflows is described variously as real exchange-rate instability, real appreciation, or loss of competitiveness.

[12] Of course, capital outflows are not necessary to force devaluation. If the domestic economy still demands more goods and services than it supplies to the rest of the world, the exchange rate cannot be maintained without a monetary contraction.

[13] Countries that face *balance of payments surpluses*—the desire to purchase fewer goods and services from the rest of the world at the fixed-exchange rate—would restrict capital inflows, rather than outflows, to reduce demand for their own assets.

[14] If there is still excess demand for foreign goods, the fixed-exchange rate will still be only temporarily sustainable.

[15] A *nominal appreciation* is a rise in the foreign exchange value of a country's currency. A *real appreciation* is a rise in the relative price of domestic goods and services compared to foreign goods and services. This may result from a nominal appreciation, domestic inflation that is higher than foreign inflation, or some combination of the two.

[16] Empirically, Edwards (1998b) finds a consistent, but limited, tendency toward real appreciation from capital inflows.

Countries have a number of policy options to prevent real appreciation in the face of capital inflows (Goldstein, 1995; Corbo and Hernandez, 1996). Permitting the exchange rate to change still results in nominal and real appreciation but avoids domestic inflation. A very common tactic for fixed exchange-rate regimes is to sterilize the monetary effects of the inflows, preventing an expansion of the money supply by reversing the effect on the domestic money market (Edwards, 1998b). It generally is believed that sterilization is not very effective in recapturing monetary independence as it keeps domestic real interest rates high and leads to continued inflows. Sterilization of inflows also is a potentially expensive strategy for the government as the domestic bonds that the central bank sells may pay higher interest than the foreign bonds the central bank buys. Fiscal contraction is an effective way to prevent real appreciation because it lowers domestic interest rates, and likewise, the demand for domestic assets; but raising taxes and/or reducing government spending may be politically unpalatable. Because of the problems associated with these first three policies, countries like Brazil, Chile, and Columbia chose to use capital controls—restricting purchase of domestic assets (inflows)—to try to prevent real appreciation and substitute for fiscal policy flexibility in the face of heavy inflows.

Theories of the Second Best

More recently, economists have considered other circumstances—other than balance-of-payments needs or real appreciation—under which capital controls might be a useful policy. As a rule, economists emphasize that restrictions on trade and investment impose costs on the economy. There are exceptions to that rule, however. Taxes and quantitative restrictions may be good for the economy—welfare improving, in technical jargon—if they are used to correct some other, pre-existing distortion to free markets that cannot be corrected otherwise. The idea that a tax or quantitative restriction can improve economic wel-

fare in this way is called a "theory of the second best."[17]

Capital controls preserve domestic savings for domestic use. From a national point of view, there might be benefits from a greater rate of domestic investment that do not fully accrue to the investors. For example, domestic savers might invest disproportionately overseas because of political risk of expropriation or a desire to escape taxation. In either case, the nation as a whole could be made better off by limiting or taxing domestic investment abroad (Harberger, 1986).

The infant industry argument—an old idea often used to justify tariffs in goods markets —has been resurrected to rationalize the use of capital controls on both inflows and outflows. This idea starts with the premise that small, domestic firms are less efficient than larger, foreign firms and so will be unable to compete on an equal basis. To permit small domestic firms to grow to the efficient scale that they need to enjoy to compete in world markets, they must be protected temporarily from international competition by trade barriers. As applied to capital markets, the argument urges that capital controls be used to protect underdeveloped financial markets from foreign competition. The problem with this argument, as in goods markets, is that protected industries often never grow up and end up seeking perpetual protection.

Financial Sector Distortions

In reality, capital controls rarely have been imposed in a well-thought-out way to correct clearly defined pre-existing distortions. Instead, capital controls most often have been used as a tool to postpone difficult decisions on monetary and fiscal policies. Recently, however, the case has been made that capital controls may be the least disadvantageous solution to the destabilizing effects of capital flows on inadequately regulated financial systems.

Recall that when a country with a fixed-exchange rate has a net capital outflow, the increase in relative demand for foreign assets means that there is insuffi-

[17] The classic example of a tax that improves welfare is the one imposed on a polluting industry. Because a polluting industry imposes non-market costs on others, for which it does not compensate them, a government could improve everyone's well being if it were to tax pollution. The factory would produce less pollution and people would be happier. The desire to meet employment goals with an independent monetary policy is really a version of a "second best" story in which the pre-existing distortion is price inertia (or a similar friction) in the real economy that causes monetary policy to have real effects (Dooley, 1996).

cient demand for domestic goods and assets at the fixed-exchange rate. The domestic monetary authorities may conduct contractionary monetary policy—raise domestic interest rates to make their assets more attractive—or lower the prices of their goods and assets (devalue the currency). This is a special case of a balance-of-payments deficit and presents the same choice—to raise interest rates or devalue—but in the case of a sudden capital outflow, the crisis manifests itself in large sudden capital outflows rather than more gradual balance-of-payments pressures from other causes. Governments must choose between high interest rates coupled with some capital outflows or an exchange rate devaluation that provokes fear of inflation and policy instability leading to greater capital outflows. In either case, a serious recession seems unavoidable.

The recent case for capital controls recognizes that a monetary contraction not only slows economic activity through the normal interest-rate channels, but also can threaten the health of the economy through the banking system (Kaminsky and Reinhart, 1999). If the monetary authorities raise interest rates, they increase the costs of funds for banks and—by slowing economic growth—reduce the demand for loans and increase the number of nonperforming loans. Choosing to devalue the currency rather than raising interest rates does not necessarily help banks either, as they may have borrowed in foreign currency. A devaluation would increase the banks' obligations to their foreign creditors. Thus, capital outflows from the banking system pose special problems for the monetary authorities, as banks' liabilities are usually implicitly or explicitly guaranteed by the government.

Indeed, the very nature of the financial system creates perverse incentives (distortions) that international capital flows often exacerbate (Mishkin, 1998). For example, in a purely domestic context, banks have incentives to make risky loans, as their losses are limited to the owners' equity capital, but their potential profits are unlimited. The existence of deposit insurance worsens this problem by reducing

depositors' incentive to monitor their banks' loan portfolio for excessive risk. Deposit insurance, in turn, exists precisely because depositors can not easily monitor the riskiness of their banks. In the absence of deposit insurance, depositors would find it difficult to tell good banks from bad banks and would withdraw their money at any sign of danger to the bank. Once some depositors began to withdraw their money from the bank, all depositors would try to do so, forcing the bank to close, even if its underlying assets were productive (Diamond and Dybvig, 1983). This puts the whole banking system at risk.

To avoid this problem, most developed countries combine implicit or explicit insurance of bank deposits with government regulation of depository institutions, especially their asset portfolios (loans). In emerging markets, however, banking regulation is much more difficult as the examiners are less experienced, have fewer resources and less strict accounting standards by which to operate. Thus, banking problems are more serious in emerging markets.

Large international capital inflows, especially short-term foreign borrowing, can exacerbate these perverse incentives and pose a real danger to banking systems. Domestic banks often view borrowing from abroad in foreign currency sources as a low-cost source of funds—as long as the domestic currency is not devalued. With this additional funding, banks expand into unfamiliar areas, generating risky loans that potentially create systemic risk to the banking system (Eichengreen, 1999; Garber, 1998; Goldstein, 1995; Dornbusch, 1998). If capital outflows force a devaluation, the foreign-currency denominated debts of the banking system increase when measured in the domestic currency, possibly leading to bank failures. The banking system is a particularly vulnerable conduit by which capital flows can destabilize an economy because widespread bank failures impose large costs on taxpayers and can disrupt the payments system and the relationships between banks and firms who borrow from them (Friedman and Schwartz, 1963; Bernanke, 1983). The

difficulty of effective banking regulation creates an argument for capital controls as a second-best solution to the existence of the distorted incentives in the banking system.[18]

There are two ways in which capital controls might be imposed to limit capital flow fluctuations and achieve economic stability. First, capital controls may be used to discourage capital *outflows* in the event of a crisis—as Malaysia did in September 1998—permitting looser domestic monetary policy. Controls on outflows are ideally taken as a transitional measure to buy time to achieve goals, as an aid to reform rather than as a substitute (Krugman, 1998). Second, controls can prevent destabilizing outflows by discouraging or changing the composition of capital *inflows*, as Chile did for most of the 1990s.

The second method—to discourage or change the composition of capital inflows with controls—requires some explanation. A prime fear of those who seek to limit capital flows is that sudden outflows may endanger economic stability because investors are subject to panics, fads, and bubbles (Kindleberger, 1978; Krugman, 1998; Wade and Veneroso, 1998). Investors may panic because they, as individuals, have limited information about the true value of the assets they are buying or selling. They can, however, infer information from the actions of others. For example, one might assume that a crowded restaurant serves good food, even if one has never eaten there. In financial markets, participants learn about other participants' information by watching price movements. An increase in the price of an asset might be interpreted as new information that the asset had been underpriced, for example. Such a process might lead to "herding" behavior, in which asset price changes tend to cause further changes in the same direction, creating a boom-bust cycle and instability in financial markets, potentially justifying capital controls. By discouraging inflows of foreign capital, governments can limit the pool of volatile capital that may leave on short notice.

Instead of limiting the total quantity of capital inflows, some would argue that changing the composition of that inflow is just as important. For example, it often is claimed that direct investment is likely to be more stable than portfolio investment because stocks or bonds can be sold more easily than real assets (like production facilities) can be liquidated (Dixit and Pindyck, 1994; Frankel and Rose, 1996; Dornbusch, 1998). In contrast, Garber (1998) argues that tracking portfolio and direct investment data may be misleading; derivatives (options, futures, swaps, etc.) can disguise the source of a crisis, making it look like the source is excessive short-term debt. Goldstein (1995) says there is little evidence that direct investment is less "reversible" than portfolio investment. For example, foreign firms with domestic production facilities abroad can use those facilities as collateral for bank loans that then can be converted to assets in another currency, effectively moving the capital back out of the country.

TYPES OF CAPITAL CONTROLS

To meet the many possible objectives described for them, there are many types of capital controls, distinguished by the type of asset transaction they affect and whether they tax the transaction, limit it, or prohibit it outright. This section distinguishes the many types of capital controls by this taxonomy.

Capital controls are not, strictly speaking, the same as exchange controls, the restriction of trade in currencies, although the two are closely related (Bakker, 1996). Although currency and bank deposits are one type of asset—money—exchange controls may be used to control the current account rather than the capital account. For example, by requiring importers to buy foreign exchange from the government for a stated purpose, exchange controls may be used to prohibit the legal importation of "luxury" goods, thereby rationing "scarce" foreign exchange for more politically desirable purposes. So, while exchange controls are inherently a type of limited capital control, they are neither necessary to restrict capital

[18] The capital adequacy standards of the Basle accords penalize long-term international interbank lending relative to short-term lending, exacerbating the problem (Corsetti, Pesenti, and Roubini, 1998b). Although banks are important and heavily regulated almost everywhere, banks play an especially important role in developing countries because information problems tend to be more important in the developing world than they are in the developed world.

MALAYSIA'S CAPITAL CONTROLS: 1998-99

The devaluation of the Thai baht in July 1997 sparked significant capital outflows from Southeast Asia, leading to a fall in local equity prices and plunging exchange rates. To counter these outflows of capital, the IMF urged many of the nations of the region to raise interest rates, making their securities more attractive to international investors. Unfortunately, the higher interest rates also slowed the domestic economies.[1]

In response to this dilemma, Malaysia imposed capital controls on September 1, 1998. The controls banned transfers between domestic and foreign accounts and between foreign accounts, eliminated credit facilities to offshore parties, prevented repatriation of investment until September 1, 1999, and fixed the exchange rate at M3.8 per dollar. Foreign exchange transactions were permitted only at authorized institutions and required documentation to show they were for current account purposes. The government enacted a fairly intrusive set of financial regulations designed to prevent evasion. In February 1999, a system of taxes on outflows replaced the prohibition on repatriation of capital. While the details are complex, the net effect was to discourage short-term capital flows but to freely permit longer-term transactions (Blustein, 1998). By imposing the capital controls, Malaysia hoped to gain some monetary independence, to be able to lower interest rates without provoking a plunge in the value of the currency as investors fled Malaysian assets.

The Malaysian government and business community claimed to be pleased with the effect of the controls in increasing demand and returning stability to the economy. Even economists who oppose capital controls believe that they may have been of some use in buying time to implement fundamental reforms (Barro, 1998).

Others fear, however, that the capital controls have replaced reform, rather than buying time for reform. As of May 1999, the Malaysian government does not appear to be using the breathing space purchased by the capital controls to make fundamental adjustments to its fragile and highly leveraged financial sector. Rather, Prime Minister Mahathir has sacked policymakers who advocate reform while aggressively lowering interest rates, loosening nonperforming loan classification regulation and setting minimum lending targets for banks. This strategy may prove short-sighted, as much of the capital outflow was caused by the recognition that asset prices were overvalued and the banking sector was weak (*Global Investor*, 1998b). Although monetary stimulus may be helpful in the short run, it may exacerbate the underlying problems. In addition, the government must be concerned about the long-term impact that the controls will have on investors' willingness to invest in the country.

[1] For an overview of the causes and policy options in the Asian financial crisis, see Corsetti, Pesenti, and Roubini (1998a and 1998b).

movement nor are they necessarily intended to control capital account transactions.

Controls on Inflows vs. Outflows

Capital controls on some long-term (more than a year) inflows—direct invest- ment and equity—often are imposed for different reasons than those on short-term inflows—bank deposits and money market instruments. While the recent trend has been to limit short-term capital flows because of their allegedly greater volatility and potential to destabilize the economy,

bans on long-term capital flows often reflect political sensitivity to foreign ownership of domestic assets. For example, Article 27 of the Mexican constitution limits foreign investment in Mexican real estate and natural resources.

Controls on capital inflows and outflows provide some slack for monetary policy discretion under fixed exchange rates, but in opposite directions. Controls on capital inflows, which allow for higher interest rates, have been used to try to prevent an expansion of the money supply and the accompanying inflation, as were those of Germany in 1972-74 (Marston, 1995) or Chile during the 1990s.[19] In contrast, controls on capital outflows permit lower interest rates and higher money growth than otherwise would be possible (Marston, 1995). They most often have been used to postpone a choice between devaluation or tighter monetary policy, as they have been in Malaysia, for example (see the shaded insert).

Price vs. Quantity Controls

Capital controls also may be distinguished by whether they limit asset transactions through price mechanisms (taxes) or by quantity controls (quotas or outright prohibitions). Price controls may take the form of special taxes on returns to international investment (like the U.S. interest equalization tax of the 1960s—see the shaded insert), taxes on certain types of transactions, or a mandatory reserve requirement that functions as a tax.

One type of price mechanism to discourage short-term capital flows is the "Tobin" tax. Proposed by Nobel laureate James Tobin in 1972, the Tobin tax would charge participants a small percentage of all foreign exchange transactions (ul Haq, Kaul and Grunberg, 1996; Kasa, 1999). Advocates of such a tax hope that it would diminish foreign exchange market volatility by curtailing the incentive to switch positions over short horizons in the foreign exchange market. There are many problems with a Tobin tax, however. The tax might reduce liquidity in foreign exchange markets or be evaded easily through derivative instruments. It is uncertain who

would collect the tax or for what purposes the revenue would be used. And, most dauntingly, a Tobin tax would have to be enacted by widespread international agreement to be successful.

A *mandatory reserve requirement* is a price-based capital control that commonly has been implemented to reduce capital inflows. Such a requirement typically obligates foreign parties who wish to deposit money in a domestic bank account—or use another form of inflow—to deposit some percentage of the inflow with the central bank for a minimum period. For example, from 1991 to 1998, Chile required foreign investors to leave a fraction of short-term bank deposits with the central bank, earning no interest.[20] As the deposits earn no interest and allow the central bank to buy foreign money market instruments, the reserve requirement effectively functions as a tax on short-term capital inflows (Edwards, 1998b). See the shaded insert on the Chilean *encaje* of the 1990s.

Quantity restrictions on capital flows may include rules mandating ceilings or requiring special authorization for new or existing borrowing from foreign residents. There may be administrative controls on cross-border capital movements in which a government agency must approve transactions for certain types of assets. Certain types of investment might be restricted altogether as in Korea, where the government has, until recently, restricted long-term foreign investment (Eichengreen, et al. 1999). Forbidding or requiring special permission for repatriation of profits by foreign enterprises operating domestically may restrict capital outflows. Capital controls may be more subtle: Domestic regulations on the portfolio choice of institutional investors also may be used as a type of capital control, as they have been in Italy and in South Korea in the past (Bakker, 1996; Park and Song, 1996).

EVALUATING CAPITAL CONTROLS

The conventional wisdom of the economics profession has been—whatever the problems with destabilizing capital flows or

[19] Recall that capital inflows entail foreign purchases of domestic assets or foreign loans to domestic residents while outflows entail domestic purchases of foreign assets or loans to foreign residents by domestic residents. Under a fixed exchange rate, persistent capital inflows will require an expansion of the money supply or a revaluation while substantial capital outflows will require a contraction.

[20] The Chilean reserve requirement applied not only to bank deposits but to many types of capital inflows.

THE U.S. INTEREST EQUALIZATION TAX: 1963-74

During the late 1950s and early 1960s, the United States had both a fixed-exchange rate regime for the dollar (the Bretton-Woods system) and chronic pressures toward balance-of-payments deficits. These strains resulted partly from the fact that interest rates in the rest of the world—especially those in Europe—tended to be higher than those in the United States, making foreign assets look attractive to U.S. residents.[1] Faced with the unpalatable alternatives of devaluing the dollar or conducting contractionary policies, on July 19, 1963, President Kennedy proposed the Interest Equalization Tax (IET) to raise the prices that Americans would have to pay for foreign assets (*Economist*, 1964a).[2]

The IET imposed a variable surcharge, ranging from 1.05 percent on one-year bills to 15 percent on equity and bonds of greater than 28.5 years maturity, on U.S. purchases of stocks and bonds from Western Europe, Japan, Australia, South Africa, and New Zealand (Congressional Quarterly Service, 1969). Canada and the developing world were exempted from the tax out of consideration for their special dependence on U.S. capital markets. By raising the prices of foreign assets, it was hoped that demand for those assets—and the consequent balance-of-payments deficit—would be reduced or eliminated.

The IET reduced direct outflows to the targeted countries but didn't change total outflows much because investors were able to evade the tax through third countries, like Canada (Pearce, 1995; Kreinen, 1971; Stern, 1973). In addition, because the tax did not cover loans, investment was diverted initially from bond and stock purchases to bank loans. Loans from American banks to firms in Europe and Japan jumped from $150 million during the first half of 1963 to $400 million during the second half (*Economist*, 1964b).

To check bank loans to foreign countries, the U.S. Congress enacted the Voluntary Foreign Credit Restraint Program (VFCRP) in February 1965, broadening it in 1966 to limit U.S. short-term capital outflows to other developed countries. In addition, U.S. corporations were asked to voluntarily limit their direct foreign investment. The program was made mandatory in 1968 (Laffer and Miles, 1982; Kreinin, 1971). U.S. capital controls were relaxed in 1969 and phased out in 1974, after the United States left the Bretton-Woods system of fixed-exchange rates (Congressional Quarterly Service, 1973a, 1973b, 1977).

One unintended consequence of the IET was the growth of foreign financial markets—at the expense of U.S. markets—as they inherited the job of intermediating international capital flows. For example, the volume of international borrowing in London rose from $350 million in 1962 to more than $1 billion in 1963 while the volume of foreign flotations in New York fell from $2 billion in the first half of 1963 to just over $600 million in the next nine months (*Economist*, 1964b).

---

[1] At this time, the United States had a current account surplus that failed to fully offset private demand for foreign assets—the capital account deficit—resulting in the need for temporary measures to close the gap.

[2] In August 1964, the U.S. Congress enacted this tax, making it retroactive to the date it was proposed.

CHILE'S *ENCAJE*: 1991-98

During the late 1980s and early 1990s, international capital began to return to Chile as a result of slow growth and low interest rates in the developed world and sound macroeconomic policies, including reduced debt, in Chile (Edwards, 1998a). The Chilean authorities feared that these capital inflows would complicate monetary policy decisions—perhaps causing real appreciation of the exchange rate—and they also were wary of the danger of building up short-term debt.

Chile had long restricted capital flows and these limits were updated in the early 1990s to deal with the surge in capital inflows. Direct investment was made subject to a 10-year stay requirement in 1982; this period was reduced to three years in 1991 and to one-year in 1993. Portfolio flows were made subject to the *encaje*—a one-year, mandatory, non-interest paying deposit with the central bank—created in 1991 to regulate capital inflows.[1] The *encaje* was initially 20 percent but was increased to 30 percent in 1992. The penalty for early withdrawal was 3 percent.[2]

The effect of the *encaje* was to tax foreign capital inflows, with short-term flows being taxed much more heavily than long-term flows. For example, consider the choice of an American buying a one-year discount bond with a face value of 10,000 pesos for a price of 9,091 pesos, or a 10-year discount bond with the same face value and a price of 3,855 pesos. Either bond, if held to maturity, would yield a 10 percent per annum return.[3] In the presence of a 30 percent one-year reserve requirement, however, the one-year bond's annual yield would be 7.7 percent and the 10-year bond's annual yield would be 9.7 percent. Hence, the *encaje* acted as a graduated tax on capital inflows.

Researchers disagree about the effectiveness of Chile's capital controls. Valdes and Soto (1996) concluded that they changed the composition but not the magnitude of the inflows. In other words, investors substituted from heavily taxed short-term flows to more lightly taxed long-term inflows. They also found that the controls were ineffective in preventing a real

appreciation of the exchange rate. Larraín B., Labán M., and Chumacero (1997) studied the same issue with different methods and found that, although there was considerable substitution in the short run, the controls did change the magnitude of the inflows in the long run.

There is even more disagreement about whether the capital controls were important in keeping Chile insulated from the Asian crisis. Many observers have cited Chile's capital controls in advocating more widespread restrictions on capital controls for other developing countries (Bhagwati, 1998). Edwards (1998a), on the other hand, points out that Chile also had substantial capital controls during the late 1970s and early 1980s, before its major banking crisis that cost Chileans more than 20 percent of GDP during 1982-83. The major difference between then and now is that Chile now has a modern and efficient system of banking regulation. Others credit the participation of foreign banks in strengthening the Chilean banking system by providing experience and sophistication in assessing risks and making loans. At the time of the crisis, Chile had a high percentage of domestic loans from foreign-owned banks—20 percent, about the same as the United States and far higher than South Korea, Thailand, and Indonesia (5 percent) (*Economist*, 1997). In addition, Edwards (1998a) claims that the *encaje* harmed the domestic financial services industry and the small firms that could not borrow long term on international markets to avoid the tax. If Chile's capital controls helped, it was to buy time for structural reforms and effective financial regulation.

---

[1] The word *encaje* means "strongbox" in Spanish.

[2] The *encaje* was reduced to 10 percent and the early withdrawal penalty from 3 percent to 1 percent in June 1998. The *encaje* was eliminated entirely on September 16, 1998 (Torres, 1998).

[3] The yield to maturity on a bond equates the initial outlay with the present discounted value of its payoffs. The yields on the bonds are determined by solving the following equations for i:
$9091 = 10,000/(1+i)$, $3855 = 10,000/(1+i)^{10}$,
$1.3*9091 = .3*9091/(1+i)+10,000/(1+i)$ and
$1.3*3855 = .3*3855/(1+i) + 10,000/(1+i)^{10}$.

fixed exchange rates—that capital controls are ineffective and impose substantial costs on economies that outweigh any benefits. That generalization ignores distinctions among types of capital controls and varied criteria for success, however. Capital controls have many potential purposes and thus many potential standards by which to judge their efficacy. Difficulties in separating the effects of capital controls from the balance of payments or capital flow problems they were intended to alleviate complicates the empirical study of the effects of capital controls (Johnston and Tamirisa, 1998). Also, generalizing about the effectiveness of capital controls from one country—or even one period—to another is risky because the effectiveness of capital controls depends on the rigor with which they are enforced (Obstfeld and Taylor, 1998). Governments that control substantial aspects of their citizens' lives (e.g., Cuba) find it easier to enforce controls on trade in assets (Minton, 1999).

Are Capital Controls Effective?

Keeping in mind these difficulties, there are several possible ways to gauge the effectiveness of capital controls. Perhaps the most direct way is to measure whether the imposition of capital controls changes the magnitude or composition of capital flows, using some assumption about what flows would have been without the capital controls. Measuring the composition of capital flows always has been difficult, however, and it has become more so since the advent of derivatives that can be used to disguise capital flows. For example, a U.S. firm may build a production facility in Mexico but hedge the risk that the peso will decline—reducing the dollar value of the investment—by buying put options on the peso, which will increase in value if the peso falls.[21] The direct investment will be measurable as an inflow, but the corresponding outflow—the put contract to potentially sell pesos—may not be (Garber, 1998).

If capital controls are designed to permit monetary autonomy, one can examine the extent to which onshore interest

rates—subject to capital controls—differ from those found in offshore markets or domestic currency returns on foreign assets. Such tests assume that returns on comparable investments in the same currency should be equal in the absence of effective capital controls. To the extent that they differ, the capital controls are effective (Harberger, 1980; Edwards, 1998b). This research has shown that capital controls have been able to create modest "wedges" of one to several percentage points between returns on similar domestic and international assets (Marston, 1995).[22] A related test to determine monetary autonomy is to measure the effectiveness of sterilization in preventing an appreciation of the real exchange rate.

Generally, controls on inflows have been found to be more effective than those on outflows because there is less incentive to evade controls on inflows (Reinhart and Smith, 1998; Eichengreen, et al. 1999).[23] Evading controls on inflows ordinarily will provide only marginal benefits for foreign investors, as the expected risk-adjusted domestic return usually will be comparable to that on alternative international investments. On the other hand, in the event of an expected devaluation, there is enormous incentive to avoid such a loss by evading controls on capital outflows. The expected loss on holding domestic assets can be several hundred percent in annualized terms over a short horizon. For example, if one expects the Malaysian ringgit to be devalued 10 percent in one week, the expected continuously compounded annual return associated with holding the currency through such a devaluation is almost –550 percent.[24] Therefore, researchers like Obstfeld (1998) and Eichengreen (1999) have found the idea of preventing destabilizing outflows by limiting inflows to be more promising than directly trying to stop outflows.

In sum, the consensus of the research on capital controls has been that they can alter the composition of capital flows or drive a small, permanent wedge between domestic and offshore interest rates but they cannot indefinitely sustain inconsis-

[21] An (American) put option confers on the holder the right, but not the obligation, to sell a specified quantity of pesos at a specified price, called the strike price or exercise price, on or before a given date.

[22] Fieleke (1994) finds that capital controls were of very limited effectiveness in creating interest differentials during the European Monetary System crises of 1992-93.

[23] As will be discussed in the next subsection, capital controls often are evaded by changing from prohibited to permitted assets or by falsifying invoices for traded goods.

[24] The continuously compounded annual return is computed from $52 * \ln(.9/1)$.

tent policies, and their effectiveness tends to erode over time as consumers and firms become better at evading the controls (Marston, 1995). Outflow restrictions, in particular, may buy breathing space, but that is all. There are more researchers willing to defend inflow restrictions, however. Eichengreen (1999) argues that, to restrain inflows, controls do not have to be perfect, they just need to make avoidance costly enough to reduce destabilizing flows.

How Are Capital Controls Evaded?

Over time, consumers and firms realize that they can evade capital controls through the channels used to permit trade in goods. Firms, for example, may evade controls on capital flows by falsifying invoices for traded goods; they apply to buy or sell more foreign exchange than the transaction calls for. For example, a domestic firm wishing to evade limits on capital inflows might claim that it exported $10 million worth of goods when it only, in fact, exported $9 million. It may use the excess $1 million to invest in domestic assets and split the proceeds with the foreign firm providing the capital.

Perhaps the most common method to evade controls on capital flows is through "leads and lags" in which trading firms hasten or delay payments for imports or exports (Einzig, 1968). To evade controls on outflows, for example, importers pay early for imports (leads), in exchange for a discount, and exporters allow delayed payments for their goods (lags), in return for a higher payment. This permits importers and exporters to effectively lend money to the rest of the world, a capital outflow. To evade controls on inflows, importers delay payments while exporters demand accelerated payments. Thus, leads and lags permit trade credit to substitute for short-term capital flows. Governments often attempt to close the leads/lags loophole on short-term capital flows with administrative controls on import/export financing.

Travel allowances for tourists are another method by which capital controls may be evaded (Bakker, 1996). More

recently, financial innovation has spawned financial instruments—derivatives—that may be used to mislead banking and financial regulators to evade prudential regulation and/or capital controls (Garber, 1998). For example, derivatives may contain clauses that change payouts in the event of defaults or the imposition of exchange controls (Garber, 1998). Improvements in information technology make it easier to buy and sell assets and reduce the effectiveness of capital controls (Eichengreen, et al. 1999).

Capital controls also induce substitution from prohibited to permitted assets (Goldstein, 1995). So, for example, the U.S. interest equalization tax was evaded through trade in assets with Canada while heavy Chilean taxes on short-term inflows may have induced a (desired) substitution to more lightly taxed longer-term inflows (Valdes, 1998). Capital controls have been more successful in changing the composition of asset trade than the volume.

Costs of Capital Controls

Although they often are evaded successfully, capital controls nonetheless impose substantial costs in inhibiting international trade in assets. Foremost among these costs are limiting the benefits of capital flows as described in Section 2: risk-sharing, diversification, growth, and technology transfer (*Global Investor*, 1998a). Capital exporting countries see a lower return on their savings while capital importers receive less investment and grow more slowly with capital controls. Krugman (1998) argues that capital controls do the most harm when they are used to defend inconsistent policies that produce an overvalued currency—a currency that would tend to depreciate or be devalued in the absence of the controls. This attempt to free governments from the discipline of the market permits poor or inconsistent policies to be maintained longer than they otherwise would, increasing the costs of these policies.

Poorly designed or administered capital controls often adversely affect direct investment and the ordinary financing of

trade deals (*Economist*, 1998). Controls can even worsen the problem of destabilizing capital flows. For example, the Korean government has acknowledged that the restriction on offshore borrowing by Korean corporations contributed to its balance of payments and banking crises in 1997 (*Global Investor*, 1998a). The corruption created by evasion and the administrative costs of controls also is an unintended cost of the controls. Even as the costs accumulate and their original purpose has ended, capital controls, like any regulation, develop their own constituencies and become difficult to phase out. The resumption of free capital flows does not always end the costs of capital controls. Specifically, blocking the departure of capital temporarily subsidizes investment but raises the perception of risk, increasing a risk premium and/or deterring future investment (*Economist*, 1998; Goldstein, 1995).

Partly because the costs of capital controls are serious and tend to worsen over time, economists have suggested attacking problems at their source rather than with capital controls (Krugman, 1998; Mishkin, 1998). For example, to cope with banks' incentive to take on excessive risk, a government might concentrate on reforming and strengthening the domestic financial structure—especially regulations on foreign borrowing—as it slowly phases out capital controls to derive the benefits of capital flows (Goldstein, 1995).[25] Or, to fight a real appreciation brought on by a capital inflow, a government might conduct contractionary fiscal policy. In all circumstances, better macroeconomic policy is needed to avoid financial crises, such as those that affected Asia in 1997. Countries must eschew overvalued currencies, excessive foreign debt, and unsustainable consumption.

CONCLUSION

Recently, a number of opinion leaders, including some prominent economists, have suggested that developing countries should reconsider capital controls. This article has reviewed the issues associated with capital controls. Controls most often have been used to permit more freedom for monetary policy during balance of payments crises in the context of fixed exchange rates. Restrictions on inflows have been implemented to prevent real appreciation of the exchange rate or to correct other pre-existing distortions, like the incentives for financial institutions to take excessive risk. Although controls on capital flows may change the composition of flows, they impose substantial costs on the economy and cannot be used to indefinitely sustain inconsistent policies. Under most circumstances, it is better to attack the source of the distortion or inconsistent policy at the source rather than treating symptoms through capital controls.

Although the worst of the Asian financial crisis seems to be over, it—like the peso crisis of December 1994—has been a sobering lesson in the volatility of capital flows and the fragility of emerging market financial systems. It also has raised questions for future research: Are limits on capital inflows the best solution to protecting domestic financial systems from the distortions inherent in banking? What is the proper sequence for economic reforms of the capital account, the current account, and the banking system?

REFERENCES

Alesina, Alberto, Vittorio Grilli, and Gian Maria Milesi-Ferretti. "The Political Economy of Capital Controls," in *Capital Mobility: The Impact on Consumption, Investment, and Growth*, Leonardo Leiderman and Assaf Razin, eds., Cambridge University Press, 1994, pp. 289-321.

Bakker, Age F.P. *The Liberalization of Capital Movements in Europe*, Kluwer Academic Publishers, 1996.

Barro, Robert. "Malaysia Could Do Worse than this Economic Plan," *Business Week*, November 2, 1998, p. 26.

Bernanke, Ben S. "Nonmonetary Effects of the Financial Crisis in Propagation of the Great Depression," *American Economic Review* (June 1983), pp. 257-76.

Bhagwati, Jagdish. "Why Free Capital Mobility may be Hazardous to Your Health: Lessons from the Latest Financial Crisis," NBER Conference on Capital Controls, November 7, 1998.

Blustein, Paul. "Is Malaysia's Reform Working? Capital Controls Appear to Aid Economy, but Doubts Remain," *Washington Post*, November 21, 1998, p. G01.

[25] An important but unresolved issue is the sequencing of reforms of the current account, the capital account, and the financial sector. Many blame the recent Asian crisis on the fact that the Asian governments moved faster to liberalize international capital flows than they did to regulate their financial system. In contrast, Sweeney (1997) argues for early liberalization of the capital account to provide accurate pricing information for business decisions.

Congressional Quarterly Service. "Interest Equalization Tax Background," *Congress and the Nation, 1965-1968, Vol. II*, 1969, p. 144.

_____. "Foreign Investment Controls," *Congress and the Nation, 1969-1972, Vol. III*, 1973a, p. 123.

_____. "1971 Smithsonian Agreement Broke Down in 14 Months," *Congress and the Nation, 1969-1972, Vol. III*, 1973b, p. 133.

_____. "Interest Equalization Tax," *Congress and the Nation, 1973-1976, Vol. IV*, 1977, p. 85.

Corbo, Vittorio, and Leonardo Hernandez. "Macroeconomic Adjustment to Capital Inflows: Lessons from Recent Latin American and Asian Experiences," *World Bank Research Observer* (February 1996), pp. 61-85.

Corsetti, Giancarlo, Paolo Pesenti, and Nouriel Roubini. "What Caused the Asian Currency and Financial Crisis? Part I: A Macroeconomic Overview," NBER Working Paper 6833, December 1998a.

_____. "What Caused the Asian Currency and Financial Crisis? Part II: The Policy Debate," NBER Working Paper 6834, December 1998b.

Department of Commerce. *The Balance of Payments of the United States: Concepts, Sources and Estimation Procedures*, U.S. Government Printing Office, 1990.

Diamond, Douglas W., and Philip H. Dybvig. "Bank Runs, Deposit Insurance, and Liquidity," *Journal of Political Economy* (June 1983), pp. 401-19.

Dixit, Avinash K., and Robert S. Pindyck. *Investment Under Uncertainty*, Princeton University Press, 1994.

Dooley, Michael P. "A Survey of Literature on Controls of International Capital Transactions," IMF Staff Papers (December 1996), pp. 639-87.

Dornbusch, Rudiger. "Capital Controls: An Idea Whose Time is Past," in *Should the IMF Pursue Capital-Account Convertibility?*, Princeton Essays in International Finance No. 207, May 1998, pp. 20-27.

Economist. "One Year Old and Not Yet Born?" July 18, 1964a, p. 283.

_____. "Where Will all the Borrowers Go?" August 4, 1964b, pp. 565-67.

_____. "How Far is Down?" November 15, 1997, pp. 19-21.

_____. "The Perils of Global Capital," April 11, 1998, pp. 52-54.

Edwards, Sebastian. "Capital Controls Are Not the Reason for Chile's Success," *Wall Street Journal*, April 3, 1998a, p. A19.

_____. "Capital Flows, Real Exchange Rates and Capital Controls: Some Latin American Experiences," NBER Working Paper 6800, November 1998b.

Eichengreen, Barry. *Toward a New International Financial Architecture: A Practical Post-Asia Agenda*, Institute for International Economics, 1999.

_____, Michael Mussa, Giovanni Dell'Ariccia, Enrica Detragiache, Gian Maria Milesi-Ferretti, and Andrew Tweedie. "Liberalizing Capital Movements: Some Analytical Issues," IMF Economic Issue No. 17, February 1999.

Einzig, Paul. *Leads and Lags*, MacMillan and Company, 1968.

Fieleke, Norman S. "International Capital Transactions: Should They Be Restricted?" *New England Economic Review* (March/April 1994), pp. 27-39.

Frankel, Jeffrey A., and Andrew K. Rose. "Currency Crashes in Emerging Markets: An Empirical Treatment," *Journal of International Economics* (November 1996), pp. 351-66.

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

Garber, Peter. "Derivatives in International Capital Flow," NBER Working Paper 6623, June 1998.

Global Investor. "Managing Currency Risk in Tumultuous Times," *Emerging Market Currencies: A Guide for Investors Supplement* (October 1998a), pp. 2-4.

_____. "Malaysia's Exchange Controls: Delaying the Inevitable," *Emerging Market Currencies: A Guide for Investors Supplement* (October 1998b), pp. 12-14.

Goldstein, Morris. "Coping with Too Much of a Good Thing: Policy Responses for Large Capital Inflows to Developing Countries," Institute for International Economics, 1995.

Grilli, Vittorio, and Gian Maria Milesi-Ferretti. "Economic Effects and Structural Determinants of Capital Controls," *IMF Staff Papers* (September 1995), pp. 517-51.

Harberger, Arnold C. "Vignettes on the World Capital Market," *American Economic Review* (May 1980), pp. 331-37.

_____. "Economic Adjustment and the Real Exchange Rate," in *Economic Adjustment and Exchange Rates in Developing Countries*, Sebastian Edwards and Liaquat Ahamed, eds., University of Chicago Press, 1986, pp. 371-414.

Johnston, Barry R., and Natalia T. Tamirisa. "Why Do Countries Use Capital Controls?" IMF Working Paper 98-181, December 1998.

Kaminsky, Graciela L., and Carmen M. Reinhart, "The Twin Crises: The Causes of Banking and Balance of Payments Problems," *American Economic Review* (June 1999), pp. 473-99.

Kasa, Kenneth. "Time for a Tobin Tax?" FRBSF Economic Letter 99-12, April 9, 1999.

Kindleberger, Charles P. *Manias, Panics, and Crashes: A History of Financial Crises*, Macmillan, 1978.

Kreinin, Mordechai. *International Economics: A Policy Approach*, Harcourt, Brace, Jovanovich, 1971.

Krugman, Paul. "An Open Letter to Prime Minister Mahathir," September 1, 1998.

Laffer, Arthur, and Marc A. Miles. *International Economics in an Integrated World*, Scott, Foresman and Company, 1982.

Larraín B., Felipe, Raúl Labán M., and Rómulo A. Chumacero. "What Determines Capital Inflows?: An Empirical Analysis for Chile," Faculty Research Working Paper Series, Kennedy School of Government, Harvard University, 97-09, April 1997.

Marston, Richard C. *International Financial Integration: A Study of Interest Differentials Between the Major Industrial Countries*, Cambridge University Press, 1995.

McKinnon, Ronald I., and Wallace E. Oates. "The Implications of International Economic Integration for Monetary, Fiscal and Exchange Rate Policies," The International Finance Section, Princeton University, 1966.

Minton, Zinny, "Global Finance Survey: A Wealth of Blueprints," *Economist*, January 30, 1999, pp. S5-S8.

Mishkin, Frederic S. "International Capital Movements, Financial Volatility and Financial Instability," NBER Working Paper 6390, January 1998.

Mitchell, B.R. *International Historical Statistics: Africa, Asia & Oceania 1750-1993*, Macmillan Reference, 1998.

New York Times. "How U.S. Wooed Asia to Let Cash Flow In," February 16, 1999, p. A1.

Obstfeld, Maurice. "The Global Capital Market: Benefactor or Menace?" Working Paper, 1998.

_____ and Alan M. Taylor. "The Great Depression as a Watershed: International Capital Mobility over the Long Run," in *The Defining Moment: The Great Depression and the American Economy in the Twentieth Century*, Michael D. Bordo, Claudia D. Goldin, and Eugene N. White, eds., University of Chicago Press, 1998, pp. 353-402.

Park, Yung Chul, and Chi-Young Song. "Managing Foreign Capital Flows: The Experiences of Korea, Thailand, Malaysia and Indonesia," Jerome Levy Economics Institute Working Paper 163, May 1996.

Pearce, David W., ed. *MIT Dictionary of Modern Economics*, MIT Press, 1995.

Reinhart, Carmen M., and R. Todd Smith. "Too Much of a Good Thing: The Macroeconomic Effects of Taxing Capital Inflows," in *Managing Capital Flows and Exchange Rates: Perspectives from the Pacific Basin*, Reuven Glick, ed., Cambridge University Press, 1998, pp. 436-64.

Stern, Robert. *The Balance of Payments: Theory and Economic Policy*, Aldine Publishing Company, 1973.

Sweeney, Richard J. "The Information Costs of Capital Controls," in *Capital Controls in Emerging Economies*, Christine P. Ries and Richard J. Sweeney, eds., Westview Press, 1997, pp. 45-61.

Torres, Craig. "Chilean Bid to Boost Confidence Lauded," *Wall Street Journal*, June 29, 1998, p. A14.

ul Haq, Mahbub, Inge Kaul, and Isabelle Grunberg, eds. *The Tobin Tax: Coping with Financial Volatility*, Oxford University Press, 1996.

Valdes, Salvador. "Capital Controls in Chile Were a Failure," *Wall Street Journal*, December 11, 1998, p. A15.

_____ and Marcelo Soto. "New Selective Capital Controls in Chile: Are they Effective?" Catholic University of Chile Working Paper, 1996.

Wade, Robert, and Frank Veneroso. "The Gathering Support for Capital Controls," *Challenge* (November/December 1998), pp. 14-26.

The Role of Supervisory Screens and Econometric Models in Off-Site Surveillance

R. Alton Gilbert, Andrew P. Meyer, and Mark D. Vaughan

Banking is one of the more closely supervised industries in the United States, reflecting the view that bank failures have stronger adverse effects on economic activity than other business failures. Bank failures can disrupt the flow of credit to local communities (Gilbert and Kochin, 1989), interfere with the operation of the payments system (Gilbert and Dwyer, 1989), and reduce the money supply (Friedman and Schwartz, 1963). Bank failures also can have lingering effects on the real economy. Indeed, a growing body of literature blames the length of the Great Depression on the disruption of credit relationships that followed the wave of bank failures during the early 1930s (Bernanke, 1983; Bernanke, 1995; and Bernanke and James, 1991).

The existence of unfairly priced deposit insurance bolsters the case for bank supervision. Without insurance, depositors have strong incentives to monitor and discipline risky institutions by withdrawing funds or demanding higher interest rates. Insured depositors, in contrast, have little incentive to monitor and

discipline risk (Flannery, 1982). Moreover, deposit insurance premiums established under the Federal Deposit Insurance Corporation Improvement Act of 1991 (FDICIA) do not appear to punish risk adequately. The spread between the premiums paid by the riskiest and safest banks is only 27 basis points, and just 562 of the 10,486 FDIC-insured institutions paid any premiums during the first half of 1999 (Barancik, 1999). As a result, bank supervisors must act as agents of the taxpayers to limit risk. Supervisory limits on bank risk reduce the likelihood that failures will exhaust the deposit insurance fund and impose direct costs on the taxpayers.[1]

Bank supervisors use on-site examination and off-site surveillance to identify banks likely to fail. Supervisors then can take steps to reduce the likelihood that these institutions will fail. The most useful tool for identifying problem institutions is on-site examination, in which examiners travel to a bank and review all aspects of its safety and soundness. On-site examination is, however, both costly and burdensome: costly to supervisors because of its labor-intensive nature and burdensome to bankers because of the intrusion into day-to-day operations. As a result, supervisors also monitor bank condition off-site. Off-site surveillance yields an ongoing picture of bank condition, enabling supervisors to schedule and plan exams efficiently. Off-site surveillance also provides banks with incentives to maintain safety and soundness between on-site visits.

In off-site surveillance, supervisors rely primarily on two analytical tools: supervisory screens and econometric models. Supervisory screens are combinations of financial ratios, derived from bank balance sheets and income statements, that have, in the past, given forewarning of safety-and-soundness problems. Supervisors

[1] See White (1991) for a discussion of the role of lax government supervision in the thrift debacle of the 1980s.

draw on their experience to weigh the information content of these ratios. Econometric models also combine information from bank financial ratios. These models, however, rely on a computer rather than judgement to combine ratios, boiling the information about bank condition in the financial statements down to one number. In some models this number represents the likelihood that a bank will fail. In others, the number represents the supervisory rating that would be awarded if the bank were examined today.

In past statistical comparisons, econometric models have outperformed supervisory screens, yet screens continue to enjoy considerable popularity in the surveillance community. Cole, Cornyn, and Gunther (1995) demonstrated that the Federal Reserve's econometric model, the System for Estimating Examination Ratings (SEER), outperformed a surveillance approach based on screens (the Uniform Bank Surveillance System or UBSS), both as a predictor of failures and as an identifier of troubled institutions. Nonetheless, analysts at the Board of Governors and in each of the Reserve Banks continue to generate a variety of screens to aid in exam scheduling and scoping. To economists who are not involved in day-to-day surveillance, the continuing popularity of screens is somewhat puzzling.

We explore two possible explanations for the popularity of screens: (1) perhaps the extra precision of econometric models is not worth the added cost, or (2) perhaps the flexibility of screens makes them particularly attractive in today's dynamic banking environment. Although models can tease information out of bank financials that the human eye might overlook, they are more costly to operate than screens, requiring surveillance analysts to learn to interpret complex statistical output. If models only marginally outperform screens in flagging banks headed for problems, then the marginal benefit of the extra precision might not exceed the marginal learning costs. Another possible explanation for the attachment to screens is the ease with which they can be adapted to new environ-

ments. The last 15 years have witnessed remarkable change in the banking industry. In such a fluid environment, screens can be adapted to reflect changes in the sources of safety-and-soundness problems faster than econometric models.

We demonstrate that econometric models still significantly outperform supervisory screens in statistical horse races, implying that the marginal benefit of using models does indeed outweigh any marginal learning costs. Specifically, we use data from the 1980s and 1990s to compare the performance of supervisory screens and econometric models as tools for predicting failures 12 to 24 months in the future. We highlight the resource savings associated with using each approach rather than random examination. We also estimate an econometric model designed to predict the likelihood that a bank, currently considered safe and sound, will suffer a significant slip in its supervisory rating in 12 to 24 months. Finally, we demonstrate how econometric models can be used to pinpoint the source of developing problems.

Despite the statistical advantages of using econometric models, screens can still add tremendous value in off-site surveillance. In today's fast-changing world of banking, supervisors can modify screens well before econometric models can be re-estimated. Moreover, experience with new screens then can inform the respecification of econometric models. In short, supervisory screens and econometric models play important complementary roles in allocating examination resources.

ON-SITE AND OFF-SITE SURVEILLANCE: A CLOSER LOOK

To appreciate the roles of models and screens in off-site surveillance, it is important to first place these tools in the overall framework of bank supervision. Bank supervisors rely principally on regular on-site examinations to maintain bank safety and soundness. Examinations ensure the integrity of bank financial statements and identify banks that should be subject to

supervisory sanctions.[2]   During a routine exam, examiners assess six components of safety and soundness—capital protection (C), asset quality (A), management competence (M), earnings strength (E), liquidity risk (L) and market risk (S)—and assign a grade of 1 (best) through 5 (worst) to each component. Examiners then use these six scores to award a composite rating, also expressed on a 1 through 5 scale.[3]   At present, most banks boast 1 or 2 CAMELS composites. Indeed, at year-end 1998, only 285 of 8,264 U.S. banks carried 3, 4, or 5 composite ratings.

Although on-site examination is the most effective tool for constraining bank risk, it is both costly to supervisors and burdensome to bankers. As a result, supervisors face continuous pressure to limit exam frequency.   During the 1980s, supervisors yielded to this pressure, and many banks escaped yearly examination (Reidhill and O'Keefe, 1997). In 1991, however, the Federal Deposit Insurance Corporation Improvement Act (FDICIA) required annual examinations for all but a handful of small, well-capitalized, highly rated banks, and even these institutions must be examined every 18 months. This new mandate reflected the lessons learned from the wave of failure during the late 1980s, namely that more frequent exams, though likely to increase the up-front costs of supervision, reduce the down-the-road costs of resolving failures by revealing problems at an early stage.

Although recent changes in public policy have mandated greater exam frequency, supervisors still can use off-site surveillance tools to flag banks for accelerated exams and to plan regularly scheduled, as well as accelerated exams. Bank condition can deteriorate rapidly between on-site visits (Cole and Gunther, 1998). In addition, the Federal Reserve now employs a "risk-focused" approach to exams, in which supervisors allocate on-site resources according to the risk exposures of the bank (Board of Governors, 1996). Off-site surveillance helps supervisors allocate on-site resources efficiently by identifying institutions that need immediate attention

and by pinpointing risk exposures for regularly scheduled as well as accelerated exams. For these reasons, an interagency body of bank and thrift supervisors—the Federal Financial Institutions Examinations Council (FFIEC)—requires banks to submit quarterly Reports of Condition and Income, often referred to as call reports. Surveillance analysts then use call report data to conduct financial statement analysis between exams.

Using their field experience as a guide, supervisors have developed rules of thumb for exam scheduling and scoping with call report data.[4] These rules of thumb are called supervisory screens. To give an example of the use of screens, supervisors might flag a bank for an accelerated examination (or plan to allocate more resources to a given area on a scheduled exam) if a certain financial ratio, like a risk-based capital

Table 1	

How to Interpret CAMELS Composite Ratings

CAMELS Composite Rating	Description
1	Financial institutions with a composite-1 rating are sound in every respect and generally have individual component ratings of 1 or 2.
2	Financial institutions with a composite-2 rating are fundamentally sound. In general, a 2-rated institution will have no individual component ratings weaker than 3.
3	Financial institutions with a composite-3 rating exhibit some degree of supervisory concern in one or more of the component areas.
4	Financial institutions with a composite-4 rating generally exhibit unsafe and unsound practices or conditions. They have serious financial or managerial deficiencies that result in unsatisfactory performance.
5	Financial institutions with a composite-5 rating generally exhibit extremely unsafe and unsound practices or conditions. Institutions in this group pose a significant risk for the deposit insurance fund and their failure is highly probable.

Source: *Federal Reserve Commercial Bank Examination Manual*

[2] See Flannery and Houston (1999) for evidence that holding company inspections help ensure the integrity of financial statements. See Gilbert and Vaughan (1998) for a discussion of the sanctions available to bank supervisors.

[3] See Hall, King, Meyer, and Vaughan (1999) for a discussion of the factors used to assign individual and composite ratings.

[4] See Putnam (1983) for a description of the use of supervisory screens in off-site surveillance during the late 1970s and early 1980s.

ratio, is suspect. Another example might be a rule that flags a bank if 10 out of 15 ratios either exceed or fall short of desired levels. This approach offers two advantages: simplicity and flexibility. An experienced supervisor can detect emerging problems easily, as well as the sources of these problems, without sophisticated statistical analysis. An experienced supervisor also can easily modify the screens in changing banking environments. On the negative side, supervisors who rely only on subjective judgment to "screen" might miss subtle but important interactions among financial ratios.

Econometric models offer a more systematic way to combine call report data for scheduling and scoping. A common type of model used in surveillance estimates the marginal impact of a change in a financial ratio on the probability that a bank will fail, holding all other ratios constant. These models can examine ratios simultaneously, capturing subtle but important interactions. The Federal Reserve uses two different models in off-site surveillance. One model combines financial ratios to estimate the probability that each Fed-supervised bank will fail within the next two years. Another model estimates the CAMELS rating that would be awarded based on the bank's latest financial statements. Every quarter, economists at the Board of Governors feed the latest call report data into these models and forward the results to each of the 12 Reserve Banks. Surveillance analysts in the Reserve Banks then investigate the institutions that the models flag as "exceptions."

SPECIFYING REPRESENTA-TIVE VERSIONS OF SUPER-VISORY SCREENS AND ECONOMETRIC MODELS

To compare the performance of supervisory screens and econometric models, we first specified a representative version of each surveillance tool. To specify a set of supervisory screens, we interviewed safety-and-soundness officers and

examiners in the Eighth Federal Reserve District. To specify an econometric model, we reviewed the academic literature. After conducting interviews and reviewing literature, we identified a set of financial ratios common to both approaches. We included only these common ratios in our representative screens and models to facilitate a comparison of relative performance. The financial ratios common to both the screens and the models reflect the individual components of bank condition in the CAMEL framework. (Bank regulators added the "S" to the CAMEL framework on January 1, 1997. During our sample period, however, examiners explicitly graded only five aspects of safety and soundness.) Although our screens and models are representative of the screens and models regularly used in off-site surveillance, they are not identical to the tools currently used by the Board of Governors or the individual Reserve Banks.

In both the screens and models, we used the ratio of total equity to total assets (EQUITY) to assess capital adequacy. Higher levels of capital protection provide a larger buffer against losses and increase the owners' stake in the bank. We expect, therefore, that higher levels of capital will reduce the likelihood of safety-and-soundness problems. A safety-and-soundness problem first is defined as an outright failure; later in the paper we define a safety-and-soundness problem as a downgrade from a CAMEL-1 or CAMEL-2 rating to a CAMEL-3, CAMEL-4, or CAMEL-5 rating.

We gauged asset quality with three different measures: the ratio of nonperforming loans to total loans (BAD-LOANS), the ratio of consumer loans to total assets (CONSUMER), and the ratio of other real estate owned to total loans (OREO). Nonperforming loans are loans that are 90 or more days past due or in nonaccrual status. (In bank accounting, loans are either classified as accrual or nonaccrual. As long as a loan is classified as accrual, the interest due is counted as current revenue, even if the borrower falls behind on interest payments.) We used the nonperforming loan ratio as a measure of asset quality because banks ultimately charge off

relatively high percentages of nonperforming loans. We used the consumer loan ratio because the charge-off rate for consumer loans has been higher historically than for other types of loans. For example, nationwide, the average charge-off rate for all types of bank loans from 1990 through 1997 was 0.86 percent; for consumer loans, the average was 2.08 percent. Finally, we included "other real estate owned" because the term generally applies to collateral seized after loan defaults; banks with higher OREO ratios tend to have more credit risk exposure. We expect that banks with higher values of these ratios will experience more safety-and-soundness problems.

As proxies for managerial competence, we used noninterest expense as a percentage of total revenue (OVERHEAD), insider loans as a percentage of total assets (INSIDER), and occupancy expense as a percentage of average assets (OCCUPANCY). Because well-managed banks hold down overhead costs, avoid excessive lending to insiders, and pay reasonable amounts for office space, we expect that banks with higher values of these ratios will suffer more safety-and-soundness problems.

We measured earnings strength with the ratio of net income to total assets (return on assets, or ROA), and the ratio of interest income accrued, but not collected, to total loans (UNCOLLECTED). All other things being equal, higher earnings provide a greater cushion for withstanding adverse economic shocks. We expect, therefore, that higher returns on assets will reduce the likelihood of safety-and-soundness problems. Banks with high levels of interest income accrued but not collected are vulnerable to large restatements of earnings and capital because the loans generating accrued-interest-that-has-not-been-collected could be reclassified as nonaccrual. We expect, therefore, that higher levels of uncollected interest income point to future safety-and-soundness problems.

We gauged liquidity risk with three measures: liquid assets (cash, securities, federal funds sold, and reverse repurchase agreements) as a percentage of total assets (LIQUID), large time deposits as a percentage of total assets (LARGE-TIME), and core deposits as a percentage of total assets (CORE). A larger stock of liquid assets indicates greater ability to meet unexpected liquidity needs. Larger stocks of liquid assets, therefore, should translate into fewer safety-and-soundness problems. Liquidity risk also depends on the division of bank liabilities between volatile and core funding. Large time deposits represent a volatile source of funding because they are not fully insured by the FDIC; a sudden jump in market interest rates or a sudden deterioration in bank condition could raise funding costs dramatically. All other things being equal, greater reliance on large time deposits implies a greater likelihood of safety-and-soundness problems. Similarly, the smaller a bank's volume of nonvolatile or core deposits, the greater the likelihood of safety-and-soundness problems.

Finally, we included control variables for bank size and holding company affiliation in the representative versions of the screens and models. We added the natural logarithm of total assets (SIZE) because larger banks should be better able to diversify across product lines and geographic regions and, therefore, avoid safety-and-soundness problems. We also added a control variable to capture the effect of holding company affiliation. This variable, BHCRATIO, equaled the ratio of total assets in the sample bank to total assets in all banks in the parent-holding company. Because holding companies are better able to serve as a source of strength for their smaller members, we expect that lower values of BHCRATIO imply fewer safety-and-soundness problems in the future. (The shaded insert discusses the holding company control variable in more detail.) Table 2 presents a complete list of the variables used in this article as the supervisory screens and as independent variables in the econometric models. The table also includes a positive or negative sign indicating the hypothesized relationship between each variable and the likelihood of outright failure or CAMEL downgrade from CAMEL 1 or 2 to CAMEL 3, 4, or 5.

WHY CONTROL FOR HOLDING COMPANY MEMBERSHIP?

It may seem curious that we included a variable related to holding company membership in the supervisory screens and the econometric model. We included this variable because theory and evidence suggest that small banks belonging to large holding companies are less likely to fail or suffer supervisory downgrades.

To see why small banks belonging to large holding companies are less likely to encounter safety-and-soundness problems, suppose that such a bank is facing serious asset quality problems. The owners of the holding company must confront a trade-off when deciding whether to inject equity into this subsidiary. On the one hand, alternative investments are likely to offer higher returns because loan losses will absorb some of the injections. On the other hand, not injecting equity into the troubled subsidiary could lead to a failure, which, in turn might taint the reputation of the holding company in the eyes of financial markets or bank supervisors. Because the bank is small, the injection is more likely to prevent a failure and the attendant reputational damage. In short, when a subsidiary bank is relatively small, the holding company is better able to serve as a source of strength.

For this reason, we added BHCRATIO, the assets of the sample bank divided by the total assets of all bank subsidiaries of its holding company, to the list of screens and explanatory variables. BHCRATIO assumed a value of unity when the sample bank did not belong to a holding company or was the only bank in the holding company. All other things being equal, the smaller the assets of the sample bank relative to the assets of the holding company, the smaller the value of BHCRATIO. We expect to observe a positive relationship between BHCRATIO and future safety-and-soundness problems (failures or downgrades of CAMEL ratings to problem status).

Empirical studies confirm that BHCRATIO helps explain both bank failures and capital injections into troubled holding company subsidiaries. Belongia and Gilbert (1990) found that a variable constructed like BHCRATIO enhanced the explanatory power of a model of agricultural bank failures: the smaller the agricultural banks relative to the size of their parent organizations, the lower their probabilities of failure. Gilbert (1991) also found that a variable constructed like BHCRATIO helped explain equity injections into undercapitalized banks; the smaller the under-capitalized banks relative to the size of their parent organizations, the larger the equity injections into the undercapitalized banks.

Taken together, our empirical evidence supports the hypothesis that BHCRATIO is positively related to both failures and CAMEL downgrades. When used as a screen, the means differed in the hypothesized direction in two of the three failure samples (1988 and 1989) and six of the seven downgrade samples. When used in the econometric model, the coefficient on BHCRATIO was positive and significant in only one of the three failure prediction models (1987), but it was positive and statistically significant in all the CAMEL downgrade equations. The lack of supporting evidence from the failure prediction screens and models may be the result of the Texas bank failures of the late 1980s. In several prominent cases, regulators shut down entire holding companies even when many of the subsidiary banks were safe and sound. See Cannella, et. al. (1995) for additional discussion of the closure of these holding companies and banks.

Table 2

What Variables Help Predict Bank Failures or CAMEL Downgrades?

This table lists the single-variable screens and independent variables used in our econometric models. The sign indicates the hypothesized relationship between the variable and the likelihood of a safety-and-soundness problem. For example, the negative sign for the equity-to-assets ratio indicates that a higher capital ratio would reduce the likelihood of a failure or CAMEL downgrade.

Symbol	Description	Hypothesis about sign of coefficient for predicting failure or CAMEL downgrades (positive sign indicates positive correlation with probability of failure or rating downgrade).
EQUITY	Equity as a percentage of total assets.	−
BAD-LOANS	Nonperforming loans as a percentage of total loans.	+
OREO	Other real estate owned (real estate other than bank premises) as a percentage of total loans.	+
CONSUMER	Consumer loans as a percentage of total assets.	+
INSIDER	The value of loans to insiders (officers and directors of the bank) as a percentage of total assets.	+
OVERHEAD	Noninterest expense as a percentage of total revenue.	+
OCCUPANCY	Occupancy expense as a percentage of average assets.	+
ROA	Net income as a percentage of total assets.	−
UNCOLLECTED	Interest accrued as revenue but not collected as a percentage of total loans.	+
LIQUID	Liquid assets (sum of cash, securities, federal funds sold, and reverse repurchase agreements) as a percentage of total assets.	−
LARGE-TIME	Large denomination time deposit liabilities as a percentage of total assets.	+
CORE	Core deposits (transactions, savings and small time deposits) as a percentage of total assets.	−
SIZE	Natural logarithm of total assets, in thousands of dollars.	−
BHCRATIO	The ratio of each bank's total assets to the total assets of its holding company. Banks without holding companies have BHCRATIO ≡ 1.	+

Number of Commercial Bank Failures by Year 1934-97
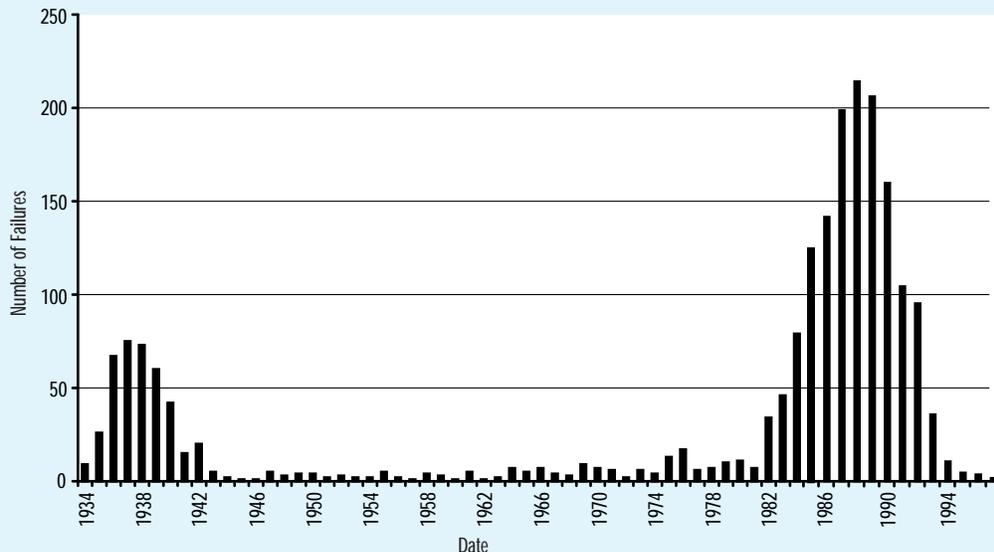
This figure shows that U.S. commercial bank failures peaked in 1988 and dropped precipitously during the 1990s.

GAUGING SUPERVISORY SCREENS AND ECONOMETRIC MODELS AS PREDICTORS OF BANK FAILURE

We began by using the representative supervisory screens on historical data to gauge how well they would have predicted bank failures during 1989, 1990, and 1991. To conduct these tests, we partitioned a list of all U.S. banks during those years into failures and survivors for each year. The sample ended in 1991 because so few banks failed after the early 1990s (see Figure 1). We then used 1987, 1988, and 1989 call report data to generate screen values for the sample banks two years before the observation of failure or survival. An individual screen would provide early warning if the mean value of the screen for the failed banks differed significantly from the mean value for the survivor banks in the direction hypothesized. The capital screen, for example, would meet this condition if the mean equity-to-asset (EQUITY) ratio for the failed banks was significantly below the mean ratio for the

surviving banks two years before the observation of failure or survival. Table 3 presents the means and standard deviations of the screen ratios for both banks that failed and banks that survived.

Overall, the individual screens would have done a good job predicting bank failures during 1989, 1990, and 1991. For 11 of the 14 variables, the average screen values for the failed and surviving banks differed significantly in the hypothesized direction across all three years. Indeed, only the consumer loans screen, the core deposit screen, and the size control variable failed to correlate consistently with future failures. The capital screen clearly illustrates the signaling value of individual supervisory screens. In all three years, the differences in means were economically large and statistically significant—banks with weaker capital ratios were more likely to fail. For example, the fourth-quarter 1987 equity-to-asset ratio for banks that would fail during 1989 (4.30 percent) was well below the ratio for banks that would survive that year (8.50 percent).

Table 3

How Well Do the Individual Screens Predict Bank Failures?

This table presents evidence about the failure prediction record of individual supervisory screens. The far-left and right columns for each year contain the mean values of the screens; standard deviations appear in parentheses below the means. An asterisk indicates a significant difference (at the 5-percent level) between the means for failed and survivor banks. Shading highlights screens with significant predictive power in all three years. The center column for each year (‡) shows the number of survivor banks with screen values worse than those of the average failed bank; the larger this number, the worse the performance of the screen. Taken together, this evidence shows that screens warn of potential failures but also can lead to many unnecessary exams.

	Data as of 1987:4 for:			Data as of 1988:4 for:			Data as of 1989:4 for:		
	149 banks that failed in 1989	‡	11,838 banks that survived 1989	115 banks that failed in 1990	‡	11,446 banks that survived 1990	82 banks that failed in 1991	‡	11,246 banks that survived 1991
EQUITY	4.30* (2.23)	359	8.50 (3.09)	3.38* (3.82)	180	8.58 (3.22)	4.24* (2.36)	273	8.69 (3.38)
BAD-LOANS	8.19* (6.02)	612	2.54 (2.95)	8.22* (5.20)	386	2.16 (2.52)	6.79* (4.03)	493	2.02 (2.56)
OREO	6.85* (9.71)	360	1.28 (2.26)	7.36* (7.29)	317	1.24 (2.32)	5.24* (5.68)	631	1.22 (2.46)
CONSUMER	10.54 (8.68)	4,817	10.79 (7.82)	12.72* (10.33)	3,361	10.74 (7.98)	12.63 (12.03)	3,380	10.71 (7.97)
INSIDER	1.09* (2.20)	1,724	0.52 (0.92)	1.51* (2.35)	1,074	0.54 (1.10)	1.00* (1.12)	1,850	0.53 (0.96)
OVERHEAD	46.62* (26.12)	1,277	35.04 (22.33)	49.79* (16.59)	741	34.11 (10.50)	41.36* (12.33)	1,423	32.05 (9.89)
OCCUPANCY	0.66* (0.39)	2,276	0.49 (0.31)	0.80* (0.42)	1,185	0.49 (0.30)	0.76* (0.41)	1,383	0.48 (0.31)
ROA	-2.16* (3.19)	358	0.67 (1.16)	-2.55* (2.73)	177	0.80 (1.11)	-1.28* (1.67)	336	0.87 (1.03)
UNCOLLECTED	0.96* (0.62)	2,037	0.67 (0.39)	0.94* (0.47)	2,418	0.71 (0.40)	0.97* (0.42)	2,594	0.76 (0.43)
LIQUID	32.99* (13.86)	2,702	45.36 (15.24)	32.76* (11.84)	2,776	44.43 (15.14)	27.82* (10.33)	1,469	43.87 (14.84)
LARGE-TIME	22.98* (13.04)	757	9.27 (7.90)	17.07* (8.25)	1,666	9.65 (7.41)	14.77* (7.83)	2,390	10.06 (7.30)
CORE	69.42* (14.32)	1,466	79.41 (9.73)	77.33 (9.22)	3,796	78.90 (9.53)	77.23 (10.98)	3,904	78.38 (9.42)
SIZE	10.98 (1.35)	7,374	10.79 (1.24)	10.72 (1.09)	5,876	10.84 (1.26)	11.26* (1.55)	7,717	10.89 (1.27)
BHCRATIO	0.62* (0.44)	8,517	0.75 (0.39)	0.83* (0.31)	7,849	0.75 (0.39)	0.92* (0.23)	7,571	0.75 (0.39)

Table 4

What Were the CAMEL Ratings of Banks that Failed in 1989, 1990, and 1991?

This table shows that supervisors already were aware of problems in most of the banks that failed in 1989, 1990, and 1991. Shading highlights the failure record of problem banks (CAMEL 3, 4, or 5). Supervisors recognize that these banks are significant failure risks and, therefore, monitor them closely. CAMEL-1 or -2 banks rarely fail, so they are not monitored as closely.

Rate of Bank Failure by Prior CAMEL Rating

Date of Rating (Calendar Year of Failure)	CAMEL Rating	Number of Banks	Number of Failures	Percentage Failed
March 1988 (1989)	1	1,908	0	0.00%
	2	5,029	6	0.12
	3	1,493	30	2.01
	4	643	52	8.09
	5	115	27	23.48
March 1989 (1990)	1	2,409	0	0.00
	2	6,130	10	0.16
	3	1,585	19	1.20
	4	673	48	7.13
	5	139	36	25.90
March 1990 (1991)	1	2,573	0	0.00
	2	6,423	9	0.14
	3	1,474	14	0.95
	4	629	31	4.93
	5	158	27	17.09

A better measure of the value added by individual screens, however, is their record in identifying failure candidates that were not already on supervisors' watch lists. Suppose, for example, that it is March 1988, and supervisors are scheduling and staffing exams for the rest of the year. Most of the banks with CAMEL composite ratings below 2 already are under scrutiny, so supervisors would like to use the latest call report data (year-end 1987) to identify CAMEL 1 or 2-rated banks that are significant failure risks in 1989. A tool that accurately predicted the 1989 failures of CAMEL 3, 4, and 5-rated banks, but did a poor job predicting the failures of CAMEL 1 or 2-rated banks, would not add much value in off-site surveillance because it would give supervisors little new information.

With this standard in mind, we looked again at the failure prediction record of the single-variable screens for 1989, 1990, and 1991. First, we identified all the CAMEL-2 banks as of March 1988, 1989, and 1990 and partitioned that set into banks that failed and banks that did not fail the following calendar year. We then generated the corresponding screen values using call report data from the previous December. Finally, we calculated the percentage of CAMEL-2 banks that would have to be examined, using each screen as a guide, to flag one-half of the CAMEL-2 banks that failed the next year. We selected one-half of the failures as a threshold because catching all of the CAMEL-2 failures would require, in some cases, examining most of the CAMEL-2 banks. We looked at only CAMEL-2 banks because no banks rated CAMEL 1 as of March 1988, March 1989, or March 1990 failed during the following calendar year. Table 4 puts the CAMEL-2 failure numbers in perspective by showing the failure rates for each CAMEL cohort,

while Table 5 shows the percentage of CAMEL-2 banks that must be examined, using each screen, to catch one-half of the failures the next year.

The evidence for 1989, 1990, and 1991 failures shows that single-variable screens would have improved significantly over random examination of CAMEL-2 banks. In each of the years, several screens were particularly informative. The large-time-deposits-to-total-assets ratio, for example, outperformed the other 13 screens as a tool for identifying 1989 failures. Had supervisors used the fourth-quarter 1987 value of this ratio as a guide, they would have caught one-half of the 1989 failures after examining only 1.7 percent of the CAMEL-2 banks. For 1990 failures, the return-on-asset screen was dominant; had supervisors scheduled exams using fourth-quarter 1988 values of this screen they would have caught one-half of 1990's failures after visiting only 0.9 percent of the CAMEL-2 banks. Finally, for 1991 failures, the non-performing loan screen turned in the best performance. Supervisors could have identified one-half of that year's failures by examining only 2.2 percent of CAMEL-2 banks. To put these numbers in perspective, if supervisors scheduled examinations randomly, on the average examiners would have had to visit 50 percent of CAMEL-2 banks to catch one-half of those that failed during the next 12 to 24 months. The average three-year performance of every single-variable screen except the consumer loan screen and the size control variable was well below 50 percent.

Next, we fit an econometric model to the data on bank failures and the measures used as screens to gauge how well it would have predicted failures. Again, we partitioned U.S. banks into failures and survivors for each year, assigning a "1" to banks that failed and a "0" to banks that survived. This binary observation served as the dependent variable in the model. As independent variables, we used the two-year lagged screen values, including the size and holding company control variables. We estimated a logit model—a specific type of econometric model used when the

dependent variable is a "0" or "1"—year by year; that is, we fit the model to 1985 screen values and 1987 failure observations, then to 1986 screen values and 1988 failure observations, and finally to 1987 screen values and 1989 failure observations. Table 6 presents the estimation results.

The econometric model would also have done a good job identifying failures in 1987, 1988, and 1989. For all three years, we could reject the hypothesis that the model had no explanatory power. Moreover, six individual coefficients differed statistically from zero with the hypothesized signs across all three equations. Specifically, low capital ratios (EQUITY), low liquid-asset ratios (LIQUID), high nonperforming-loan ratios (BAD-LOANS), high other-real-estate-owned ratios (OREO), high interest-accrued-but-not-collected ratios (UNCOLLECTED), and high large-time-deposit ratios (LARGE-TIME) correlated strongly with future failures. Overall, the econometric model implies that capital protection, asset quality, and liquidity positions are the most important determinants of failure risk.

Next, we used the econometric model to identify failure candidates that were not already on supervisors' watch lists. The evidence from 1989, 1990, and 1991 (which appears in Table 7) shows that the econometric model also would have improved significantly over random examination. Specifically, if the sample banks had been examined from the highest to the lowest estimated probability of failure (based on year-end 1987 data), 55 banks would have had to be examined to catch three of the six that would fail in 1989. To flag five of the 10 banks that would fail in 1990, 51 examinations would have been necessary. To identify five of the nine failures in 1991, 155 banks would have had to be examined. At first glance these numbers might seem high, but 55 banks represented only 1.1 percent of all CAMEL-2 banks in 1988; 51 represented only 0.8 percent of CAMEL-2 banks in 1989; and 155 represented a mere 2.4 percent of all CAMEL-2 banks in 1990. In short, the econometric model improves significantly on the random examination of CAMEL-2 banks.

Table 5

Do Individual Supervisory Screens Improve Over Random Examination of CAMEL-2 Banks?

This table demonstrates that individual supervisory screens improve over random examination of CAMEL-2 banks. To catch one-half of the following year's failures using a random examination strategy, supervisors would have to order, on average, visits to one-half of the CAMEL-2 banks. Only the consumer loan screen and the size control variable had average performance ratios above 50 percent.

Note, however, the considerable variance in the performance of individual supervisory screens. The performance ranking of individual screens changed significantly from year to year. Shading highlights screens that placed among the top five predictors in all three years. Only two screens placed consistently among the top five predictors.

Single-variable screen	For each year, the first column shows the percentage of CAMEL-2 banks that must be examined to include one-half of the banks that failed in the following calendar year. The second column indicates the rank of each screen from best (1) to worst (14).					
	Banks that Failed in:					
	1989		**1990**		**1991**	
	Percent based on 1987:4 data	Rank of screen	Percent based on 1988:4 data	Rank of screen	Percent based on 1989:4 data	Rank of screen
EQUITY	4.6	3	2.3	2	4.0	2
BAD-LOANS	16.9	6	7.3	4	2.2*	1
OREO	8.6	5	21.6	8	17.6	6
CONSUMER	26.9	10	37.0	11	86.1	14
INSIDER	44.8	12	9.3	6	37.1	11
OVERHEAD	22.6	7	5.6	3	56.7	12
OCCUPANCY	69.6	14	8.7	5	14.2	5
ROA	5.3	4	0.9*	1	4.7	3
UNCOLLECTED	25.5	8	37.2	12	31.1	10
LIQUID	25.9	9	15.9	7	5.0	4
LARGE-TIME	1.7*	1	23.7	10	29.1	8
CORE	3.9	2	46.6	14	30.7	9
SIZE	42.0	11	41.7	13	70.4	13
BHCRATIO	55.9	13	21.9	9	20.0	7

*Lowest among the screens.

EQUITY	Equity as a percentage of total assets.		UNCOLLECTED	Interest accrued as revenue but not collected as a percentage of total loans.
BAD-LOANS	Nonperforming loans as a percentage of total loans.		LIQUID	Liquid assets (sum of cash, securities, federal funds sold, and reverse repurchase agreements) as a percentage of total assets.
OREO	Other real estate owned (real estate other than bank premises) as a percentage of total loans.		LARGE-TIME	Large denomination time deposit liabilities as a percentage of total assets.
CONSUMER	Consumer loans as a percentage of total assets.			
INSIDER	The value of loans to insiders (officers and directors of the bank) as a percentage of total assets.		CORE	Core deposits (transactions, savings and small time deposits) as a percentage of total assets.
OVERHEAD	Noninterest expense as a percentage of total revenue.		SIZE	Natural logarithm of total assets, in thousands of dollars.
OCCUPANCY	Occupancy expense as a percentage of average assets.		BCHRATIO	The ratio of each bank's total assets to the total assets of its holding company. Banks without holding companies have BHCRATIO ≡ 1.
ROA	Net income as a percentage of total assets.			

Table 6

How Well Does the Econometric Model Fit the Bank Failure Data?

This table presents the estimated regression coefficients for the failure prediction logit. The model predicts in-sample failures ("1" represents failure; "0" denotes survivor) for calendar year *t* with year *t-2* call report data. Standard errors appear in parentheses below each coefficient. Three asterisks denote significance at the 1 percent level; two asterisks denote significance at the 5-percent level. Shading highlights coefficients that were significant with the correct sign in all three years. Overall, the evidence in this table suggests that the econometric model predicted in-sample failures well.

Independent Variables	Banks that Failed or Survived in:		
	1987	**1988**	**1989**
Intercept	-0.994 (2.801)	-2.588 (2.525)	-6.479 (3.499)
EQUITY	-0.303*** (0.055)	-0.314*** (0.056)	-0.285*** (0.051)
BAD-LOANS	0.107*** (0.018)	0.099*** (0.020)	0.095*** (0.023)
OREO	0.097*** (0.031)	0.047** (0.024)	0.122*** (0.019)
CONSUMER	0.007 (0.012)	0.002 (0.012)	-0.018 (0.012)
INSIDER	0.041 (0.023)	0.084 (0.048)	0.102 (0.054)
OVERHEAD	-0.014 (0.014)	-0.012 (0.013)	0.001 (0.002)
OCCUPANCY	0.710** (0.314)	0.450 (0.374)	-0.069 (0.308)
ROA	-0.061 (0.052)	0.007 (0.065)	0.007 (0.050)
UNCOLLECTED	0.935*** (0.132)	0.608*** (0.160)	0.828*** (0.215)
LIQUID	-0.041*** (0.010)	-0.019** (0.008)	-0.033*** (0.009)
LARGE-TIME	0.072*** (0.021)	0.074*** (0.016)	0.115*** (0.026)
CORE	0.003 (0.022)	0.007 (0.018)	0.034 (0.025)
SIZE	-0.356*** (0.111)	-0.120 (0.101)	-0.011 (0.116)
BHCRATIO	1.075*** (0.340)	-0.119 (0.236)	-0.348 (0.260)
Number of Observations	12,645	12,345	11,987
Pseudo-R^2	0.375	0.275	0.403
-2 log likelihood testing whether all coefficients (except the intercept) = 0	633.108***	453.035***	645.996***

EQUITY	Equity as a percentage of total assets.	UNCOLLECTED	Interest accrued as revenue but not collected as a percentage of total loans.
BAD-LOANS	Nonperforming loans as a percentage of total loans.	LIQUID	Liquid assets (sum of cash, securities, federal funds sold, and reverse repurchase agreements) as a percentage of total assets.
OREO	Other real estate owned (real estate other than bank premises) as a percentage of total loans.		
CONSUMER	Consumer loans as a percentage of total assets.	LARGE-TIME	Large denomination time deposit liabilities as a percentage of total assets.
INSIDER	The value of loans to insiders (officers and directors of the bank) as a percentage of total assets.	CORE	Core deposits (transactions, savings and small time deposits) as a percentage of total assets.
OVERHEAD	Noninterest expense as a percentage of total revenue.	SIZE	Natural logarithm of total assets, in thousands of dollars.
OCCUPANCY	Occupancy expense as a percentage of average assets.	BCHRATIO	The ratio of each bank's total assets to the total assets of its holding company. Banks without holding companies have BHCRATIO = 1.
ROA	Net income as a percentage of total assets.		

Table 7

How Well Does the Econometric Model Identify CAMEL-2 Failure Candidates?

This table quantifies the supervisory value added by the econometric model. Specifically, it shows how many CAMEL-2 banks must be examined in each year, based on logit probability estimates using data from the previous year, to catch each potential failure. For example, in 1988, supervisors would have had to examine 18 (or 0.4 percent) of the 2-rated banks to catch one of the 1989 failures. Catching one-half of the 1989 failures would have required examining 55 (or 1.1 percent) of the 2-rated banks. To catch all six failures, supervisors would have had to examine 650 (or 12.9 percent) of the 2-rated banks. Shading highlights the number of banks that must be examined to catch one-half of the failures in each year. Overall, the evidence suggests that the econometric model improved significantly on random examinations of CAMEL-2 banks.

Among the CAMEL-2 rated banks, rank based on probability of failure:		Estimated probability of failure	Percentage of CAMEL-2 rated banks that must be examined to include this failed bank
Among those that failed	Among all CAMEL-2 rated banks		
Among banks rated CAMEL 2 as of March 1988, six that failed during 1989:			
1	18	5.2%	0.4%
2	20	4.9	0.4
3	55	2.9	1.1
4	82	2.2	1.6
5	547	0.6	10.9
6	650	0.5	12.9
Among banks rated CAMEL 2 as of March 1989, 10 that failed during 1990:			
1	4	33.8	0.1
2	8	12.5	0.1
3	34	5.1	0.6
4	43	4.8	0.7
5	51	4.4	0.8
6	58	4.1	0.9
7	206	2.1	3.4
8	544	1.2	8.9
9	1,324	0.7	21.6
10	3,488	0.3	56.9
Among banks rated CAMEL 2 as of March 1990, nine that failed during 1991:			
1	34	4.7	0.5
2	72	3.4	1.1
3	101	2.9	1.6
4	141	2.5	2.2
5	155	2.3	2.4
6	212	2.0	3.3
7	523	1.1	8.1
8	1,913	0.4	29.8
9	5,774	0.0	89.9

At first glance, the resource-savings benchmark—the number of CAMEL-2 banks that must be examined to catch one-half of the following year's failures—appears to suggest that the screens and the model would have been comparable tools for allocating on-site examination resources. The comparison appears in Table 8, which combines data from Tables 5 and 7. In each year, the performance of the dominant screen is relatively close to the performance of the econometric model. For example, using the econometric model as a guide, supervisors would have had to examine 1.1 percent of all CAMEL-2 banks (as of March 1988) to catch one-half of the 1989 failures. If supervisors had used the dominant screen instead—the large-time-deposit ratio—they would have had to examine 1.7 percent of the CAMEL-2 banks. For 1990, the econometric model would have identified one-half of the failures after 0.8 percent of 2-rated banks had been examined; the comparable figure for the dominant screen (return on assets) was 0.9 percent. Finally, for 1991 failures, the dominant screen outperformed the econometric model. The nonperforming-loan screen identified one-half of the failures after examining 2.2 percent of the CAMEL-2 banks; the figure for the econometric model was 2.4 percent.

A closer look, however, reveals that the screens and the model would not have been equally effective surveillance tools. Although during each year the performance of the dominant screen is close to that of the econometric model, the dominant screens vary from year to year. Moreover, only two screens ranked among the top five in all three years, and in only one of those six cases (two screens, three years) did a screen beat the model. On average during the three-year period, the model significantly outperformed all of the individual screens. On average, supervisors could have caught one-half of the surprise failures by examining only 1.4 percent of the CAMEL-2 banks. The lowest average for the supervisory screens—the return-on-asset screen and the equity screen—was 3.6 percent. To put this evidence in perspective, suppose supervisors decided on the basis of 1989

screen performance to use the large-time-deposits-to-total-assets ratio as a guide for predicting 1990 failures. With such a guide, they would have had to examine 23.7 percent of the banks rated CAMEL-2 as of March 1989 to catch one-half of the failures. The comparable percentage using the econometric model is 0.8 percent. In summary, for single-variable screens to be as effective as the model, supervisors would have to know at the beginning of each year which screen would perform relatively well—an unrealistic information requirement.

It also is important to compare the performance of the screens and the model for a broader range of type-1 and type-2 errors. Put another way, the resource savings benchmark, while intuitively appealing, represents only one possible type-1/type-2 error trade-off. Type-1 errors, in this context, are missed failures; these errors impose unexpected costs on the deposit insurance fund and the real economy. Type-2 errors are missed survivors; these errors waste scarce examination resources and impose undue burdens on banks. Consider a concrete example of type-2 error using the individual capital screen. Suppose bank supervisors scheduled 1989 exams for all banks (CAMEL 1 through 5) using only fourth quarter 1987 values of the capital screen. Because the distributions of capital screen values for the failed and survivor banks overlap considerably (see Figure 2), this approach would lead to a large number of type-2 errors. For example, 359 survivor banks had weaker equity ratios than the average ratio for all the failed banks (see Table 3).

The evidence from a broader range of type-1/type-2 error trade-offs confirms the statistical dominance of the econometric model. An econometric model would dominate a set of screens as devices for identifying failures if it produced fewer type-2 errors (missed survivors) for any desired level of type-1 errors (missed failures). In pictures, meeting this condition implies that a curve tracing the trade-off between the two types of errors for the econometric model lies completely below

Table 8

How Do the Individual Supervisory Screens and the Econometric Model Compare as Tools for Allocating On-Site Examination Resources?

This table illustrates the superior performance of the econometric model as a tool for allocating on-site examination resources. It combines data from Tables 5 and 7. The columns show the percentage of banks that must be examined, using either the econometric model or a specific supervisory screen as a guide, to catch one-half of the banks that will fail that year. In each year, the dominant screen comes close to the model's performance, but the dominant screen varies year to year. Moreover, the three-year average for the model is well below the averages for the single variable screens.

Method of ranking banks by probability of failure.	Among banks rated CAMEL 2, the percentage that must be examined to include one-half of the banks that failed in the following calendar year.			
	Banks that failed in:			
	1989	1990	1991	Mean Percentage
Model	1.1%	0.8%	2.4%	1.4%
Screens				
EQUITY	4.6	2.3	4.0	3.6
BAD-LOANS	16.9	7.3	2.2*	8.8
OREO	8.6	21.6	17.6	15.9
CONSUMER	26.9	37.0	86.1	50.0
INSIDER	44.8	9.3	37.1	30.4
OVERHEAD	22.6	5.6	56.7	28.3
OCCUPANCY	69.6	8.7	14.2	30.8
ROA	5.3	0.9*	4.7	3.6
UNCOLLECTED	25.5	37.2	31.1	31.3
LIQUID	25.9	15.9	5.0	15.6
LARGE-TIME	1.7*	23.7	29.1	18.2
CORE	3.9	46.6	30.7	27.1
SIZE	42.0	41.7	70.4	51.4
BHCRATIO	55.9	21.9	20.0	32.6

*Lowest among the screens for that year.

EQUITY	Equity as a percentage of total assets.	UNCOLLECTED	Interest accrued as revenue but not collected as a percentage of total loans.
BAD-LOANS	Nonperforming loans as a percentage of total loans.	LIQUID	Liquid assets (sum of cash, securities, federal funds sold, and reverse repurchase agreements) as a percentage of total assets.
OREO	Other real estate owned (real estate other than bank premises) as a percentage of total loans.		
CONSUMER	Consumer loans as a percentage of total assets.	LARGE-TIME	Large denomination time deposit liabilities as a percentage of total assets.
INSIDER	The value of loans to insiders (officers and directors of the bank) as a percentage of total assets.	CORE	Core deposits (transactions, savings and small time deposits) as a percentage of total assets.
OVERHEAD	Noninterest expense as a percentage of total revenue.	SIZE	Natural logarithm of total assets, in thousands of dollars.
OCCUPANCY	Occupancy expense as a percentage of average assets.	BHCRATIO	The ratio of each bank's total assets to the total assets of its holding company. Banks without holding companies have BHCRATIO = 1.
ROA	Net income as a percentage of total assets.		

the trade-off curves for every single-variable screen.[5] Figure 3 presents the 1990 failure trade-off curve for the econometric model and the four best single variable screens, using the sample of CAMEL-2 banks. With only two exceptions, the trade-off curve for the econometric model does indeed lie below the curves for the individual screens. For small ranges of values, trade-off curves for the return-on-assets and the capital screens dip below the curve for the econometric model. Similar curves for 1989 and 1991 failures reveal similar patterns— the trade-off curve for the econometric lies below the curves for the individual screens with only a few exceptions. In those cases, one or two screens outperform the model for a small range of type-1/type-2 error trade-offs, but no one screen consistently outperforms the model. In summary, only by correctly guessing which screen will dominate at the beginning of the year and by preselecting a desired type-1 error rate from a small range of values can a supervisor beat the econometric model with a single-variable individual screen. These conditions are clearly difficult to meet.

GAUGING SUPERVISORY SCREENS AND ECONOMETRIC MODELS AS PREDICTORS OF CAMEL DOWNGRADES

Because failures have fallen off sharply since the early 1990s, supervisors have become interested in developing tools for flagging safe-and-sound banks that will develop problems. For this reason, we estimated an econometric model designed to capture the likelihood that a bank's CAMEL rating will be downgraded from CAMEL 1 or 2 to CAMEL 3, 4, or 5. Because such downgrades remained relatively common through 1997, we have a large enough sample to conduct a meaningful comparison of the resource savings obtained with the screens and the econometric model. (Figure 4 and Table 9 provide data on the frequency of these downgrades.)

To estimate a downgrade model, we changed the definition of a safety-and-

---

Figure 2

Hypothetical Distributions of Equity Ratios

This figure demonstrates type-2 error (the problem of missed survivors) using supervisory screens. When the distributions of the screen ratios for failing and surviving banks overlap considerably, supervisors who rely only on screens to schedule exams will devote a large quantity of on-site resources to banks unlikely to fail. Suppose that the figures below are capital screen (equity-to-total-asset ratio) distributions. In the top panel, the distribution for failures lies completely to the left of the distribution for survivors. If the actual distributions looked like this, supervisors could allocate on-site resources efficiently by examining banks with the lowest capital ratios. Unfortunately, the actual distributions are more like those in the bottom panel. For example, in late 1987, 359 of the 11,838 banks that would survive through 1989 had equity-to-asset ratios below the mean for the 149 banks that would fail that year. If supervisors flagged banks with the lowest equity-to-asset ratios in late 1987, their watch list would have included many more survivor banks than failed banks.

Nonoverlapping distributions


Overlapping distributions

---
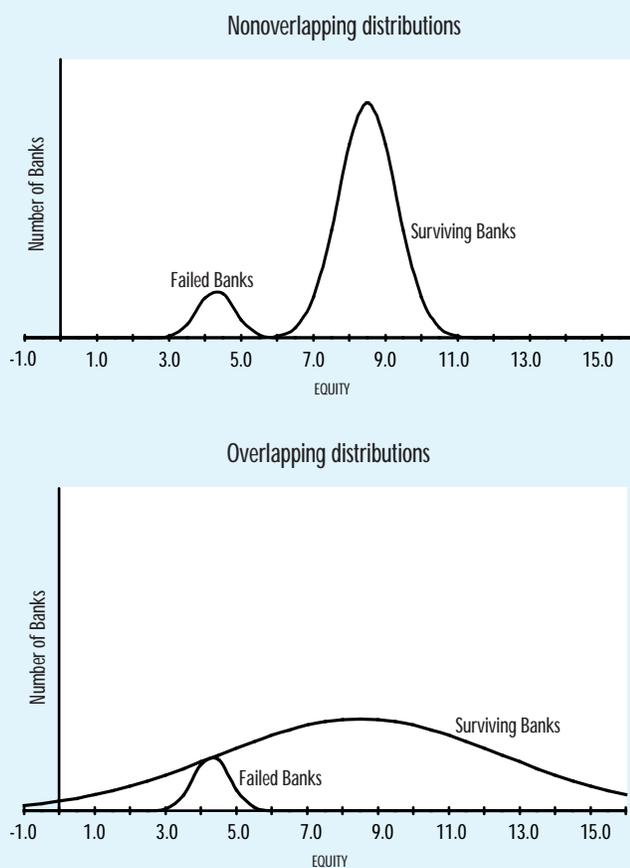
soundness problem and the sample selection criteria. Now, in the econometric model, we assigned a "1" to banks that suffered a downgrade from safe-and-sound status (CAMEL 1 or 2) to problem status (CAMEL 3, 4, or 5) and a "0" to all other

[5] Our graphical analysis of error trade-offs follows the approach used by Cole, Cornyn, and Gunther (1995).

Figure 3

What is the Trade-Off Between False Negatives and False Positives in the Failure-Prediction Model Compared to the Individual Screens?

1990 Failure Predictions Using Year-End 1988 Data (CAMEL-2 Banks only)
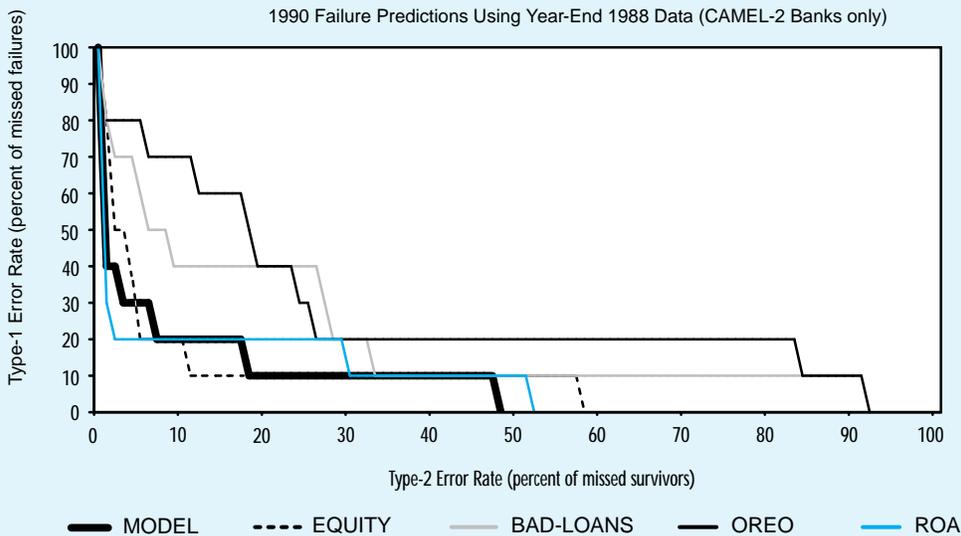
This figure shows the trade-off between the type-1 error rate (missed failures) and the type-2 error rate (missed survivors). The type-1 error rate is the percentage of banks rated CAMEL 2 that subsequently failed but were not identified by the model (or screen). The type-2 error rate is the percentage of banks rated CAMEL 2 that did not subsequently fail but were misidentified by the model (or screen) as a failure risk. This graph shows that for any level of type-1 error rate tolerated by supervisors, the econometric model (in bold) leads to fewer type-2 errors than most individual screens. Moreover, even in years when individual screens dominate the logit model over some ranges of the type-1 versus -2 trade-off, the dominant screens are not consistently the same. (For clarity, only the four best screens are shown.)

Figure 4

Number of Commercial Bank Downgrades by Year 1989-97

This figure shows that downgrades to problem status (CAMEL 3,4, or 5) are still relatively common, although the absolute number has declined since the early 1990s.

Table 9

How Many CAMEL-1 and CAMEL-2 Banks Suffered Downgrades to CAMEL 3, 4, or 5 from 1991 to 1997?

This table shows the number of our sample banks that were downgraded to problem status in each year. We excluded banks receiving downgrades to problem status the same year as the CAMEL 1 or 2 observation from the sample to avoid biasing comparisons against supervisory screens. Note: As overall banking performance improved in the 1990s, the percentage of banks suffering downgrades fell, but downgrades were still much more common than failures.

Date of Rating (Year of Downgrade)	CAMEL Rating	Number of Banks	Number of Banks Downgraded	Percentage Downgraded
March 1990 (1991)	1	2,057	79	3.84%
	2	5,036	987	19.60
March 1991 (1992)	1	1,956	51	2.61
	2	4,985	670	13.44
March 1992 (1993)	1	1,972	17	0.86
	2	5,212	292	5.60
March 1993 (1994)	1	2,041	14	0.69
	2	5,030	185	3.68
March 1994 (1995)	1	2,359	13	0.55
	2	4,446	127	2.86
March 1995 (1996)	1	2,583	13	0.50
	2	3,940	135	3.43
March 1996 (1997)	1	1,931	9	0.47
	2	2,420	103	4.26

banks. All of the sample banks were examined during the year of the downgrade. We excluded banks receiving downgrades the same year as the CAMEL 1 or 2 obser-vation. Without this exclusion, the predictive power of both the screens and the model would be seriously weakened. A simple example illustrates the problem. Suppose we are selecting sample banks for 1990. We begin with all CAMEL-1 and CAMEL-2 banks as of March 1990. If we did not exclude banks receiving downgrades during the remainder of 1990, the predictive power of 1989 screens for 1991 downgrades would be weakened because banks reclassified as problems in 1990 would not be in the set of 1991 downgrades. Apart from the change in the dependent variable and the sample selection criteria, the empirical tests were identical to those conducted on failures. Our dataset, however, now includes CAMEL downgrades from 1991

through 1997 and the corresponding lagged call report ratios.

The supervisory screens and the econometric model would each have done a good job predicting CAMEL downgrades. (Due to space constraints, the tables con-taining means and coefficient values can be found on the Research Department website of the Federal Reserve Bank of St. Louis <www.stls. frb.org/publications>.) For seven of the 14 individual screens, the differences between the means for the downgraded and non-downgraded banks were statistically significant with the hypothesized sign across all seven years. At the same time, the hypothesis that the econometric model had no explanatory value could be soundly rejected for all seven years. More specifically, across the seven equations, coefficients on five of the 14 independent variables were consistently significant with the hypothe-sized sign. In each approach, several
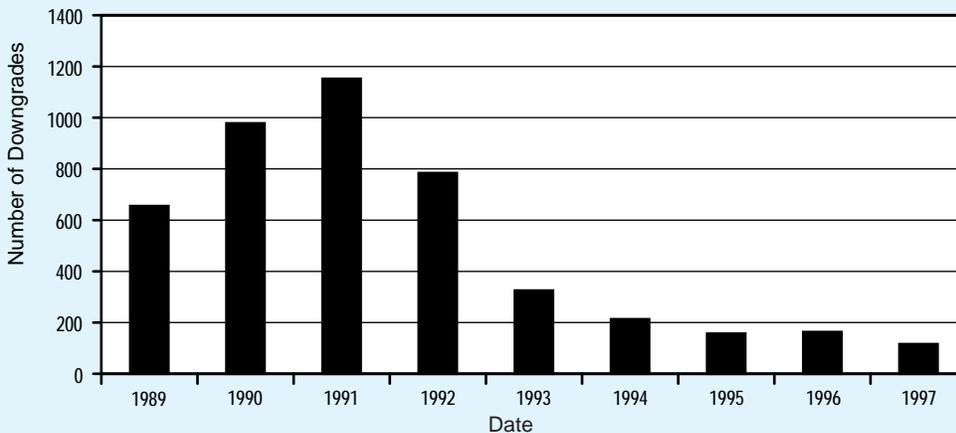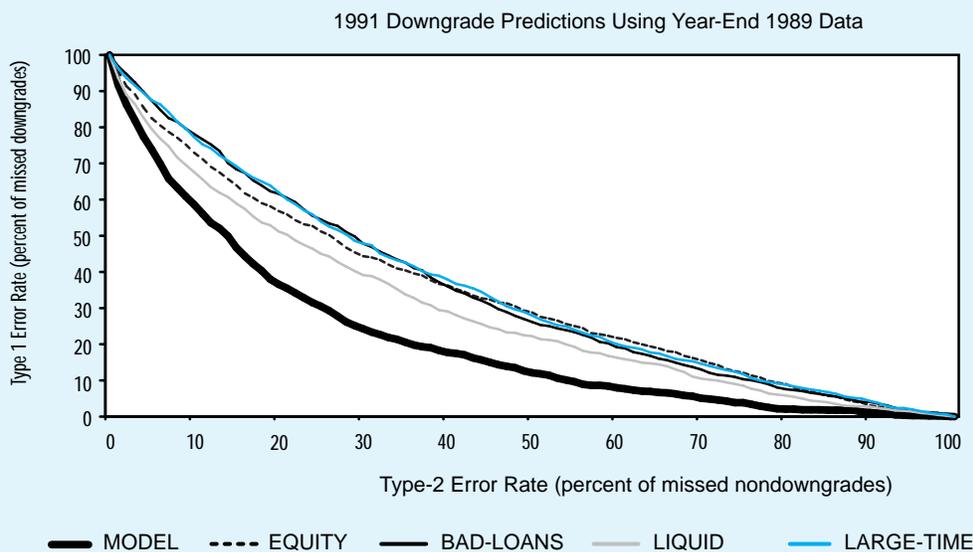
Figure 5

What is the Trade-Off Between False Negatives and False Positives in the Downgrade-Prediction Model Compared to the Individual Screens?

1991 Downgrade Predictions Using Year-End 1989 Data

This figure shows the trade-off between the type-1 error rate (missed downgrades) and the type-2 error rate (missed nondowngrades). The type 1-error rate is the percentage of banks rated CAMEL-1 or -2 that were subsequently downgraded by supervisors but were not identified by the model (or screen). The type-2 error misidentified by the model (or screen) as a downgrade risk. A desirable early-warning system minimizes the increase in type-2 errors for any given decrease in type-1 errors. This graph shows that for any level of type-1 error rate tolerated by supervisors, the econometric model (in bold) leads to fewer type-2 errors than any individual screen. For clarity, only the four best screens are shown.

additional variables also were factors in downgrades during most of the years. Looking at the evidence from the screens and the model, the credit and liquidity risk variables appear most closely correlated with future downgrades.

Next, we directly compared the performance of the screens and the model using the resource benchmark and the error trade-off benchmark. Table 10 contains the comparison for CAMEL-1 banks that will be downgraded, while Table 11 contains the comparison for CAMEL-2 banks. Figure 5 illustrates the type-1 versus type-2 error trade-offs for 1991 downgrades. (Due to space constraints, the error trade-off figures for 1992 through 1997 are available on the Research Department website of the Federal Reserve Bank of St. Louis.)

By the resource savings benchmark, the model would have outperformed the screens for both 1- and 2-rated institutions.

For CAMEL-1 banks, the econometric model posted lower exam percentages than any of the screens during four of the seven years. Moreover, as was the case for failure predictions, the rankings of the screens varied considerably from year to year. Finally, to catch one-half of the downgrades during the seven-year sample, supervisors would have had to examine only 16.9 percent of the CAMEL-1 banks using the econometric model. The lowest average for the supervisory screens—shared by the nonperforming-loan screen and the uncollected-interest-income screen—was 27.1 percent. For the CAMEL-2 banks, the results were even stronger: The econometric model outperformed the dominant screen every year. Again, the screen rankings varied considerably from year to year, and the dominant screen one year was not necessarily dominant the next. On average, supervisors could have caught one-half of

Table 10

How Does the Econometric Model Compare with the Single-Variable Screens as a Tool for Predicting CAMEL-1 Downgrades?

This table compares the econometric model and the individual screens as tools for predicting which CAMEL-1 banks will be downgraded to problem status. The columns show the percentage of banks that must be examined, using either the econometric model or a specific supervisory screen as a guide, to catch one-half of the downgrades the following year. In each year, the dominant screen comes close to the model's performance, but the dominant screen varies year to year. Moreover, on average, the model is clearly superior. The evidence suggests that the econometric model is the better tool for allocating on-site resources.

Method of Ranking Banks by Probability of Downgrade	Among banks rated CAMEL 1, the percentage of banks that must be examined to include one-half of the banks that were downgraded in the following calendar year.							
	Banks that were downgraded in:							
	1991	1992	1993	1994	1995	1996	1997	Mean Percentage
Model	12%	11%	9%	23%	31%	23%	9%	16.9%
EQUITY	29	46	51	31	49	34	24	37.7
BAD-LOANS	31	17*	25	22	23	35	37	27.1
OREO	50	42	39	46	74	52	67	52.9
CONSUMER	54	54	59	36	48	27	17	42.1
INSIDER	46	45	33	17	43	56	42	40.3
OVERHEAD	35	24	22	14*	45	28	64	33.1
OCCUPANCY	34	31	39	31	55	48	39	39.6
ROA	41	37	30	32	24	21	92	39.6
UNCOLLECTED	37	42	30	32	21*	17*	11*	27.1
LIQUID	19*	29	17*	64	59	37	15	34.3
LARGE-TIME	30	25	24	16	39	23	35	27.4
CORE	34	30	35	42	49	38	53	40.1
SIZE	73	52	40	21	32	26	54	42.6
BHCRATIO	56	44	19	27	36	36	58	39.4

*Lowest number among single-variable screens that year.

EQUITY	Equity as a percentage of total assets.	UNCOLLECTED	Interest accrued as revenue but not collected as a percentage of total loans.
BAD-LOANS	Nonperforming loans as a percentage of total loans.	LIQUID	Liquid assets (sum of cash, securities, federal funds sold, and reverse repurchase agreements) as a percentage of total assets.
OREO	Other real estate owned (real estate other than bank premises) as a percentage of total loans.		
CONSUMER	Consumer loans as a percentage of total assets.	LARGE-TIME	Large denomination time deposit liabilities as a percentage of total assets.
INSIDER	The value of loans to insiders (officers and directors of the bank) as a percentage of total assets.	CORE	Core deposits (transactions, savings and small time deposits) as a percentage of total assets.
OVERHEAD	Noninterest expense as a percentage of total revenue.	SIZE	Natural logarithm of total assets, in thousands of dollars.
OCCUPANCY	Occupancy expense as a percentage of average assets.	BCHRATIO	The ratio of each bank's total assets to the total assets of its holding company. Banks without holding companies have BHCRATIO = 1.
ROA	Net income as a percentage of total assets.		

Table 11

How Does the Econometric Model Compare with the Single-Variable Screens as a Tool for Predicting CAMEL-2 Downgrades?

This table compares the econometric model and the individual screens as tools for predicting which CAMEL-2 banks will be downgraded to problem status. The columns show the percentage of banks that must be examined, using either the econometric model or a specific supervisory screen as a guide, to catch one-half of the downgrades the following year. In each year, the dominant screen comes close to the model's performance, but the dominant screen varies year to year. Moreover, on average, the model is clearly superior. The evidence in this table suggests that the econometric model is the better tool for allocating on-site resources.

Method of Ranking Banks by Probability of Downgrade	Among banks rated CAMEL 2, the percentage of banks that must be examined to include one-half of the banks that were downgraded in the following calendar year.							
	Banks that were downgraded in:							
	1991	1992	1993	1994	1995	1996	1997	Mean Percentage
Model	24%	18%	13%	19%	21%	15%	16%	18.0%
EQUITY	35	39	45	48	51	52	42	44.6
BAD-LOANS	37	35	26	38	33	35	30	33.4
OREO	45	44	39	44	39	43	50	43.4
CONSUMER	50	47	47	53	45	45	36	46.1
INSIDER	51	46	42	42	44	47	45	45.3
OVERHEAD	42	38	24*	39	35	42	43	37.6
OCCUPANCY	38	34	27	35	34	39	40	35.3
ROA	40	39	26	32*	37	41	32	35.3
UNCOLLECTED	47	40	44	43	32	29*	34	38.4
LIQUID	29*	25*	25	34	38	35	28*	30.6
LARGE-TIME	34	34	32	41	31*	36	35	34.7
CORE	37	39	38	45	36	41	40	39.4
SIZE	62	58	51	39	37	33	36	45.1
BHCRATIO	49	42	36	32	28	33	40	37.1

*Lowest number among single-variable screens that year.

EQUITY	Equity as a percentage of total assets.		UNCOLLECTED	Interest accrued as revenue but not collected as a percentage of total loans.
BAD-LOANS	Nonperforming loans as a percentage of total loans.		LIQUID	Liquid assets (sum of cash, securities, federal funds sold, and reverse repurchase agreements) as a percentage of total assets.
OREO	Other real estate owned (real estate other than bank premises) as a percentage of total loans.			
			LARGE-TIME	Large denomination time deposit liabilities as a percentage of total assets.
CONSUMER	Consumer loans as a percentage of total assets.			
INSIDER	The value of loans to insiders (officers and directors of the bank) as a percentage of total assets.		CORE	Core deposits (transactions, savings and small time deposits) as a percentage of total assets.
OVERHEAD	Noninterest expense as a percentage of total revenue.		SIZE	Natural logarithm of total assets, in thousands of dollars.
OCCUPANCY	Occupancy expense as a percentage of average assets.		BCHRATIO	The ratio of each bank's total assets to the total assets of its holding company. Banks without holding companies have BHCRATIO = 1.
ROA	Net income as a percentage of total assets.			

the CAMEL-2 downgrades by examining only 18 percent of the CAMEL-2 banks. The lowest average for the supervisory screens—the liquid asset screen—was 30.6 percent.

Broadening the desired range of type-1 errors to other values besides 50 percent confirms the dominance of the econometric model. Figure 5 contains the 1991 error trade-off curves for the model and the individual supervisory screens, based on a pooled sample of CAMEL-1 and CAMEL-2 banks. For all ranges of type-1 errors, the econometric trade-off curve model lies below the curves for the individual supervisory screens. The curves for 1992 through 1997 reveal a similar pattern. In one year, the trade-off curves for the return-on-asset screen and the holding company control variable dipped below the econometric-model curve for a small range of values. Again, to beat the model with a single screen, supervisors would have had to guess correctly which screen would turn in a superior performance and the appropriate level of type-1 error. In summary, by the resource savings benchmark or the error trade-off benchmark, the econometric model clearly outperforms individual supervisory screens as a tool for predicting CAMEL downgrades.

RISK-SCOPING WITH ECONOMETRIC MODELS

To be useful in risk-focused supervision, an off-site surveillance tool must go beyond identifying institutions that are likely to develop safety-and-soundness problems and pinpoint the source of the developing problems. Armed with this information, supervisors then can determine the appropriate size and experience level of the examination team. Screens are attractive for risk-scoping because the specific financial ratios are designed to conform to the CAMELS framework. With some minor tweaking, however, supervisors also can use the output from econometric models to scope exams.

Table 12 demonstrates how the downgrade model can reveal the source of developing safety-and-soundness problems

for a randomly selected bank in our sample. (Currently, the Board of Governors provides similar information to each Reserve Bank to support SEER. This information is contained in the Risk Profile Analysis Report.) The table presents the actual values of the regression variables for a sample bank with a sizable downgrade probability (column 2), along with average values for all the sample banks (column 3). Overall, this bank has an 11.31 percent chance of suffering a downgrade to problem status during the next 12 to 24 months, roughly three times the average downgrade probability for the sample. In addition, the actual values for the regression variables at this bank are weaker than the sample average in every case except uncollected revenue and core deposits.

Asset quality and management competence appear to be the principal sources of weakness at this bank. We isolated these sources of weakness by calculating the downgrade probability that we would obtain for each independent variable if it were set equal to the peer average and all the other independent variables remained at their actual values. These numbers appear in column four of the table. Column five of Table 12 then shows the difference between this hypothetical probability and the overall probability of a downgrade. A large positive number in column five indicates that the screen value makes a relatively large contribution to the downgrade probability. For example, OREO is the largest single contributor to risk for this bank: The ratio is 4.10 compared with a 0.23 average figure for the sample. If that OREO ratio were set equal to the average ratio for the sample, the overall downgrade probability for the bank would fall 5.07 percentage points, from 11.31 percent to 6.24 percent. Viewed another way, the high OREO ratio at this bank accounts for nearly one-half of its overall downgrade probability. The nonperforming loan ratio and the overhead expense ratio also contribute substantially to the downgrade probability. Supervisors risk-scoping this exam would assign more examiners to loan review and discussions with management.

Table 12

What Does the Econometric Model Tell Us About the Factors Contributing to a Downgrade?

This table shows how the econometric model can be used to isolate the variables most responsible for a likely downgrade. Column one lists the explanatory variables in the model. The second column gives the value of each variable for a sample bank with an 11.31 percent downgrade probability. The third column shows the average value of each variable among all the sample banks. Column four shows what the predicted downgrade probability would be if the selected variable were set equal to the sample peer average and all the other variables were kept at their actual values. The final column shows the difference between this hypothetical probability and the actual downgrade probability (11.31 percent). A large positive number in column 5 indicates that the given variable makes a significant contribution to the bank's risk. For example, the largest single contributor to risk at this bank is the OREO ratio (4.10 compared with the peer average of 0.23). In contrast, favorable core deposit and uncollected interest income rates, relative to peer, improve the bank's standing by 0.34 and 0.13 percentage points.

Random Bank from the Downgrade Regression Sample
Downgrade Probability: 11.31 percent

Regression Variable (1)	Most recent value of bank's ratio (in %) (2)	Average sample value for variable (in %) (3)	Downgrade probability with variable set to sample average (4)	Difference from bank's actual downgrade probability (5)
EQUITY	8.55	9.94	10.64	0.67
BAD-LOANS	1.75	0.75	9.32	1.99
OREO	4.10	0.23	6.24	5.07
CONSUMER	10.93	9.00	11.11	0.20
INSIDER	0.32	1.27	11.25	0.06
OVERHEAD	49.73	35.49	8.05	3.26
OCCUPANCY	0.56	0.39	11.23	0.08
ROA	0.75	1.10	10.61	0.70
UNCOLLECTED	0.57	0.59	11.44	-0.13
LIQUID	28.62	38.55	8.60	2.71
LARGE-TIME	9.05	8.50	11.00	0.32
CORE	80.47	77.67	11.65	-0.34
SIZE	10.15	11.27	8.60	2.71
BHCRATIO	0.99	0.65	6.94	4.37

EQUITY	Equity as a percentage of total assets.	UNCOLLECTED	Interest accrued as revenue but not collected as a percentage of total loans.
BAD-LOANS	Nonperforming loans as a percentage of total loans.	LIQUID	Liquid assets (sum of cash, securities, federal funds sold, and reverse repurchase agreements) as a percentage of total assets.
OREO	Other real estate owned (real estate other than bank premises) as a percentage of total loans.	LARGE-TIME	Large denomination time deposit liabilities as a percentage of total assets.
CONSUMER	Consumer loans as a percentage of total assets.		
INSIDER	The value of loans to insiders (officers and directors of the bank) as a percentage of total assets.	CORE	Core deposits (transactions, savings and small time deposits) as a percentage of total assets.
OVERHEAD	Noninterest expense as a percentage of total revenue.	SIZE	Natural logarithm of total assets, in thousands of dollars.
OCCUPANCY	Occupancy expense as a percentage of average assets.	BCHRATIO	The ratio of each bank's total assets to the total assets of its holding company. Banks without holding companies have BHCRATIO = 1.
ROA	Net income as a percentage of total assets.		

Supervisors also can use information provided by the control variables in exam planning. For example, both of the control variables, SIZE and BHCRATIO, make significant contributions to the downgrade probability for this bank. Recall that all other things being equal, both large banks and small banks that are members of large holding companies are less likely to encounter safety-and-soundness problems. In the example, the large values for SIZE and BHCRATIO imply that the management and loan quality problems demand more examiner attention because this bank is not a large, well-diversified institution and cannot rely on a parent company as a source of strength.

SUPERVISORY SCREENS AND ECONOMETRIC MODELS AS COMPLEMENTS

Our statistical evidence does not, however, imply that the screens currently employed by supervisors add no value in off-site surveillance. First, as noted earlier, our screens and models are not the actual screens and models currently used by the surveillance community. Second, our tests are biased in favor of econometric models. Finally, our tests do not measure the potential value of screens in a rapidly changing banking environment.

Our comparisons contain several biases against screens. As noted, in practice, supervisory screens typically are weighted averages of financial ratios. Our representative screens, in contrast, are single-variable screens. In addition, supervisors modify their screens regularly based on feedback from field examiners. Our approach implicitly assumes that supervisors used the same single-variable screens throughout the entire sample. A better approach would rely on a time series of the actual multiple-variable screens used, but unfortunately, no such series exists. Finally, it is possible that successful use of the screens weakened their predictive power. Suppose supervisors ignored the output of econometric models, relying exclusively on screens to identify the banks that were most vulner-

able to failure or ratings downgrades. Suppose further that supervisors then intervened to prevent these failures or downgrades. From a statistical standpoint, the more successful the use of screens, the weaker their predictive power would be.

Our simple statistical horse races also fail to capture the value that supervisory screens can add in a dynamic banking environment. The agricultural bank problems of the 1980s demonstrate this value. Before the 1980s, the agricultural-loan-to-total-loan ratio would not have correlated positively with bank failures. That changed with the sharp declines in farm income and prices after 1981 (Belongia and Gilbert, 1990). By 1982, examination reports revealed that banks top-heavy with agricultural loans were significant failure risks. Failures did not rise sharply, however, until the second half of 1984, after declines in farm income and prices had absorbed the net worth of farmers and their banks (Kliesen and Gilbert, 1996). Because of the need to re-estimate coefficients and conduct new performance tests, new econometric models would not have been available to warn of agricultural bank vulnerability until 1985 or perhaps 1986. In short, supervisors could have developed screens for predicting agricultural bank failures long before econometric models would have signaled a rise in failure probabilities.

CONCLUSION

Off-site surveillance involves using accounting data to identify banks likely to develop safety-and-soundness problems. Early intervention, based on this information, can limit losses to the deposit insurance fund and the real economy. Supervisors rely heavily on two tools to flag developing problems: supervisory screens and econometric models. We used data from the 1980s and 1990s to compare, once again, the performance of these two approaches to off-site surveillance. As in earlier comparisons, the econometric models outperformed the supervisory screens. These results do not, however, suggest that screens should be dropped from the surveillance toolbox.

When abrupt changes in the causes of bank failures and CAMEL downgrades occur, supervisors can use their first-hand knowledge to modify screens long before models can be revised to reflect the new conditions. In short, the flexibility of supervisory screens makes them an important complement for econometric models in off-site surveillance.

REFERENCES

Barancik, Scott. "Few Banks and Thrifts Pay FDIC Premiums," *American Banker*, April 21, 1999, page 26.

Belongia, Michael T., and R. Alton Gilbert. "The Effects of Management Decisions on Agricultural Bank Failures," *American Journal of Agricultural Economics* (November 1990), pp. 901-10.

Bernanke, Ben S. "The Macroeconomics of the Great Depression: A Comparative Approach," *Journal of Money, Credit, and Banking* (February 1995), pp. 1-28.

_____. "Nonmonetary Effects of the Financial Crisis in the Propagation of the Great Depression," *American Economic Review* (June 1983), pp. 257-76.

_____ and Harold James. "The Gold Standard, Deflation, and Financial Crisis in the Great Depression: An International Comparison," in *Financial Markets and Financial Crises*, R. Glenn Hubbard, ed., University of Chicago Press, 1991, pp. 33-68.

Board of Governors of the Federal Reserve System. "Risk-Focused Safety and Soundness Examinations and Inspections," SR 96-14, May 24, 1996.

Cannella, Albert A. Jr., Douglas R. Fraser, and D. Scott Lee. "Firm Failure and Managerial Labor Markets: Evidence from Texas Banking," *Journal of Financial Economics* (June 1995), pp. 185-210.

Cole, Rebel A., and Jeffrey W. Gunther. "A Comparison of On- and Off-Site Monitoring Systems," *Journal of Financial Services Research* (April 1998), pp. 103-17.

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

Flannery, Mark J. "Deposit Insurance Creates a Need for Bank Regulation," Federal Reserve Bank of Philadelphia *Business Review* (January/February 1982), pp. 17-27.

_____ and Joel F. Houston. "The Value of a Government Monitor for U.S. Banking Firms," *Journal of Money, Credit, and Banking* (February 1999), pp. 14-34.

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

Gilbert, R. Alton. "Do Bank Holding Companies Act as Sources of Strength for Their Bank Subsidiaries," this *Review* (January/February 1991), pp. 3-18.

_____ and Gerald P. Dwyer, Jr. "Bank Runs and Private Remedies," this *Review* (May/June 1989), pp. 43-61.

_____ and Levis A. Kochin. "Local Economic Effects of Bank Failures," *Journal of Financial Services Research* (December 1989), pp. 333-45.

_____ and Mark D. Vaughan. "Does the Publication of Supervisory Enforcement Actions Add to Market Discipline?" *Research in Financial Services Public and Private Policy* (1998), pp. 259-80.

Hall, John R., Thomas B. King, Andrew P. Meyer, and Mark D. Vaughan. "Do Certificate of Deposit Holders and Supervisors View Bank Risk Similarly? A Comparison of the Factors Affecting CD Yields and CAMEL Composites," Supervisory Policy Analysis Working Paper, October 1999.

Kliesen, Kevin L., and R. Alton Gilbert. "Are Some Agricultural Banks Too Agricultural?" this *Review* (January/February 1996), pp. 23-35.

Putnam, Barron H. "Early Warning Systems and Financial Analysis in Bank Monitoring: Concepts of Financial Monitoring," Federal Reserve Bank of Atlanta *Economic Review* (November 1983), pp. 6-13.

Reidhill, Jack, and John O'Keefe. "Off-Site Surveillance Systems," in *History of the Eighties: Lessons for the Future, Volume I.* Washington: Federal Deposit Insurance Corporation, 1997.

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

Supplemental Table 1: How Well Do the Individual Supervisory Screens Predict CAMEL Downgrades?

This table presents evidence about the downgrade prediction record of individual supervisory screens. The far left and right columns for each year contain the mean values of the screens; standard deviations appear in parentheses below the means. An asterisk indicates a significant difference (at the 5 percent level) between the means for downgraded and non-downgraded banks. Shading highlights screens with significant predictive power in all seven years. The center column for each year (‡) shows the number of non-downgraded banks with screen values worse than those of the average downgraded bank; the larger this number, the worse the performance of the screen. Taken together, the evidence in this table shows that screens warn of potential downgrades but can also lead to many unnecessary exams.

	Data as of 1989:4 for:			Data as of 1990:4 for:		
	1,066 banks downgraded in 1991	‡	6,027 banks not downgraded in 1991	721 banks downgraded in 1992	‡	6,220 banks not downgraded in 1992
EQUITY	8.03* (2.37)	2,208	9.43 (3.38)	8.20* (2.40)	2,717	9.28 (3.10)
BAD-LOANS	2.20* (1.90)	1,210	1.40 (1.93)	2.19* (1.87)	1,270	1.38 (1.49)
OREO	0.97* (1.59)	1,206	0.65 (1.25)	1.04* (2.03)	1,241	0.66 (1.19)
CONSUMER	10.79 (7.57)	2,244	10.49 (7.69)	10.61 (7.27)	2,232	10.14 (7.71)
INSIDER	0.65* (1.13)	1,315	0.46 (0.84)	0.62* (0.93)	1,434	0.47 (0.85)
OVERHEAD	32.08* (8.54)	1,961	29.95 (7.69)	34.08* (9.09)	1,566	30.31 (7.60)
OCCUPANCY	0.53* (0.28)	1,441	0.43 (0.28)	0.55* (0.29)	1,340	0.42 (0.27)
ROA	0.92* (0.62)	1,801	1.14 (0.52)	0.81* (0.59)	1,557	1.07 (0.52)
UNCOLLECTED	0.82* (0.44)	1,864	0.74 (0.42)	0.86* (0.46)	1,663	0.71 (0.41)
LIQUID	36.11* (13.36)	1,591	46.71 (14.80)	34.46* (12.67)	1,486	46.06 (14.93)
LARGE-TIME	12.84* (8.08)	1,366	9.09 (6.69)	12.52* (7.72)	1,452	9.00 (6.53)
CORE	75.00* (10.92)	1,429	78.79 (8.80)	76.35* (9.90)	1,756	78.95 (8.74)
SIZE	11.37* (1.57)	4,415	10.85 (1.18)	11.05* (1.28)	3,754	10.91 (1.18)
BHCRATIO	0.78 (0.36)	4,279	0.77 (0.37)	0.82* (0.34)	4,343	0.77 (0.37)

Supplemental Table 1, Continued: How Well Do the Individual Supervisory Screens Predict CAMEL Downgrades?

This table presents evidence about the downgrade prediction record of individual supervisory screens. The far left and right columns for each year contain the mean values of the screens; standard deviations appear in parentheses below the means. An asterisk indicates a significant difference (at the 5 percent level) between the means for downgraded and non-downgraded banks. Shading highlights screens with significant predictive power in all seven years. The center column for each year (‡) shows the number of non-downgraded banks with screen values worse than those of the average downgraded bank; the larger this number, the worse the performance of the screen. Taken together, the evidence in this table shows that screens warn of potential downgrades but can also lead to many unnecessary exams.

	Data as of 1991:4 for:			Data as of 1992:4 for:		
	309 banks downgraded in 1993	‡	6,875 banks not downgraded in 1993	199 banks downgraded in 1994	‡	6,872 banks not downgraded in 1994
EQUITY	8.59* (2.30)	3,491	9.34 (3.27)	8.89* (2.71)	3,659	9.47 (3.16)
BAD-LOANS	2.53* (2.11)	1,048	1.37 (1.84)	1.81* (1.73)	1,491	1.20 (1.57)
OREO	1.12* (1.74)	1,290	0.67 (1.23)	1.02* (1.83)	1,375	0.63 (1.10)
CONSUMER	9.17 (5.93)	2,759	9.49 (7.18)	8.92 (6.81)	2,621	9.04 (7.22)
INSIDER	0.69* (1.09)	1,405	0.46 (0.79)	0.60* (0.76)	1,435	0.41 (0.77)
OVERHEAD	38.34* (10.14)	1,340	32.79 (8.60)	42.53* (11.70)	1,644	37.56 (9.58)
OCCUPANCY	0.57* (0.31)	1,245	0.42 (0.31)	0.52* (0.27)	1,679	0.42 (0.25)
ROA	0.71* (0.74)	1,098	1.08 (0.65)	0.97* (0.64)	1,631	1.25 (0.55)
UNCOLLECTED	0.79* (0.52)	1,957	0.67 (0.43)	0.71* (0.54)	1,729	0.57 (0.38)
LIQUID	36.18* (13.36)	1,806	46.65 (14.91)	41.68* (16.56)	2,810	46.76 (15.01)
LARGE-TIME	11.20* (7.81)	1,575	8.02 (5.91)	9.22* (7.35)	1,765	7.02 (5.34)
CORE	77.93* (8.91)	2,034	79.94 (8.28)	79.94 (8.43)	2,484	80.66 (8.42)
SIZE	10.90 (1.17)	3,568	10.96 (1.17)	10.67* (1.08)	2,799	11.06 (1.23)
BHCRATIO	0.91* (0.24)	4,734	0.77 (0.37)	0.95* (0.17)	4,614	0.77 (0.37)

Supplemental Table 1, Continued: How Well Do the Individual Supervisory Screens Predict CAMEL Downgrades?

This table presents evidence about the downgrade prediction record of individual supervisory screens. The far left and right columns for each year contain the mean values of the screens; standard deviations appear in parentheses below the means. An asterisk indicates a significant difference (at the 5 percent level) between the means for downgraded and non-downgraded banks. Shading highlights screens with significant predictive power in all seven years. The center column for each year (‡) shows the number of non-downgraded banks with screen values worse than those of the average downgraded bank; the larger this number, the worse the performance of the screen. Taken together, the evidence in this table shows that screens warn of potential downgrades but can also lead to many unnecessary exams.

	Data as of 1993:4 for:			Data as of 1994:4 for:			Data as of 1995:4 for:		
	140 banks down-graded in 1995	‡	6,665 banks not down-graded in 1995	148 banks down-graded in 1996	‡	6,375 banks not down-graded in 1996	112 banks down-graded in 1997	‡	4,239 banks not down-graded in 1997
EQUITY	9.41 (3.73)	3,820	9.74 (3.05)	9.40 (3.46)	3,609	9.72 (3.23)	9.48* (4.04)	2,149	10.31 (3.83)
BAD-LOANS	1.79* (2.12)	1,199	1.05 (1.27)	1.65* (2.11)	1,026	0.91 (1.12)	1.86* (1.84)	575	0.91 (1.12)
OREO	0.86* (1.89)	1,060	0.45 (0.91)	0.66* (1.80)	964	0.34 (0.89)	0.44 (1.51)	716	0.27 (0.64)
CONSUMER	9.65 (6.68)	2,186	9.03 (7.49)	11.63* (10.13)	1622	9.33 (7.81)	12.97* (9.86)	865	9.09 (6.91)
INSIDER	1.42 (1.79)	2,027	1.22 (1.50)	1.30 (1.30)	2,254	1.27 (1.59)	1.34 (1.30)	1,431	1.27 (1.43)
OVERHEAD	47.17* (13.41)	1,584	41.75 (10.15)	46.09* (11.27)	1,850	42.04 (9.33)	41.29* (14.46)	1,234	38.17 (27.49)
OCCUPANCY	0.52* (0.25)	1,755	0.43 (0.24)	0.54* (0.30)	1,584	0.44 (0.24)	0.52* (0.28)	1,183	0.45 (0.27)
ROA	1.06* (0.74)	2,116	1.29 (0.84)	0.98* (0.56)	1,850	1.21 (1.18)	0.99* (0.57)	1,110	1.25 (0.87)
UNCOLLECTED	0.74* (0.55)	1,434	0.54 (0.36)	0.94* (0.71)	998	0.61 (0.39)	1.04* (0.64)	732	0.69 (0.46)
LIQUID	39.03* (14.16)	2,667	44.65 (15.06)	34.45* (12.55)	2,293	41.50 (14.65)	32.85* (11.74)	1,232	41.63 (13.98)
LARGE-TIME	9.83* (7.46)	1,494	6.90 (5.15)	10.07* (5.62)	1,617	7.60 (5.56)	11.69* (6.29)	1,049	8.86 (5.66)
CORE	78.09* (8.85)	1,986	79.96 (8.78)	77.73 (7.41)	2,224	78.63 (9.08)	76.21 (8.34)	1,509	77.31 (8.65)
SIZE	10.70* (1.05)	2,678	11.12 (1.28)	10.52* (1.02)	2,014	11.16 (1.27)	10.67* (0.90)	1,557	11.15 (1.23)
BHCRATIO	0.92* (0.23)	4,588	0.78 (0.36)	0.92* (0.22)	4,439	0.78 (0.36)	0.90* (0.25)	3,049	0.80 (0.34)

Supplemental Table 2: How Well Does the Logit Model Fit the CAMEL Downgrade Data?

This table presents the estimated regression coefficients for the downgrade prediction logit. The model predicts in-sample downgrades ("1" represents downgrade from safe-and-sound to problem status) for calendar year *t* with year *t*-2 call report data. Standard errors appear in parentheses below each coefficient. Three asterisks denote significance at the 1 percent level; two asterisks denote significance at the 5 percent level. Shading highlights coefficients that were significant with the correct sign in all seven years. Overall, the logit model does a good job predicting in-sample downgrades.

	Banks that were examined in:	
Independent Variables	**1989**	**1990**
Intercept	0.037	-1.192
	(1.444)	(1.161)
EQUITY	-0.135***	-0.123***
	(0.030)	(0.024)
BAD-LOANS	0.198***	0.268***
	(0.024)	(0.023)
OREO	0.171***	0.144***
	(0.029)	(0.025)
CONSUMER	-0.006	-0.018***
	(0.006)	(0.005)
INSIDER	0.181***	0.012
	(0.044)	(0.032)
OVERHEAD	-0.005	-0.003
	(0.009)	(0.007)
OCCUPANCY	0.616**	0.245
	(0.251)	(0.223)
ROA	-0.472***	-0.532***
	(0.087)	(0.081)
UNCOLLECTED	0.678***	0.297***
	(0.134)	(0.112)
LIQUID	-0.044***	-0.052***
	(0.004)	(0.004)
LARGE-TIME	0.058***	0.049***
	(0.010)	(0.009)
CORE	0.001	0.001
	(0.009)	(0.007)
SIZE	-0.077	0.138***
	(0.052)	(0.041)
BHCRATIO	0.293**	0.435***
	(0.132)	(0.111)
Number of Observations	5,495	6,672
Pseudo-R^2	0.199	0.185
-2 log likelihood testing whether all coefficients (except the intercept) = 0	786.646***	1004.216***

EQUITY	Equity as a percentage of total assets.	UNCOLLECTED	Interest accrued as revenue but not collected as a percentage of total loans.
BAD-LOANS	Nonperforming loans as a percentage of total loans.		
OREO	Other real estate owned (real estate other than bank premises) as a percentage of total loans.	LIQUID	Liquid assets (sum of cash, securities, federal funds sold, and reverse repurchase agreements) as a percentage of total assets.
CONSUMER	Consumer loans as a percentage of total assets.		
INSIDER	The value of loans to insiders (officers and directors of the bank) as a percentage of total assets.	LARGE-TIME	Large denomination time deposit liabilities as a percentage of total assets.
		CORE	Core deposits (transactions, savings and small time deposits) as a percentage of total assets.
OVERHEAD	Noninterest expense as a percentage of total revenue.		
		SIZE	Natural logarithm of total assets, in thousands of dollars.
OCCUPANCY	Occupancy expense as a percentage of average assets.	BHCRATIO	The ratio of each bank's total assets to the total assets of its holding company. Banks without holding companies have BHCRATIO ≡ 1.
ROA	Net income as a percentage of total assets.		

Supplemental Table 2, Continued: How Well Does the Logit Model Fit the CAMEL Downgrade Data?

This table presents the estimated regression coefficients for the downgrade prediction logit. The model predicts in-sample downgrades ("1" represents downgrade from safe-and-sound to problem status) for calendar year *t* with year *t*-2 call report data. Standard errors appear in parentheses below each coefficient. Three asterisks denote significance at the 1 percent level; two asterisks denote significance at the 5 percent level. Shading highlights coefficients that were significant with the correct sign in all seven years. Overall, the logit model does a good job predicting in-sample downgrades.

	Banks that were examined in:	
Independent Variables	**1991**	**1992**
Intercept	-0.309	-0.728
	(1.095)	(1.340)
EQUITY	-0.117***	-0.086***
	(0.022)	(0.024)
BAD-LOANS	0.244***	0.254***
	(0.023)	(0.025)
OREO	0.130***	0.089***
	(0.026)	(0.031)
CONSUMER	-0.023***	-0.015***
	(0.005)	(0.006)
INSIDER	0.081**	0.038
	(0.035)	(0.043)
OVERHEAD	-0.004	0.026***
	(0.008)	(0.008)
OCCUPANCY	0.357*	0.053
	(0.213)	(0.241)
ROA	-0.599***	-0.569***
	(0.080)	(0.089)
UNCOLLECTED	0.225**	0.447***
	(0.096)	(0.108)
LIQUID	-0.058***	-0.067***
	(0.004)	(0.004)
LARGE-TIME	0.045***	0.057***
	(0.008)	(0.010)
CORE	0.000	0.007
	(0.007)	(0.009)
SIZE	0.109***	-0.072
	(0.039)	(0.048)
BHCRATIO	0.461***	0.694***
	(0.108)	(0.134)
Number of Observations	7,093	6,941
Pseudo-R^2	0.187	0.203
-2 log likelihood testing whether all coefficients (except the intercept) = 0	1121.201***	938.837***

EQUITY	Equity as a percentage of total assets.		UNCOLLECTED	Interest accrued as revenue but not collected as a percentage of total loans.
BAD-LOANS	Nonperforming loans as a percentage of total loans.			
OREO	Other real estate owned (real estate other than bank premises) as a percentage of total loans.		LIQUID	Liquid assets (sum of cash, securities, federal funds sold, and reverse repurchase agreements) as a percentage of total assets.
CONSUMER	Consumer loans as a percentage of total assets.			
INSIDER	The value of loans to insiders (officers and directors of the bank) as a percentage of total assets.		LARGE-TIME	Large denomination time deposit liabilities as a percentage of total assets.
			CORE	Core deposits (transactions, savings and small time deposits) as a percentage of total assets.
OVERHEAD	Noninterest expense as a percentage of total revenue.			
			SIZE	Natural logarithm of total assets, in thousands of dollars.
OCCUPANCY	Occupancy expense as a percentage of average assets.		BHCRATIO	The ratio of each bank's total assets to the total assets of its holding company. Banks without holding companies have BHCRATIO $\equiv$ 1.
ROA	Net income as a percentage of total assets.			

Supplemental Table 2, Continued: How Well Does the Logit Model Fit the CAMEL Downgrade Data?
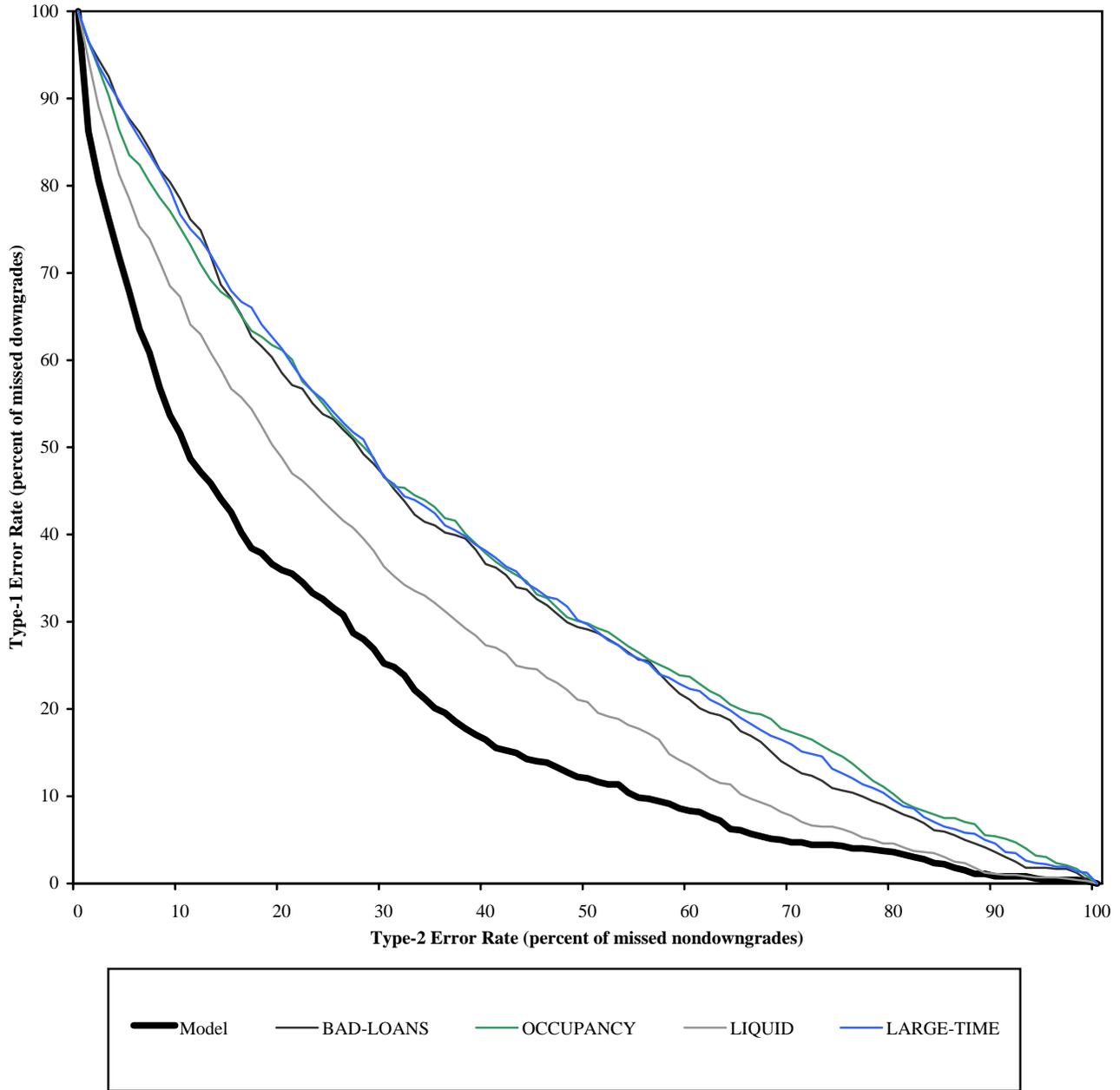
This table presents the estimated regression coefficients for the downgrade prediction logit. The model predicts in-sample downgrades ("1" represents downgrade from safe-and-sound to problem status) for calendar year *t* with year *t*-2 call report data. Standard errors appear in parentheses below each coefficient. Three asterisks denote significance at the 1 percent level; two asterisks denote significance at the 5 percent level. Shading highlights coefficients that were significant with the correct sign in all seven years. Overall, the logit model does a good job predicting in-sample downgrades.

Independent Variables	Banks that were examined in:		
	1993	**1994**	**1995**
Intercept	-2.249	-2.709	-1.768
	(2.152)	(2.807)	(2.380)
EQUITY	-0.067**	-0.081**	-0.050
	(0.033)	(0.039)	(0.038)
BAD-LOANS	0.235***	0.100***	0.216***
	(0.032)	(0.029)	(0.047)
OREO	0.084**	0.106**	0.168***
	(0.037)	(0.050)	(0.057)
CONSUMER	-0.025***	-0.002	0.011
	(0.010)	(0.012)	(0.011)
INSIDER	0.067	-0.006	-0.007
	(0.059)	(0.078)	(0.052)
OVERHEAD	0.041***	0.013	0.026***
	(0.009)	(0.011)	(0.010)
OCCUPANCY	-0.922***	0.434	0.050
	(0.281)	(0.366)	(0.438)
ROA	-0.367***	-0.446***	-0.207
	(0.102)	(0.152)	(0.171)
UNCOLLECTED	0.067	0.372**	0.519**
	(0.142)	(0.186)	(0.219)
LIQUID	-0.070***	-0.027***	-0.031***
	(0.006)	(0.006)	(0.008)
LARGE-TIME	0.057***	0.059**	0.058***
	(0.018)	(0.025)	(0.019)
CORE	0.012	0.011	-0.012
	(0.017)	(0.022)	(0.015)
SIZE	-0.136*	-0.276***	-0.272***
	(0.071)	(0.090)	(0.101)
BHCRATIO	1.905***	2.412***	1.572***
	(0.275)	(0.427)	(0.406)
Number of Observations	7,184	7,071	6,805
Pseudo-R^2	0.190	0.120	0.126
-2 log likelihood testing whether all coefficients (except the intercept) = 0	483.741***	217.959***	172.287***

EQUITY	Equity as a percentage of total assets.	UNCOLLECTED	Interest accrued as revenue but not collected as a percentage of total loans.
BAD-LOANS	Nonperforming loans as a percentage of total loans.		
OREO	Other real estate owned (real estate other than bank premises) as a percentage of total loans.	LIQUID	Liquid assets (sum of cash, securities, federal funds sold, and reverse repurchase agreements) as a percentage of total assets.
CONSUMER	Consumer loans as a percentage of total assets.		
INSIDER	The value of loans to insiders (officers and directors of the bank) as a percentage of total assets.	LARGE-TIME	Large denomination time deposit liabilities as a percentage of total assets.
		CORE	Core deposits (transactions, savings and small time deposits) as a percentage of total assets.
OVERHEAD	Noninterest expense as a percentage of total revenue.		
		SIZE	Natural logarithm of total assets, in thousands of dollars.
OCCUPANCY	Occupancy expense as a percentage of average assets.	BCHRATIO	The ratio of each bank's total assets to the total assets of its holding company. Banks without holding companies have BHCRATIO ≡ 1.
ROA	Net income as a percentage of total assets.		

Figure 5, Continued: What is the Trade-Off Between False Negatives and False Positives in the Downgrade-Prediction Model Compared to the Individual Screens?
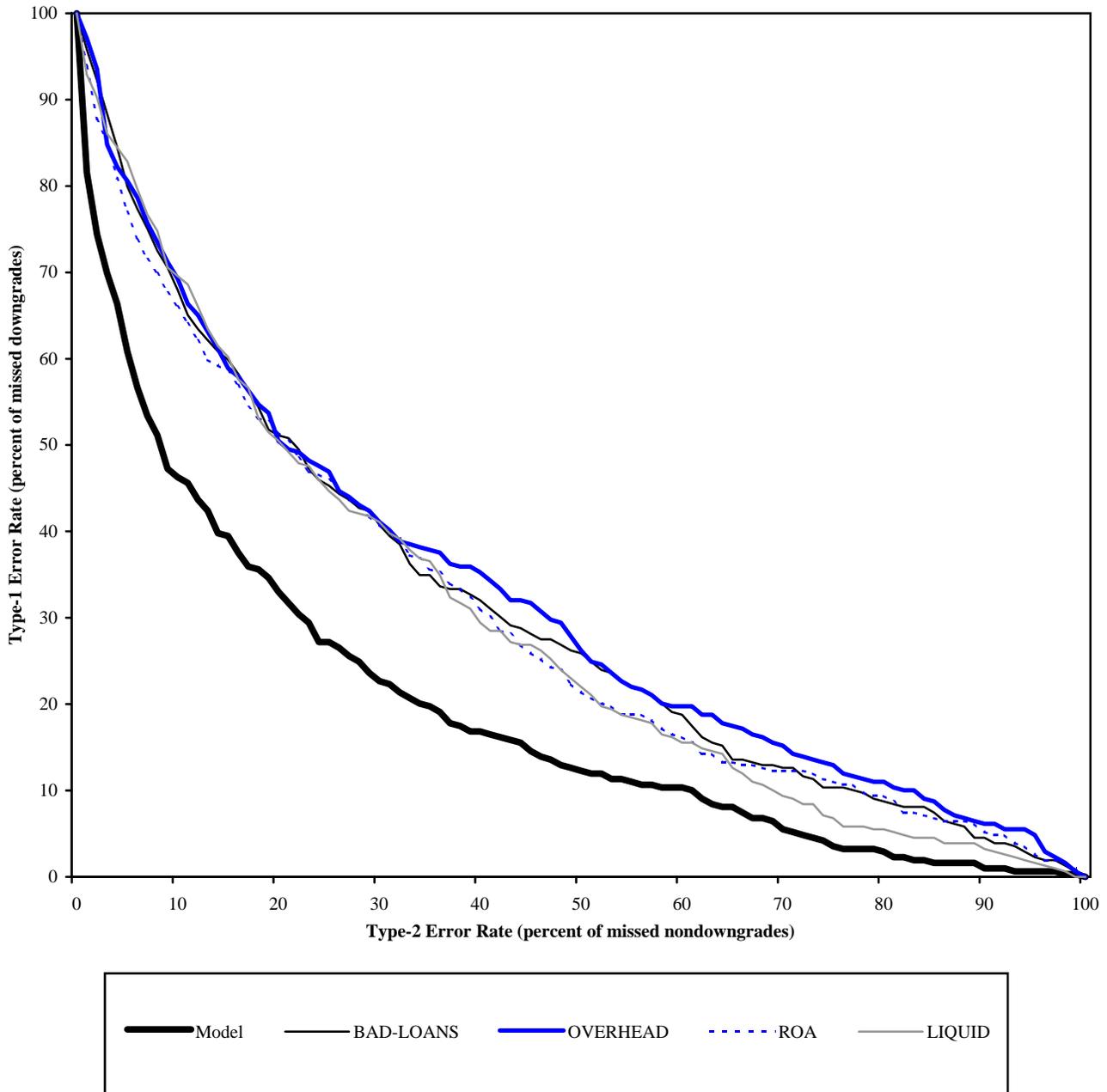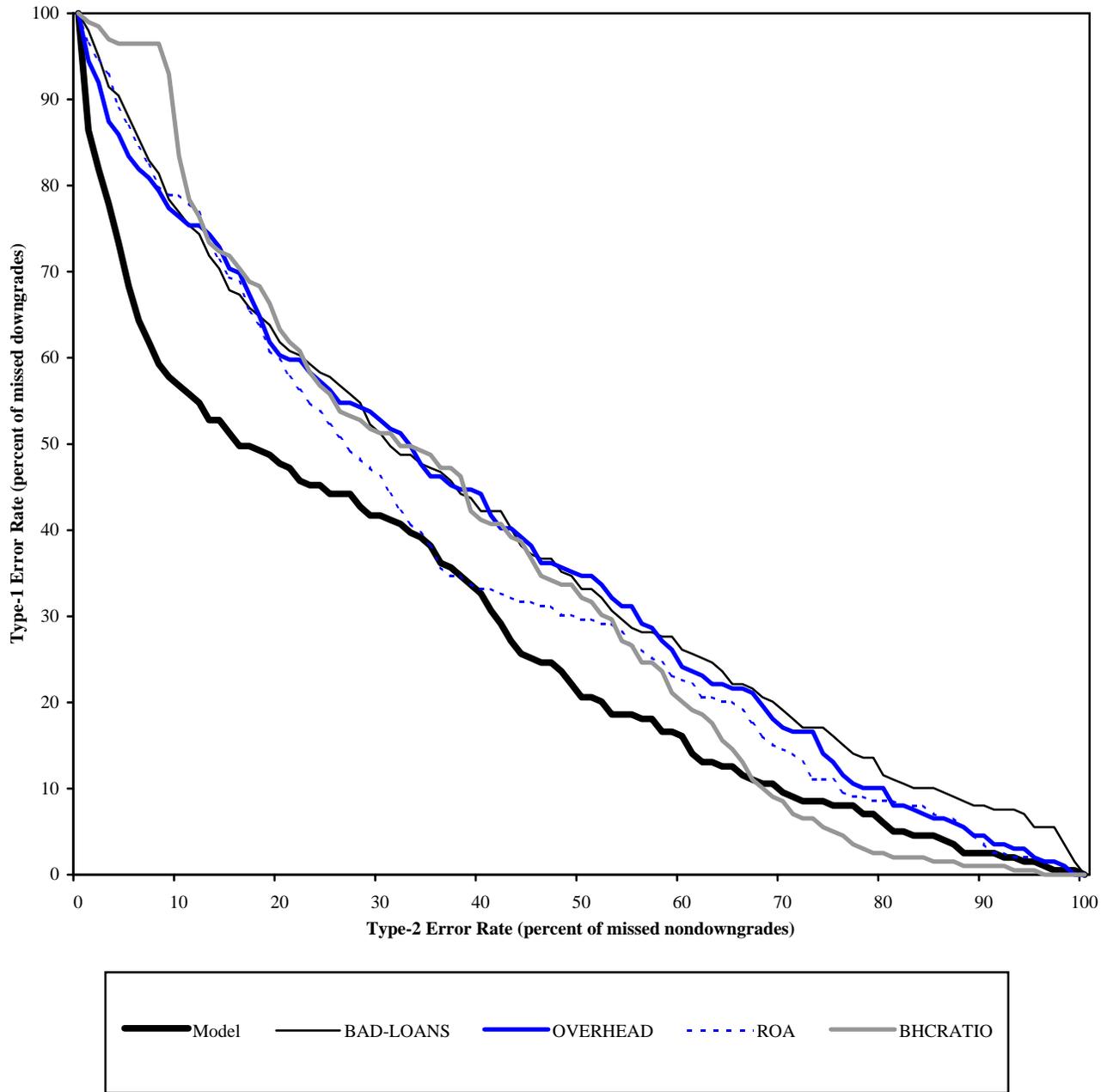
1992 Downgrade Predictions Using Year-End 1990 Data

This figure shows the trade-off between the type-1 error rate (missed downgrades) and the type-2 error rate (missed nondowngrades). The type-1 error rate is the percentage of banks rated CAMEL-1 or -2 that were subsequently downgraded by supervisors but were not identified by the model (or screen). The type-2 error rate is the percentage of banks rated CAMEL-1 or -2 that were not subsequently downgraded but were misidentified by the model (or screen) as a downgrade risk. A desirable early-warning system minimizes the increase in type-2 errors for any given decrease in type-1 errors. This graph shows that for any level of type-1 error rate tolerated by supervisors, the econometric model (in bold) leads to fewer type-2 errors than any individual screen. For clarity, only the four best screens are shown.

Figure 5, Continued: What is the Trade-Off Between False Negatives and False Positives in the Downgrade-Prediction Model Compared to the Individual Screens?
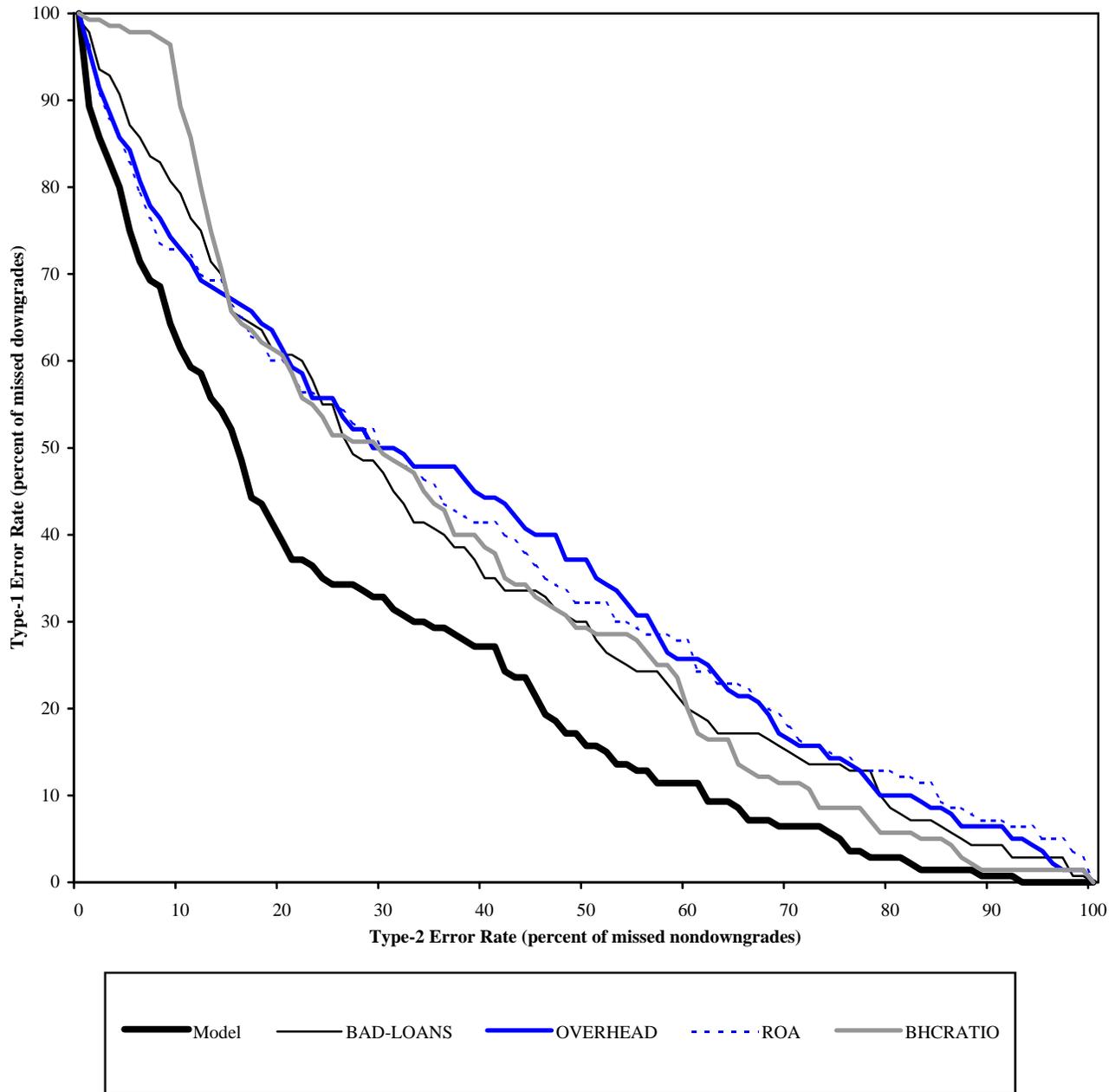
1993 Downgrade Predictions Using Year-End 1991 Data

This figure shows the trade-off between the type-1 error rate (missed downgrades) and the type-2 error rate (missed nondowngrades). The type-1 error rate is the percentage of banks rated CAMEL-1 or -2 that were subsequently downgraded by supervisors but were not identified by the model (or screen). The type-2 error rate is the percentage of banks rated CAMEL-1 or -2 that were not subsequently downgraded but were misidentified by the model (or screen) as a downgrade risk. A desirable early-warning system minimizes the increase in type-2 errors for any given decrease in type-1 errors. This graph shows that for any level of type-1 error rate tolerated by supervisors, the econometric model (in bold) leads to fewer type-2 errors than any individual screen. For clarity, only the four best screens are shown.

Figure 5, Contitnued: What is the Trade-Off Between False Negatives and False Positives in the Downgrade-Prediction Model Compared to the Individual Screens?
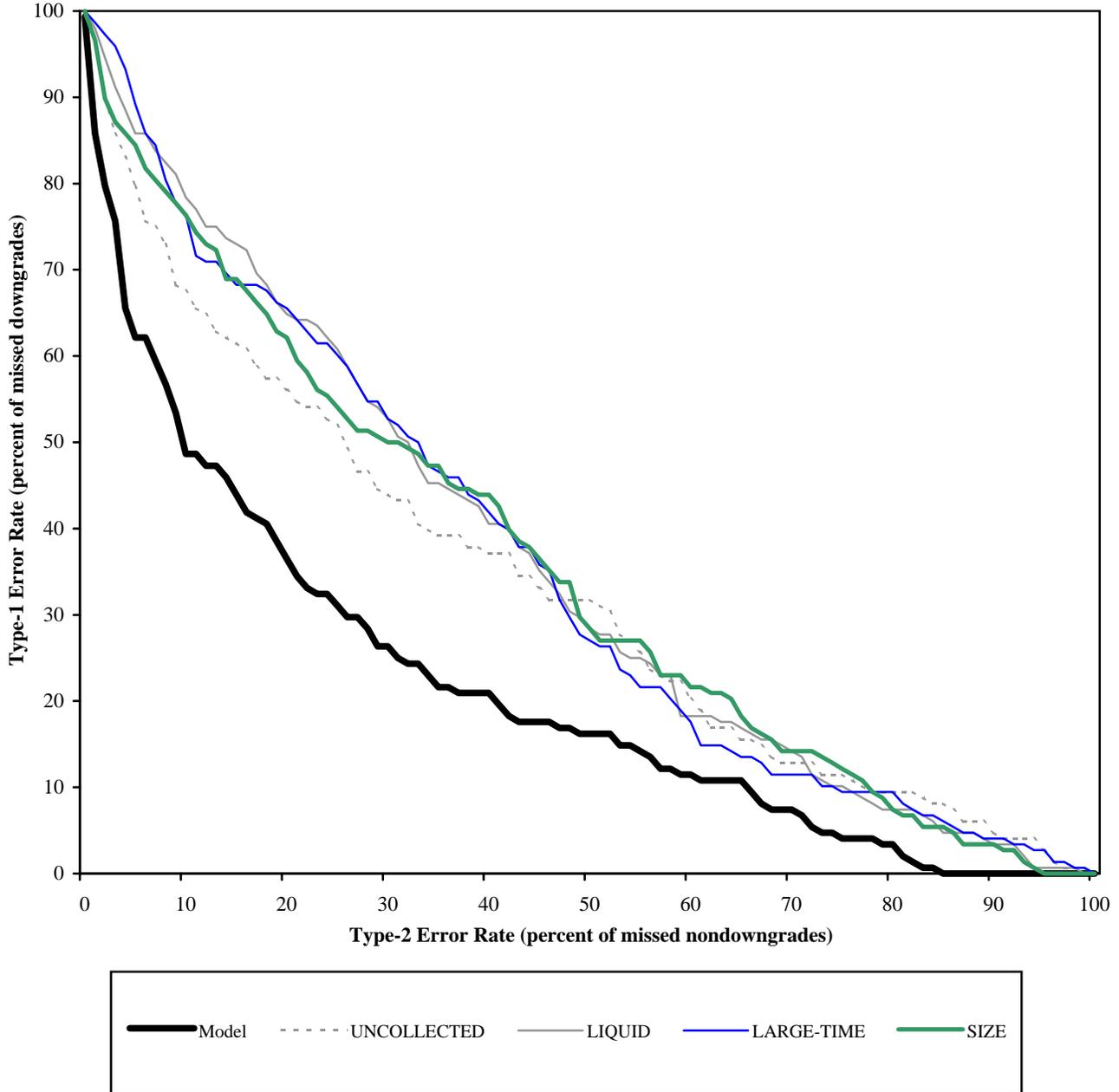
1994 Downgrade Predictions Using Year-End 1992 Data

This figure shows the trade-off between the type-1 error rate (missed downgrades) and the type-2 error rate (missed nondowngrades). The type-1 error rate is the percentage of banks rated CAMEL-1 or -2 that were subsequently downgraded by supervisors but were not identified by the model (or screen). The type-2 error rate is the percentage of banks rated CAMEL-1 or -2 that were not subsequently downgraded but were misidentified by the model (or screen) as a downgrade risk. A desirable early-warning system minimizes the increase in type-2 errors for any given decrease in type-1 errors. This graph shows that for any level of type-1 error rate tolerated by supervisors, the econometric model (in bold) leads to fewer type-2 errors than any individual screen. For clarity, only the four best screens are shown.

Figure 5, Continued: What is the Trade-Off Between False Negatives and False Positives in the Downgrade-Prediction Model Compared to the Individual Screens?
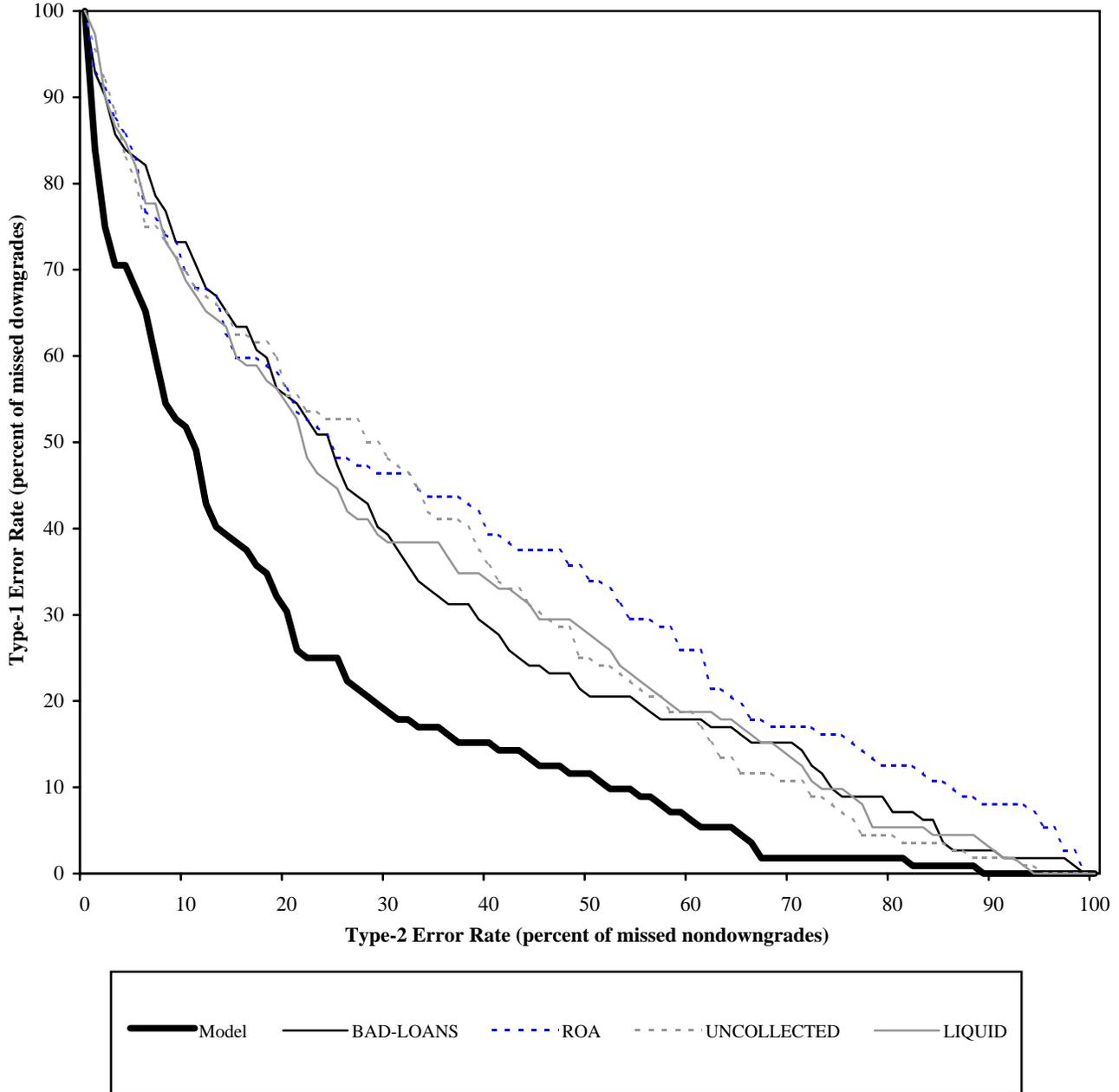
1995 Downgrade Predictions Using Year-End 1993 Data

This figure shows the trade-off between the type-1 error rate (missed downgrades) and the type-2 error rate (missed nondowngrades). The type-1 error rate is the percentage of banks rated CAMEL-1 or -2 that were subsequently downgraded by supervisors but were not identified by the model (or screen). The type-2 error rate is the percentage of banks rated CAMEL-1 or -2 that were not subsequently downgraded but were misidentified by the model (or screen) as a downgrade risk. A desirable early-warning system minimizes the increase in type-2 errors for any given decrease in type-1 errors. This graph shows that for any level of type-1 error rate tolerated by supervisors, the econometric model (in bold) leads to fewer type-2 errors than any individual screen. For clarity, only the four best screens are shown.

Figure 5, Continued: What is the Trade-Off Between False Negatives and False Positives in the Downgrade-Prediction Model Compred to the Individual Screens?

1996 Downgrade Predictions Using Year-End 1994 Data

This figure shows the trade-off between the type-1 error rate (missed downgrades) and the type-2 error rate (missed nondowngrades). The type-1 error rate is the percentage of banks rated CAMEL-1 or -2 that were subsequently downgraded by supervisors but were not identified by the model (or screen). The type-2 error rate is the percentage of banks rated CAMEL-1 or -2 that were not subsequently downgraded but were misidentified by the model (or screen) as a downgrade risk. A desirable early-warning system minimizes the increase in type-2 errors for any given decrease in type-1 errors. This graph shows that for any level of type-1 error rate tolerated by supervisors, the econometric model (in bold) leads to fewer type-2 errors than any individual screen. For clarity, only the four best screens are shown.

Figure 5, Continued: What is the Trade-Off Between False Negatives and False Positives in the Downgrade-Prediction Model Compared to the Individual Screens?

1997 Downgrade Predictions Using Year-End 1995 Data

This figure shows the trade-off between the type-1 error rate (missed downgrades) and the type-2 error rate (missed nondowngrades). The type-1 error rate is the percentage of banks rated CAMEL-1 or -2 that were subsequently downgraded by supervisors but were not identified by the model (or screen). The type-2 error rate is the percentage of banks rated CAMEL-1 or -2 that were not subsequently downgraded but were misidentified by the model (or screen) as a downgrade risk. A desirable early-warning system minimizes the increase in type-2 errors for any given decrease in type-1 errors. This graph shows that for any level of type-1 error rate tolerated by supervisors, the econometric model (in bold) leads to fewer type-2 errors than any individual screen. For clarity, only the four best screens are shown.

Testing Long-Run Monetary Neutrality Propositions: Lessons from the Recent Research

James Bullard

Monetary economists long have thought that government injections of money into a macroeconomy have a certain neutral effect. The main idea is that changes in the money stock eventually change nominal prices and nominal wages, ultimately leaving important real variables, like real output, real consumption expenditures, real wages, and real interest rates, unaffected. Since economic decision making is based on real factors, the long-run effect of injecting money into the macroeconomy is often described as *neutral*—in the end, real variables do not change and so economic decision making is also unchanged. How long such a process takes, and what might happen in the meantime, are hotly debated questions. But relatively few economists debate the merits of *long-run neutrality*. Indeed, long-run neutrality is instead taken as a given, almost an axiom, a logical consequence of suppositions made in economic theory.

Curiously, during most of the postwar period the empirical evidence on long-run monetary neutrality has been in a state of flux. No doubt this is in part because it is difficult to look at the data generated by the world's economies and come to any firm conclusion about whether monetary injections had important real effects, in the short run or in the long run. In addition, many of the empirical tests that were devised ran into important criticisms that seemed to invalidate their conclusions. These criticisms were based, at least in part, on questionable handling or interpretation of the time-series properties of the data. In recent years, however, economists have devised new tests of long-run monetary neutrality, as well as related neutrality-type propositions. A fair amount of literature has been written on the subject, and the purpose of this paper is to review this literature.[1]

The next section provides more detail concerning the background behind the current empirical tests of neutrality propositions. In the following sections, some of the recent research using the newer set of tests is reviewed, and a few related papers are discussed along with the results authors have found using somewhat different methodologies. The final section offers some comments about directions for future research.

SOME BACKGROUND

What is Long-Run Neutrality?

In discussing long-run monetary neutrality, economists typically refer to a specific, hypothetical experiment that normally is not observed directly in actual economies. The experiment is a one-time, permanent, unexpected change in the level of the money stock. If, for instance, the money stock was $5 billion one day, and had been $5 billion for a long time, then what would the effect be of suddenly changing it to $6 billion and keeping it there for a long time? According to the quantity theory of money, prices should rise eventually in proportion to the increase in the money stock, and all real variables, perhaps after some transition

[1] Not all papers dealing with neutrality issues—an enormous amount of literature—can be surveyed here. Instead, attention is restricted to those that use the newer techniques discussed later.

time, would return to their original values and stay there until some further disturbance comes along. This is long-run monetary neutrality.

In the hypothetical experiment, it is important that the new level of the money stock be maintained for some, possibly long, period of time, to allow the transition effects to vanish. Theoretically, the change in the money stock has to be "permanent." In the world's economies, we observe a high degree of persistence in many macroeconomic variables, but it is generally difficult to tell the difference between "highly persistent" and "permanent." In the empirical work surveyed below we will see the use of many tests— *unit root diagnostic tests*—intended to categorize macroeconomic variables into those that have been subject to permanent shocks and those that have not. It is important to bear in mind, however, that these tests may not accurately distinguish between the two cases—statistically speaking, the tests have limited power. The tests are used because they offer the best available method for making the distinction between highly persistent and permanent changes, but they are far from perfect.

In the hypothetical experiment, it is also important that the change be unexpected, because if the economy's participants knew that the money stock was going to increase, and therefore, that prices were about to increase, they might start changing their present behavior. For example, they might buy consumption goods today, before the price increase takes effect. Prices then might begin to rise *in advance* of the money stock change. This complicates the story, and hence, we will think in terms of unanticipated changes in the money stock level. In the discussion below, this will be approximated by the notion of a "permanent shock" to the money supply.

In the world of monetary theory, nearly all models based on standard economic assumptions embody some form of monetary neutrality.[2] Most likely this is because monetary theorists generally think long-run monetary neutrality is sensible, and, therefore, they build it into their models.

Empirical tests that convincingly documented departures from long-run monetary neutrality therefore would be quite surprising (or quite suspect!) to monetary economists.

There is a second hypothetical experiment, related to the first, that more closely resembles the types of monetary policy actions we see in actual economies. This experiment says that the government initially maintains a certain *growth rate* for the money stock for a long period of time. At some date, that growth rate is adjusted unexpectedly to some new rate, say, from 3 percent to 5 percent on an annual basis, and is kept there for another long period of time. What effect should this have on important real variables like the capital-labor ratio, real output, real consumption expenditures, and real interest rates? If the answer is that after a long period of time, nothing would happen to the real variables, we have what is referred commonly to as long-run monetary *superneutrality*. Here again, one might expect an important transition period (commonly known as "the short run") when the economy is adjusting to the new rate of monetary growth. Quite a lot could happen to real variables during this adjustment period. But the neutrality and superneutrality propositions discussed in this paper mainly concern long-run, limiting effects.

Perhaps surprisingly, there are many plausible analyses that suggest that departures from long-run monetary superneutrality might be consistent with standard economic theory. It is, in fact, relatively easy to produce such theories. Moreover, these departures could go either way; that is, a permanently higher rate of monetary growth might eventually either raise or lower the level of economic activity, or change other important real variables in either a positive or negative direction.[3] Accordingly, whereas long-run neutrality is taken almost as an axiom of monetary economics, long-run superneutrality is far more circumspect. An empirical test that convincingly showed departures from long-run superneutrality would not be too surprising, since this result

[2] The phrase "standard economic assumptions" means maintaining assumptions that markets clear at all times and that all agents behave rationally.

[3] For a description of these departures, see the survey by Orphanides and Solow (1990).

is consistent with a number of existing economic theories.[4]

It is important to note that whether the level of real output rises or falls, or whether other real variables change in a particular direction in response to a permanent increase in the money growth rate, does not have any particular connotations for social welfare. In many theories, inflation distorts a Pareto optimal equilibrium, so that as a long-run proposition the population in the economy generally prefers lower rates of money growth accompanied by lower rates of inflation. Different theories make different predictions in this regard, however, and to sort these out one would have to consider various theories and their underlying assumptions in some detail. Since this would take us too far astray, social welfare will not be addressed in this survey.

There is another side to the superneutrality question. Fischer (1996) suggests that the reason the central banks of the world's industrialized economies have avidly pursued long-run price stability is because in the long run, inflation has distortionary effects that adversely impact a real variable, or a group of real variables, that people care about. If monetary growth causes inflation, and inflation has distortionary effects, then long-run monetary superneutrality should not hold in the data. On the contrary, a permanent shock to the rate of monetary growth should have some long-run effect on the real economy; why else should we worry about it? Care needs to be taken, however, in defining which variables are supposed to be affected and which are not—this is an area of some confusion in the literature.[5] In the current paper we will try to avoid this problem through use of the language "superneutrality with respect to variable *x*."

The above discussion has referred to changes in real variables, meaning changes in the *level* of the variable, especially so with respect to the level of real output. Of course, real output in industrialized economies generally grows over time. A shift in the level would be a one-time movement, say from 100 to 90, whereupon the variable would resume growing at its previous rate.

Thus permanent effects on the level of a variable need not imply permanent effects on the growth rate of that variable. Consequently, a natural question to ask is whether permanent changes in the monetary growth rate affect a country's *rate* of economic growth; that is, is money superneutral with respect to economic growth? Many researchers in recent years have in fact investigated questions of this type (mostly with methodology outside the focus of this survey). There is much less theory concerning this issue, but some of the results I discuss later will have some bearing on this topic.
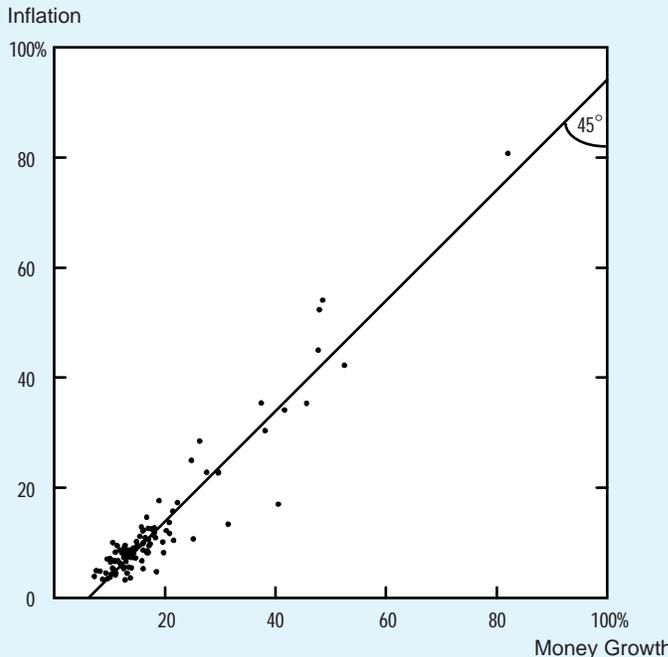
Prima Facie Evidence. In his Nobel Lecture, Lucas (1996) addresses the topic of monetary neutrality, both in the short run and the long run, and discusses theoretical developments that might reconcile the perceived short-run effects of an increase in the money supply with long-run monetary neutrality. Lucas mentions several pieces of evidence as constituting the main reasons that he would like a satisfactory theory of the real effects of monetary policy to address. Among these, he cites Friedman and Schwartz (1963) who argue that all major recessions in the United States between 1867 and 1960 were preceded by substantial contractions in the money supply, suggesting that monetary policy mistakes were a primary contributor to business cycle downturns during this period. Lucas states that severe monetary contraction seemed to play an important role especially during the Great Depression of 1929-33. But he also cites work by Sargent (1986) who argues that huge reductions in the rate of monetary expansion —reductions much larger than anything experienced in the post-Civil War United States—did not lead to any unusually large reduction in real output in the hyperinflationary post-World War I European economies. These reductions were carried out in conjunction with monetary reform. The hyperinflations ended abruptly when credible reform was announced. But these citations are subsidiary to Lucas's (1996, p. 668) main contention, that there is clear

[4] The situation described is summarized by Canova (1994, p. 123), who states, "... there are very few available models which display superneutrality, while most existing models, both in the neoclassical and neoKeynesian tradition, possess neutrality of money...."

[5] See Marty (1994) for a discussion.

Figure 1

Money Growth and Inflation

Inflation

Postwar average rates of money growth versus average inflation rates in 110 countries. Observations near the 45 degree line, which is not fitted to the data, are consistent with the quantity theory. This figure is from McCandless and Weber (1995).

theories can claim empirical success at the level exhibited in figure 1? ... The kind of monetary neutrality shown in this figure needs to be a central feature of any monetary or macroeconomic theory that claims empirical seriousness."

While Figure 1 is impressive, one should be careful to note that these results are different from the stories about long-run monetary neutrality and superneutrality outlined above. Evidently, the average rate of money growth is correlated highly with the average rate of inflation in a country. But the story about long-run monetary neutrality is about a permanent, unexpected change in the level of the money stock in a single country, and the ultimate impact of such a change. And, the story about superneutrality concerns the long-run effect of a permanent, unexpected change in the rate of monetary expansion. Taking averages over long periods of time, while informative at some level, masks the information about such events, to the extent they might have occurred in the data. To study long-run neutrality more directly, the time-series evidence on inflation and monetary growth for individual countries needs to be considered. Can we isolate permanent, or at least highly persistent, changes in the money stock (or the monetary growth rate), which are then correlated with persistent changes in the price level (or the rate of inflation) and simultaneously are uncorrelated with permanent movements in important real variables? That is the challenge of testing monetary neutrality propositions.

Time-Series Evidence. Some tests of long-run monetary neutrality during the 1960s simply regressed the level of real output on a distributed lag of observations on the money stock. In reaction to this practice, Sargent (1971) and Lucas (1972) argued that such evidence was circumspect for two related reasons. One is that Sargent and Lucas built simple and plausible reduced-form models of the macroeconomy in which long-run monetary neutrality held by construction, but which also would produce data such that, if the standard

evidence—even "decisive confirmation"—that long-run monetary neutrality holds.

Figure 1 shows the evidence that Lucas (1996) cites. This figure, from McCandless and Weber (1995), plots the average rates of monetary growth against average rates of inflation for 110 countries. The averages are taken over 30 years, 1960-90. Monetary growth is measured as the annual growth rate of M2 for a country, and inflation is measured as the annual rate of increase in the consumer price index for a country. The 45-degree line is not fit to the data, but instead represents a theoretical presumption based on the quantity theory, that the rate of inflation should correspond to the rate of money growth (adjusted for the real output growth rate in a particular economy). McCandless and Weber report a simple correlation of .95 between money growth and inflation based on this data. Lucas (1996, p. 666) asks "... how many specific economic

practice was applied, the researcher would conclude that long-run monetary neutrality failed. Thus, any evidence based on the (then) standard methodology was difficult to interpret.

The second reason—the one that is at the heart of the methods used in the recent research—was that the story of monetary neutrality involves permanent changes in the level of the money stock, and that one cannot effectively test such a theory without evidence that the actual money stock has been subject to a permanent change. The idea of permanent changes in economic variables is statistically modeled as a unit root in the autoregressive representation of a time series; a time series with a unit root has quite different properties from a stationary series.[6] During the early 1970s when Lucas and Sargent first wrote about this topic, the implications of unit roots in economic time series were only beginning to be appreciated. Later, in an influential paper, Nelson and Plosser (1982) argued that many U.S. macroeconomic time series were best characterized by a unit root in their univariate, autoregressive representations. Their results brought the issue of how to handle these nonstationary time series to the fore in macroeconometrics, and led to econometric methodologies that respected the potential for nonstationarity in important macroeconomic variables.

The nonstationarity in economic variables was viewed as something of a headache for much of macroeconometrics. But in a remarkable turn of events, it actually was a boon to testing neutrality propositions. As Lucas and Sargent had argued, one needs permanent changes in the money stock as part of the historical record to test the proposition of long-run neutrality in a time-series setting. But permanent shocks are exactly what macroeconomic time series provide.

This was exactly the line pursued by Fisher (1988) and Fisher and Seater (1989, 1993), and also in a series of papers by King and Watson (1992, 1994, 1997). These authors provided new tests of long-run neutrality propositions that respected

the Lucas-Sargent critique and required little macroeconomic structure.

TESTING NEUTRALITY PROPOSITIONS

Recent Tests Based (Mostly) on U.S. Data

Fisher and Seater (1993) work in terms of bivariate systems, with a measure of money as one of the variables. Adopting their notation, let m be the natural logarithm of the nominal money stock M. Let y be a second variable, expressed in either real or nominal terms, which is the logarithm of a variable like the price level or real output, and where the variable itself is Y.[7] Denote the order of integration of a variable by $\langle x \rangle$, so that if x is integrated of order 1, we write $\langle x \rangle = 1$. Sometimes we also will use the phrase "x is $I(1)$" to describe the order of integration. Denote a difference operator by Δ, so that Δy indicates the approximate growth rate of the variable Y. Fisher and Seater study the following system

$$(1) \qquad a(L)\Delta^{\langle m \rangle} m_t = b(L)\Delta^{\langle y \rangle} y_t + u_t$$

$$(2) \qquad d(L)\Delta^{\langle y \rangle} y_t = c(L)\Delta^{\langle m \rangle} m_t + w_t$$

where $a(L)$, $b(L)$, $c(L)$ and $d(L)$ are lag polynomials, and $a_0 = d_0 = 1$ and b_0 and c_0 are unrestricted. The error vector $(u_t, w_t)'$ is *iid* with zero mean and covariance Σ. Now let $x_t \equiv \Delta^i m_t$ and $z_t \equiv \Delta^j y_t$, with $i, j = 0$ or 1. Fisher and Seater define a certain *long-run derivative* (LRD) that is central to their findings. The LRD is a change in z with respect to a permanent change in x, given by

$$(3) \qquad LRD_{z,x} \equiv \lim_{k \to \infty} \frac{\partial z_{t+k}/\partial u_t}{\partial x_{t+k}/\partial u_t},$$

provided $\lim_{k \to \infty} \partial x_{t+k}/\partial \mu_t \neq 0$, otherwise the LRD is undefined. Fisher and Seater then define long-run neutrality and long-run

superneutrality in this framework, and for each, discuss four cases that depend on the order of integration of the variables.[8]

First of all, money is long-run neutral with respect to y if $LRD_{y,m} = 1$ when y is a nominal variable, or if $LRD_{y,m} = 0$ when y is a real variable. The four cases are: 1) $\langle m \rangle < 1$. Here the LRD is not defined because there have been no permanent shocks to the level of the money stock, and the data are uninformative concerning long-run monetary neutrality. 2) $\langle m \rangle \geq \langle y \rangle + 1 \geq 1$. Here the LRD is zero because while there have been permanent shocks to the level of the money stock, there have been none to y. If y is a nominal variable, long-run neutrality is violated, otherwise it holds. 3) $\langle m \rangle = \langle y \rangle \geq 1$. This case admits tests of long-run neutrality, in an effort to find out if the permanent shocks to the level of the money stock are correlated with the permanent shocks to the variable y. 4) $\langle m \rangle = \langle y \rangle - 1 \geq 1$. This case is more complicated. A necessary condition for long-run neutrality is that the permanent shock to money does not change the growth rate of y.

Secondly, money is long-run superneutral with respect to y if $LRD_{y,\Delta m} = 0$. The cases are 1) $\langle \Delta m \rangle < 1$. Here the LRD is not defined because there have been no permanent shocks to the growth rate of the money stock, and the data are uninformative concerning long-run monetary superneutrality. 2) $\langle \Delta m \rangle \geq \langle y \rangle + 1 \geq 1$. The LRD is zero because while there have been permanent shocks to the growth rate of the money stock, there have been none to y. Long-run superneutrality holds. 3) $\langle \Delta m \rangle = \langle y \rangle \geq 1$. This case admits tests of long-run superneutrality, in an effort to find out if the permanent shocks to the level of the money stock are correlated with the permanent shocks to the variable y. 4) $\langle \Delta m \rangle = \langle y \rangle - 1 \geq 1$. Here $LRD_{\Delta y, \Delta m} = 0$ is testable; that is, one can determine whether a permanent change in the growth rate of money is associated with a permanent change in the growth rate of y.

Fisher and Seater (1993) use these results to analyze previous research efforts testing long-run neutrality propositions, efforts that, because of the time they were written, did not take such explicit account of the time-series properties of the data. They interpret the evidence in Anderson and Karnovsky (1972), Kormendi and Meguire (1984), Lucas (1980), and Geweke (1986) mostly as consistent with long-run neutrality and not very informative about long-run superneutrality. They also provide some evidence of their own. They use the Friedman and Schwartz (1982) data on money, prices, nominal income, and real income from 1867 to 1975 in the United States. All variables are viewed as $I(1)$, making tests of long-run neutrality possible. With respect to nominal income and prices, long-run monetary neutrality holds in this data, but with respect to real output, long-run monetary neutrality fails.

As mentioned earlier, evidence of the failure of long-run monetary neutrality is either surprising or suspect among monetary theorists; the Fisher and Seater finding was no exception. In a note, Boschen and Otrok (1994) re-estimate the systems studied by Fisher and Seater, again using the Friedman and Schwartz (1982) data, but now updating the time series through 1992. They split the data set into two subsamples, 1869-1929 and 1940-92. They find that long-run neutrality holds in both of the subsamples using the Fisher and Seater methodology. They conclude that there may have been something special about the financial disruption during the Great Depression era that causes the test to fail when that period is included.

Haug and Lucas (1997) comment further on these findings. They reason that, since Canada did not experience bank failures during the Great Depression, the evidence on long-run neutrality using Canadian data might provide further evidence that something unusual happened in the United States during this period. Their data set includes real national income and the M2 money supply from 1914-94. They argue that pre-1914 data is inappropriate for this purpose because changes in the money supply were not exogenous in Canada at that time. They

---

[8] In an appendix, Fisher and Seater (1993) argue that cointegration plays no role in their bivariate tests of neutrality or superneutrality. This does not imply that one could not devise other, similar tests based on cointegration, as (in fact) has been done. See, for instance, the Boschen and Mills (1995) paper reviewed later in this section.

conclude, based on augmented Dickey-Fuller (ADF) tests, that both time series are $I(1)$. And, according to the Fisher and Seater (1993) methodology, long-run monetary neutrality with respect to real output cannot be rejected using the entire Canadian sample period. Haug and Lucas interpret this finding as independent support for the arguments of Boschen and Otrok (1994).

Olekalns (1996) similarly explored an alternative data set, 94 years of annual Australian data. The downturn of the 1930s was less severe in Australia. Olekalns uses the Fisher and Seater methodology, and two measures of money, M1 and M3, along with real gross domestic product. All variables are reasonably described as $I(1)$ according to ADF tests. Olekalns finds that long-run monetary neutrality cannot be rejected using the narrower money measure. However, using the broader money measure, long-run neutrality can be rejected for this data set, and the rejection carries even when dummy variables are used to control for the Depression period as well as the WWII period. Olekalns concludes that results can be sensitive to the measure of money used.

A recent paper by Coe and Nason (1999) also contributes to this literature. They use the Fisher and Seater (1993) test for long-run neutrality, and they employ the same U.S. data as Fisher and Seater, except that they update the data through 1997. When Coe and Nason use a broad measure of the money stock (as Fisher and Seater did), they replicate the Fisher and Seater rejection of long-run monetary neutrality with respect to real output. But when they replace the broad money measure with the monetary base, they can no longer reject long-run neutrality. They also consider about a century of data from the United Kingdom, and fail to reject long-run neutrality using either broad or narrow measures of money. Coe and Nason conclude that the Fisher and Seater rejection of long-run neutrality is not robust to a change in either the measure of money or the country of study.[9]

An important work in this literature, King and Watson (1997) also use bivariate systems, and they also take careful note of

the order of integration of the variables involved when devising tests of neutrality propositions. They study a "final form" model

$$(4) \quad \Delta y_t = \mu_y + \theta_{y\eta}(L)\varepsilon_t^\eta + \theta_{ym}(L)\varepsilon_t^m$$

$$(5) \quad \Delta m_t = \mu_m + \theta_{m\eta}(L)\varepsilon_t^\eta + \theta_{mm}(L)\varepsilon_t^m$$

where y is the logarithm of real output, the θ coefficients are lag polynomials, ε_t^m is a serially independent, zero mean shock to money, and ε_t^η is a vector of nonmonetary shocks that affect output. King and Watson show that

$$(6) \qquad \gamma_{ym} = \frac{\theta_{ym}(1)}{\theta_{mm}(1)}$$

is the long-run elasticity of real output with respect to permanent shocks to the money stock. Thus, long-run neutrality here is analogous to the Fisher and Seater definition: $\gamma_{ym} = 0$. Again, long-run neutrality can be investigated only if there have been permanent shocks to the money stock.

Importantly, King and Watson (1997) emphasize identification issues. They analyze long-run neutrality propositions across a range of possible identifications of their bivariate system, in an effort to understand the robustness of various conclusions to differing assumptions. They rewrite the equations (4) and (5) as

$$(7) \qquad \Delta y_t = \lambda_{ym}\Delta m_t + \sum_{j=1}^{p} \alpha_{j,yy}\Delta y_{t-j}$$
$$+ \sum_{j=1}^{p} \alpha_{j,ym}\Delta m_{t-j} + \varepsilon_t^\eta$$

$$(8) \qquad \Delta m_t = \lambda_{my}\Delta y_t + \sum_{j=1}^{p} \alpha_{j,my}\Delta y_{t-j}$$
$$+ \sum_{j=1}^{p} \alpha_{j,mm}\Delta m_{t-j} + \varepsilon_t^m$$
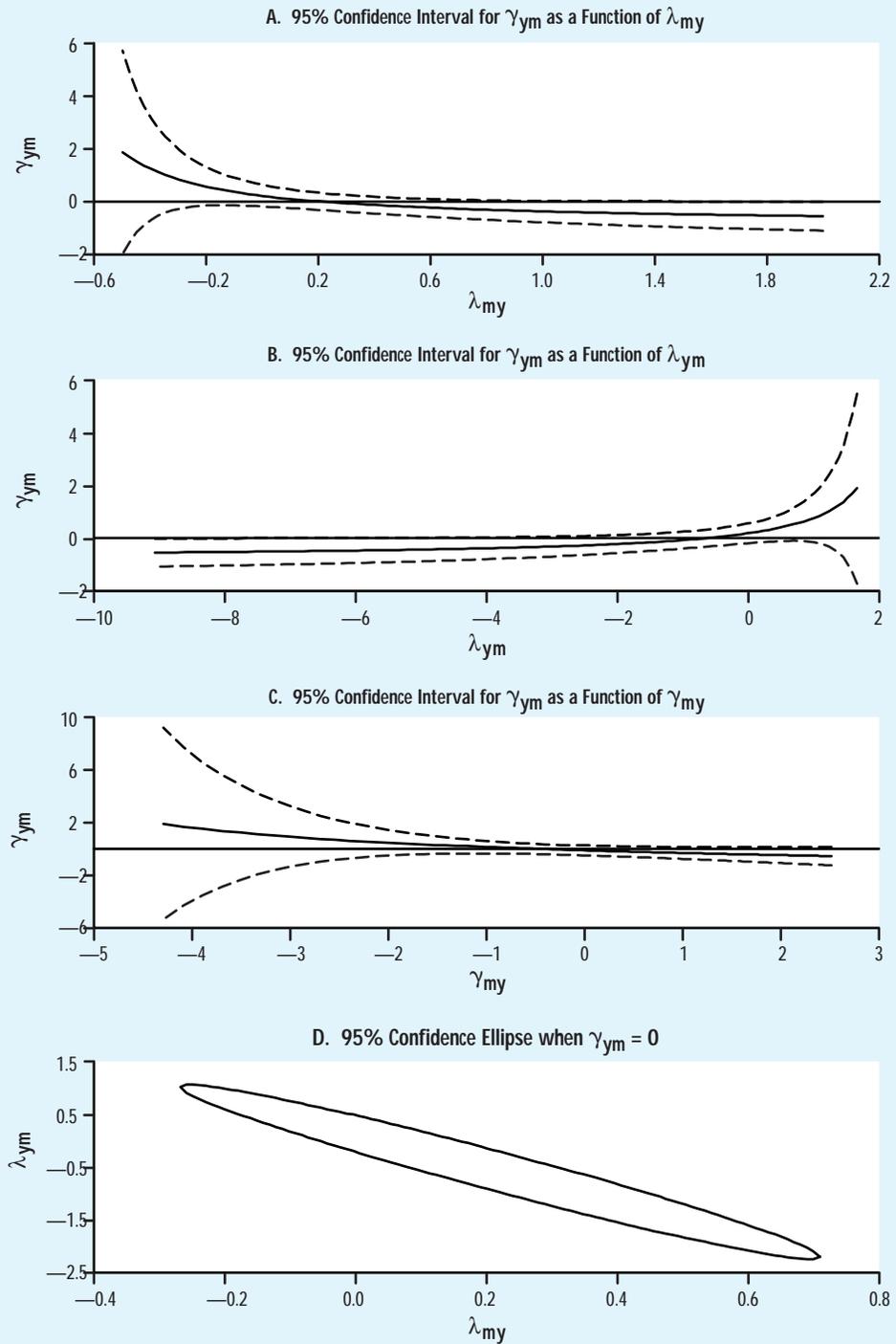
and they assume that
$$\text{cov}\left(\varepsilon_t^m, \varepsilon_t^\eta\right) = 0.$$

They note that there are several plausible ways to complete their identification of the system. One could assume $\lambda_{ym} = 0$, or that

[9] Coe and Nason also study the asymptotic power properties of the Fisher and Seater long-horizon regression test, and they conclude that the test has low power against alternative hypotheses of monetary non-neutrality. For small samples, Monte Carlo experiments reveal poor size-adjusted power especially at longer horizons. Coe and Nason conclude based on this portion of their analysis that the Fisher and Seater approach to testing long-run monetary neutrality may not be informative.

Inflation and Unemployment

A. 95% Confidence Interval for γ_{ym} as a Function of λ_{my}

B. 95% Confidence Interval for γ_{ym} as a Function of λ_{ym}

C. 95% Confidence Interval for γ_{ym} as a Function of γ_{my}

D. 95% Confidence Ellipse when $\gamma_{ym} = 0$

The evidence on long-run monetary neutrality according to King and Watson (1997). The panels show how the point estimate of γ_{ym} changes under the differing identifying restrictions, with the dotted lines indicating 95% confidence intervals. The bottom panel displays a 95% confidence ellipse for λ_{ym} and λ_{my} under the identifying restriction $\gamma_{ym} = 0$.

$\lambda_{my}=0$; this means that the impact elasticity of one variable on the other is zero. Alternatively, one could simply assume long-run monetary neutrality by imposing $\gamma_{ym}=0$. And finally, one could assume that $\gamma_{ym}=1$ where γ_{my} is the long-run elasticity of money with respect to a permanent shock to real output. King and Watson (1997, p.77) argue that this last assumption is consistent with stable prices in an economy with constant velocity.

Because King and Watson wish to investigate the robustness of neutrality results to alternative identifying assumptions, they use all four of these possibilities. Furthermore, they allow a wide variety of values for each elasticity, not just the zeroes and ones of the previous paragraph. Thus, the identifying assumptions are that either one of the impact elasticities is known to be a certain value, or that one of the long-run elasticities is known to be a certain value. They then turn to estimation and report results considering a number of neutrality propositions. The quarterly data are for the United States and cover the sample period from 1949:1 to 1990:4; except for systems with unemployment, in which case the sample period is from 1950:1 to 1990:4. The lag length p is set to six, although they experiment with values of four and eight at some points. Based on unit-root diagnostic tests, King and Watson conclude that all the series involved can reasonably be viewed as $I(1)$, so that tests of neutrality propositions can be executed.

King and Watson (1997) first investigate the long-run neutrality of money in the context of a bivariate system using real output and money (M2). They begin by estimating a value for γ_{ym} using the identifying assumption that λ_{my} is known. They find that a 95-percent confidence interval for γ_{ym} contains zero (and so supports long-run monetary neutrality) so long as $\lambda_{my} > 1.4$. If we interpret λ_{my} the parameter as a short-run elasticity of money demand, a reasonable range is $.1 \le \lambda_{my} \le .6$, so that the evidence is consistent with long-run neutrality. King and Watson complete similar calculations for identifying assumptions

involving λ_{ym} and γ_{my}. They also estimate 95-percent confidence intervals for λ_{my}, λ_{ym}, and γ_{my} using the identifying assumption that long-run neutrality holds, $\gamma_{ym}=0$, in order to see if the confidence intervals produced contain the most reasonable values for these parameters. All of this evidence comes down in favor of long-run neutrality, which is consistent with the findings of Fisher and Seater (1993) and Boschen and Otrok (1994), because the sample period here covers the postwar United States.[10] This evidence is summarized in Figure 2.

The superneutrality of money is investigated using a bivariate system with money growth (replacing the level of the money stock) and real output, and the hypothesis is that the long-run elasticity of the level of output to a permanent change in the growth rate of money, $\gamma_{y, \Delta m}$, is zero. The evidence on this question turns out to be mixed, in that for some identification schemes that King and Watson consider reasonable, the hypothesis that $\gamma_{y, \Delta m}=0$ can be rejected at the 5-percent level. Moreover, the effect can go either way: a permanently higher rate of money growth tending to permanently increase the level of real output, or to decrease the level of real output. For instance, if the identifying assumption is $\lambda_{\Delta m, y}=0$ (which King and Watson again interpret as a short-run money demand elasticity), then the estimated value of $\gamma_{y, \Delta m}$ is positive and statistically significant, while if the identifying assumption is that $\lambda_{\Delta m, y}=.6$, then the estimated value of $\gamma_{y, \Delta m}$ is negative and statistically significant. As mentioned earlier, theories exist that are consistent with both possibilities.

King and Watson go on to investigate a neutrality proposition associated with the early 20th century economist Irving Fisher. The proposition is that nominal interest rates move one-for-one with permanent changes in inflation, leaving the real interest rate unaffected. Using a system with consumer price index inflation, π, playing the role of the money variable, and the nominal interest rate on three-month Treasury bills, R, playing the

[10] Jefferson (1997) also investigates monetary neutrality questions using the King and Watson (1997) methodology, except that he considers measures of both inside money (defined as nominal checkable deposits, or M2 less currency) and outside money (defined as the monetary base). He uses nearly a century of data from the United States and finds some departures (under some identifying restrictions) from long-run neutrality when inside money is used.

role of the output variable, King and Watson (1997) investigate the hypothesis that $\gamma_{R\pi} = 1$; that is, the long-run elasticity of the nominal interest rate with respect to a permanent inflation shock is one. The evidence here again turns out to depend on the identification scheme. When statistically significant differences from the standard Fisher relation occur, they occur in a negative direction, with nominal interest rates rising less than one-for-one with perma-nent shocks to inflation. In other words, real interest rates are lowered permanently by permanent, positive shocks to the inflation rate. King and Watson find that identifying the model by assuming $\gamma_{R\pi} = 1$, and then estimating 95-percent confidence intervals for the remaining parameters, leads to the conclusion that there are reasonable configurations of parameters that are consistent with the Fisher hypothesis. Nevertheless, the main conclusion is that nominal interest rates do not adjust fully to permanent inflation shocks, and this conclusion holds across a large set of identification schemes.

Finally, King and Watson turn to estimating the slope of a long-run Phillips curve; that is, the long-run response of unemployment to permanent shocks in the inflation rate. This particular test is discussed in more detail in another paper, King and Watson (1994). The bivariate system now includes the CPI inflation rate in the role of the nominal variable, and the unemployment rate in the role of the real variable. The hypothesis is that the long-run Phillips curve is vertical, which means $\gamma_{u\pi} = 0$ in this framework. King and Watson report that a statistically significant (negative) slope for the long-run Phillips curve can be obtained only if the identifying assumption is that $\gamma_{\pi u} > 2.3$, or alternatively that $\gamma_{u\pi} < -0.7$. In particular, if either of these two impact elasticities are zero, then a vertical long-run Phillips curve cannot be rejected. King and Watson conclude that a reasonable estimate of the long-run Phillips curve based on this data is either vertical or at least "very steep."

While King and Watson's strong suit is that they can investigate the robustness of

results on neutrality for a wide variety of identification schemes, they do so only for bivariate systems, and they note the possibilities for exposure to omitted variable bias. One of the few multivariate studies available using techniques related to those of Fisher and Seater (1993) for testing neutrality is by Boschen and Mills (1995). They use the notion of permanent shocks to the level of the money stock to test long-run monetary neutrality in the U.S. data. They use a relatively high dimensional system, and they organize their research around the idea that, if long-run neutrality does not hold, there would be a nonstationary component of real output that is determined by long-term movements in the money stock. They study a vector error correction model (VECM) representation

$$(9) \quad \Delta X_t = \mu + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Pi X_{t-k} + \varepsilon_t,$$

where $X' \equiv (y, m, \upsilon)'$, and y is aggregate output, υ is a vector of real shocks, m is a vector of monetary shocks, and ε is normally distributed, *iid*, and has mean zero. Interest centers on the long-run impact coefficient matrix Π that describes the long-run relationships in the model. For each cointegrating relationship, this matrix will have a nonzero row. If there is a cointegrating relationship between the monetary variables and output, then these variables contribute to the trend shifts in output, and long-run neutrality is violated.[11]

Boschen and Mills (1995) use quarterly U.S. data from 1951:4 to 1990:4. They include variables describing productivity, real oil prices, weighted foreign real GDP of major U.S. trading partners, real government purchases, taxes, labor supply, the M1 money stock, the M2 money stock, and the nominal three-month Treasury bill rate. They use augmented Dickey-Fuller tests as diagnostics for the presence of nonstationarity in these data; they found (sometimes weak) evidence of a unit root in all the series. They test for cointegrating relationships among the blocks of real and nominal variables, and then between the nominal and real variables, as a means of testing for long-run monetary neutrality.

[11] For a related approach and analysis, see Hoffman and Rasche (1996), Chapter 7.

Based on these tests, Boschen and Mills conclude that long-run monetary neutrality holds during the postwar period in the U.S. This result confirms the findings that Fisher and Seater (1993), Boschen and Otrok (1994), and King and Watson (1997) reported regarding long-run monetary neutrality in the postwar U.S. data. This result also provides the best available evidence that omitted variable bias did not contaminate the previous results on this question.

Recent Tests Using International Data

So far, we have results that conform to the suggestions of King and Watson and Fisher and Seater only for U.S. data—certainly a natural place to start but not the true extent of the available evidence. As a first effort at generalization, Weber (1994) explicitly set out to apply the King and Watson testing procedures to G7 economies: Canada, France, Germany, Italy, Japan, the United Kingdom, and the United States. The data is quarterly, from the postwar era, but the particular years vary across countries.

Weber begins with a battery of unit-root diagnostics—using a much more elaborate procedure than the papers discussed so far—in an effort to make careful statements about the evidence for the presence of a unit root in the time series. For each country, he uses several different measures of the money stock, in part to confront the question of whether the results are sensitive to how money is defined. The combination of several diagnostic tests and many different time series produces a plethora of results that are not all the same. As a general rule, however, narrower monetary aggregates tended to be $I(1)$, while broader aggregates tended to be $I(2)$. Strictly speaking, according to the methodology outlined above, if money is $I(2)$ then superneutrality can be tested, whereas neutrality cannot. In response to this situation, Weber takes two approaches: In some cases, he performs neutrality tests anyway and warns the reader to interpret the results with caution, while in other

cases he uses the $I(1)$ aggregates to test for neutrality and the $I(2)$ aggregates to test for superneutrality. The remaining series on output, inflation, interest rates, and unemployment rates, for the most part, can be reasonably interpreted as $I(1)$.

Weber then turns to tests of long-run monetary neutrality in these countries, using the same wide variety of possible identifying assumptions that King and Watson used. His general finding is that for broader monetary aggregates, such as M2 or M3, a wide variety of identifying restrictions are consistent with (fail to reject) long-run monetary neutrality in the G7 economies during the postwar era. For narrower measures of money, the range of identifying restrictions consistent with long-run monetary neutrality is much smaller. Confidence ellipses for λ_{ym} and λ_{my} under the identifying assumption that money is long-run neutral, $\gamma_{ym}=0$, include the plausible region of the space where $\lambda_{ym}<0$, (the short-run impact of money on output is positive) and $\lambda_{my}>0$ (money reacts countercyclically to output in the short run). This is true across the G7 economies for both narrow and broad measures of money.

Superneutrality is examined using a bivariate VAR in differenced money growth and differenced real output. In general, long-run superneutrality with respect to the level of real output is rejected for a wide variety of identifying restrictions across the G7 economies.

Considering the question of whether the long-run Phillips curve is vertical, Weber proceeds using the changes in inflation and unemployment in his bivariate VAR. For six of the seven G7 economies, the hypothesis that $\gamma_{u\pi}=0$ cannot be rejected except in cases of rather extreme identifying assumptions. The exception is Italy, where this hypothesis can be rejected readily. Weber also considers a "reverse" hypothesis, with causality running from unemployment to inflation. In this case the hypothesis $\gamma_{\pi u}=0$ can be rejected easily across the non-Italian economies. For Italy this hypothesis is rejected only for extreme identification schemes. Weber

Figure 3

**Long-Run Response of the Level of Output
To a Permanent Increase in Inflation**
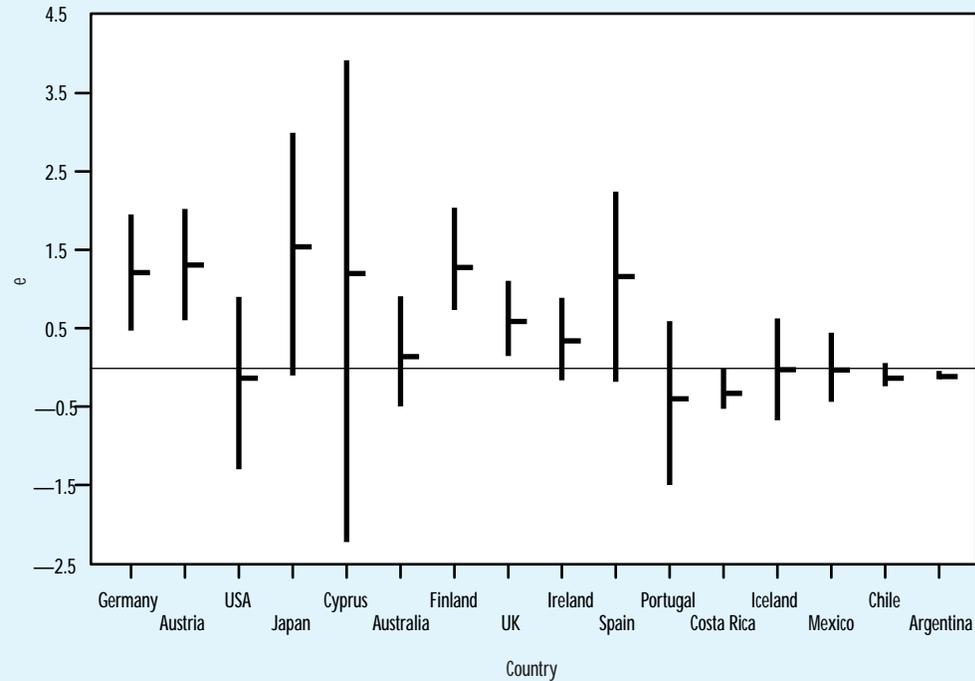
Countries ordered from lowest to highest average inflation in sample.
Horizontal lines represent point estimates. Vertical lines represent 90 percent confidence bounds.

The point estimate of the long-run response of the level of real output to a permanent inflation shock is generally positive for the low inflation countries but zero or negative for high inflation countries. This figure is reproduced from Bullard and Keating (1995) and is reprinted with permission by the *Journal of Monetary Economics*.

speculates that wage indexation in Italy during much of this period accounts for the differences between Italy and the other industrialized economies.

Finally, Weber goes on to test the Fisher relation for the G7 economies using a bivariate VAR in differenced inflation and differenced nominal interest rates. He finds that for Germany, a Fisher relation can be rejected for a wide variety of identifying restrictions, although some important benchmark restrictions do not lead to rejection. For the United States, Weber confirms the findings of King and Watson (1997) that nominal interest rates do not adjust one-for-one with permanent shocks to inflation. Even stronger evidence in this direction is found for the United Kingdom. But for Japan, Canada, Italy, and France,

the evidence is much more favorable for a Fisher relation, $\gamma_{r\pi} = 1$, to hold. The general finding in the previous literature (see for instance, Lothian, 1985, for 20 OECD countries) is that nominal interest rates adjust less than one-for-one with inflation.

While Weber considered G7 economies, Bullard and Keating (1995) consider virtually all of the countries in the world where enough data existed to formulate tests of long-run neutrality propositions in the spirit of Fisher and Seater (1993) and King and Watson (1997). In working with a large number of countries, data availability and quality impinge significantly on the analysis. Accordingly, Bullard and Keating restrict attention to countries that produce at least moderately high quality data (according to published

assessments) and have at least 25 years of consecutive annual observations (quarterly data often is not available). This leaves them with 58 countries. Bullard and Keating focus their analysis on one particular version of a neutrality proposition: the effect of a permanent shock to inflation on the level of output. If money is long-run neutral—and the evidence reported in this survey suggests that this is a reasonable assumption—then this can be viewed as a test of monetary superneutrality with respect to the level of real output. Moreover, problems with the definition of money within and across countries are avoided.

Bullard and Keating also begin with a battery of unit-root diagnostic tests for the real output and GDP deflator time series they use. They divide countries into groups based on the results of these tests, according to whether a country can be characterized as having experienced permanent shocks to inflation or not, and similarly for the level of real output. Countries that experienced permanent shocks to inflation and also to the level of output are candidates for a test based on a bivariate VAR; there were 16 countries in this group, dubbed Group A. There were also nine Group B countries for which evidence of a unit root in inflation was found, but evidence of a unit root in output was lacking. A large number of countries, 31, showed no evidence of permanent shocks in the inflation series, and were put into Group C. The two remaining countries were special cases.

Bullard and Keating (1995) then ran a two-variable VAR for the Group A countries, in differenced inflation and differenced output. They committed to the long-run identifying restriction that money is long-run neutral, $\gamma_{\pi y}=0$ in the King and Watson (1997) notation, and did not attempt to search over alternative identifying restrictions. They used the techniques of Blanchard and Quah (1989) to decompose shocks into permanent and transitory components, and consequently they considered the impulse-response functions of reactions of the two variables to both permanent and transitory shocks.

The main results for the Group A countries are as follows: The long-run response of the level of output to a permanent inflation shock was positive and statistically significant for four countries, negative and statistically significant for one country, and not statistically different from zero for the remainder. The point estimate of this long-run response generally declined as the in-sample average inflation rate increased, as shown in Figure 3.

The Group B countries, which possess permanent inflation shocks but no permanent output shocks, provide *prima facie* evidence of superneutrality. The Group C countries are uninformative because they do not possess permanent inflation shocks. Altogether, the results appear to be consistent with superneutrality for most of the countries that are informative.

However, as Figure 3 indicates, and as is borne out by the associated impulse-response functions, low-inflation countries appear to react very differently to permanent inflation shocks than high-inflation countries.[12] In particular, for low-inflation countries the point estimate of the long-run response is generally positive, while for high-inflation countries it is zero or negative. This suggests that averaging results from low and high-inflation countries may be misleading.

Bullard and Keating also comment on the prospects for permanent inflation shocks to permanently alter *rates of growth* of real output in this sample. According to ADF and other diagnostic tests for unit roots, real output growth rates are stationary in nearly all countries that experienced permanent shocks to inflation. This is direct evidence for superneutrality with respect to output growth rates. This result does not seem like one that is likely to change with data sets or countries, since the stationarity of real output growth is likely to remain under most conceivable criteria.[13]

All of the Bullard and Keating data are for postwar economies. Serletis and Krause (1996) use the Backus and Kehoe (1992) data set, which includes more than 100 years of annual observations on real output, prices, and money for Australia, Canada,

[12] Other aspects of the impulse-response functions had natural interpretations according to conventional wisdom, and also were generally consistent across countries.

[13] This result conflicts with other evidence from the cross-country growth regression literature, such as Barro (1996).

Denmark, Germany, Italy, Japan, Norway, Sweden, the United Kingdom, and the United States.[14] They test for unit roots using the procedures of Zivot and Andrews (1992), and they conclude that money is reasonably described as $I(1)$ except in Germany and Japan where it is $I(0)$; these latter two countries are therefore uninformative on neutrality questions in this data set. Serletis and Krause (1996) find that output is $I(0)$ for Australia, Canada, Denmark, Italy, the United Kingdom, and the United States. These countries, therefore, provide direct evidence in favor of long-run neutrality with respect to output. Serletis and Krause use the Fisher and Seater (1993) regression test to produce estimates for the remaining money-price or money-output combinations. These results generally support a hypothesis of long-run monetary neutrality.

The same data set is used by Serletis and Koustas (1998), who apply the King and Watson methodology to study long-run neutrality and superneutrality issues over a range of plausible identifying restrictions. They use only the money and real output series, and apply a battery of tests to determine the integration properties of the data. Except for the money series for Italy, which is $I(2)$, they conclude that all series are $I(1)$ and hence provide a reasonable dataset with which to test long-run monetary neutrality (superneutrality for Italy).[15] The results state that it is generally difficult to reject long-run monetary neutrality in this dataset under plausible identifying restrictions. An exception is the United Kingdom, when the identifying restriction is that $0 \leq \lambda_{my} \leq .6$. Superneutrality of money with respect to real output in the Italian data can be rejected under plausible identifying restrictions.

The Serletis and Krause (1996) and Serletis and Koustas (1998) results may appear to impinge on the Fisher and Seater (1993) and Boschen and Otrok (1994) findings for the United States, namely that the results for long-run monetary neutrality in the United States over the last century depend critically on inclusion or removal of the Great Depression years from the sample.

Both Serletis and Krause (1996) and Serletis and Koustas (1998) fail to reject long-run neutrality even when this period is included (under a range of plausible identification schemes in the latter case). However, Serletis and Koustas (1998) in fact reject long-run neutrality under the Fisher and Seater (1993) identifying restriction ($\gamma_{my} = 0$), but they do not reject under other, possibly more plausible, identifying restrictions. Of course, differences in results could also be due to differences in the data sets employed. Similar comments can be made concerning the results of Olekalns (1996) using a near-century of Australian data.

Crosby and Otto (1999) move away from the money-inflation-output nexus discussed in many of the papers so far, in order to analyze the long-run connection between inflation and the capital stock using the methods of Fisher and Seater (1993) and King and Watson (1997). Crosby and Otto consider a bivariate VAR with inflation playing the role of the nominal variable, and the capital stock playing the role of the real variable. They use the long-run identifying restriction that shocks to the capital stock do not have permanent effects on the rate of inflation, which is similar to the long-run restriction sometimes employed in the papers discussed earlier. They construct an annual capital stock series for 64 countries using postwar data, with differing sample periods for different countries. Their unit-root diagnostics (ADF tests) indicate that 34 of these countries have both permanent shocks to inflation and to the capital stock. For these countries they test superneutrality with respect to the capital stock using their bivariate VAR. The Crosby and Otto estimates indicate that a permanent inflation shock has no statistically significant long-run impact on the capital stock for a large majority of the countries. Departures from this result are generally on the positive side, with a permanent inflation shock tending to raise the stock of capital in a country. Crosby and Otto argue that these results are robust to a number of changes in their analysis, including an alternative identifying restriction.

[14] Some data are missing, notably 1914-24 and 1939-49 for Germany and 1941-51 for Japan. Also missing are 1915-20 for Denmark, and 1940-45 for Norway.

[15] These results on orders of integration are somewhat different from those of the previous paragraph, even though the data set is the same, because Serletis and Koustas (1998) use different (and more standard) procedures to test for the presence of unit roots than Serletis and Krause (1996). In section four of their paper, Serletis and Koustas (1998) discuss the differences when the Zivot and Andrews (1992) methodology is used.

A real variable of interest to many economists is productivity. In an attempt to understand the long-run relationship between inflation and productivity, Sbordone and Kuttner (1994) devote a portion of their analysis to bivariate VAR methodology similar to that used by King and Watson (1997). They use data from the postwar United States, and they conclude that both series are reasonably characterized as $I(1)$. Sbordone and Kuttner use the King and Watson (1997) approach to identification, setting impact multipliers and long-run multipliers to various values in an effort to learn about the sensitivity of the results to alternative identification schemes. Under many of these schemes, the long-run impact of a permanent, positive shock to inflation on productivity is negative. If the identification scheme is the monetarist one—the long-run impact of a permanent shock to productivity on inflation is zero—then the estimated effect of a permanent inflation shock on productivity is negative but is not statistically different from zero.

Koustas and Serletis (1999) use the King and Watson (1997) methodology of searching across alternative identification schemes to study the Fisher relation between nominal interest rates and inflation rates. They employ data from the postwar period for 11 industrialized countries: Belgium, Canada, Denmark, France, Germany, Greece, Ireland, Japan, the Netherlands, the United Kingdom, and the United States. According to the authors' unit root diagnostic tests, all of these countries except two (Denmark and Japan) can reasonably be interpreted as possessing the nonstationarity in interest rates and inflation rates required to use the King and Watson techniques. The basic finding is that the long-run Fisher relation can be rejected across countries for a wide range of plausible identification assumptions. The authors also argue that taking tax effects into account accentuates this finding. The Koustas and Serletis results are more consistent across countries on this question than those of Weber (1994), who found more mixed results for a similar set of countries.

Rapach (1999a) is the first author to consider a trivariate VAR in this literature. His variables are the inflation rate, the nominal interest rate, and the level of real output. The data is from the postwar period for 14 industrialized (OECD) countries, where continuous observations on all three variables are available starting from the 1960s and extending to the mid-1990s.[16] Rapach (1999a) uses long-run identifying restrictions following Blanchard and Quah (1989); he needs three for the trivariate system. Rapach first extends the often-used monetarist restriction that permanent shocks to interest rates and output cannot have permanent effects on the inflation rate. Rapach's third restriction, also motivated by theoretical considerations, is that permanent shocks to output ("permanent technology shocks") leave the real interest rate unchanged. Since the long-run response of inflation to a permanent technology shock is already set to zero, this last restriction is accomplished by making the long-run response of the nominal interest rate to a permanent technology shock equal to zero.

Rapach uses unit-root diagnostic tests to conclude that the variables are reasonably described as $I(1)$ for these countries, and runs the trivariate VAR in an effort to estimate, primarily, the long-run responses of the level of real output to a permanent inflation shock, and of the real interest rate to a permanent inflation shock (the difference between two estimated long-run responses in this system). For all countries, the point estimates indicate that real interest rates fall in response to a permanent inflation shock. Moreover, these effects are generally statistically significant (or very close) at conventional significance levels. The point estimates also indicate that the response of the level of real output to a permanent inflation shock is positive for 11 of 14 countries, and four of these are statistically significant, or nearly so, at conventional significance levels. These latter results are generally consistent with the findings of Bullard and Keating for low inflation countries in a bivariate VAR framework. The results on real interest

[16] The countries are Australia, Austria, Belgium, Canada, Denmark, France, Ireland, Italy, Japan, the Netherlands, New Zealand, Sweden, the United Kingdom, and the United States.

rates are more strikingly in favor of nonsuperneutrality than the findings of Weber for G7 economies in bivariate systems. Weber searched over identification schemes while Rapach commits to a particular scheme, but Rapach studies interactions between three variables instead of two and analyzes more countries.

Related Methodology Using U.S. Data

Ahmed and Rogers (1996, 1998) work on empirical issues related to the long-run impact of permanent inflation shocks on real variables, but with methods somewhat different from those discussed earlier. In particular, Ahmed and Rogers construct a theoretical model economy and use this economy to motivate restrictions imposed in their empirical work. The model consists of an infinitely-lived representative agent who might hold money either because it enters the utility function or because of a cash-in-advance constraint. The technology for production of private sector real output is Cobb-Douglas, multiplied by a technology shift variable and also by a function of government size. Special cases of this framework (restrictions on theoretical parameters) deliver the standard results from the theoretical literature on monetary superneutrality surveyed by Orphanides and Solow (1990).

Ahmed and Rogers (1998) use annual U.S. data from 1889 to 1995 covering inflation, real output, real consumption expenditures, real investment, and the ratio of government spending to output. Much of the data is from Kendrick (1961). Based on diagnostic testing, Ahmed and Rogers conclude that a reasonable description of the data is that these series are $I(1)$, with the exception of the size of government, which they sometimes treat as $I(0)$. The authors then estimate cointegrating relationships for two specifications of the model. Based on these estimates, a permanent, positive shock to inflation is associated with a permanent drop in the consumption-output ratio and a permanent increase in the investment-output ratio.

These effects are large and statistically significant. The Fisher effect does not hold, as Ahmed and Rogers infer that real interest rates decline in the face of permanent, positive inflation shocks according to these estimates.

Ahmed and Rogers then turn to estimation of a VECM using the cointegrating relationships implied by their theoretical model, in an effort to find out what happens to the levels of the real variables following a permanent inflation shock. For two different specifications, the estimates indicate that a permanent shock to inflation increases the level of output, consumption, and investment. Ahmed and Rogers also consider variance decompositions and note that the inflation shock only accounts for a small fraction of the forecast error variance in consumption, investment, and output.[17] They interpret these results as follows: Permanent inflation shocks do not occur very often, but when they do, they have a significant impact on the economy. Accordingly, when looking at the data historically, one might reasonably abstract from inflation in building a model, but when contemplating significant changes in inflation rates, one should not assume the effects will be negligible.[18]

Bernanke and Mihov (1998a) test long-run monetary neutrality, and, like Ahmed and Rogers (1998), they depart from the methodology described in the main portion of this survey. In particular, Bernanke and Mihov use their own, larger VAR model of short-run monetary policy which is described in more detail in another paper (Bernanke and Mihov, 1998b) as a starting point. This model uses monthly data for the United States during the postwar era, and has the following variables: total bank reserves and nonborrowed reserves, both measured as deviations from a trend and the federal funds rate (collectively the *policy variables*); interpolated monthly real GDP and interpolated monthly GDP deflator inflation, an index of spot commodity prices and real balances (with money measured as M2). They use a semi-structural approach to derive identification restrictions based on relationships between the policy

[17] Rapach (1999a) also computes variance decompositions and concludes that inflation shocks do not explain a significant fraction of output forecast error variance.

[18] Ahmed and Rogers (1998) also consider subsamples. They find that the effects of inflation on real variables move in the same direction, but are much weaker, during the postwar period as opposed to the pre-WWI or the interwar period.

and nonpolicy variables, and among the policy variables, and they estimate the VAR. They do not examine the temporary-permanent dichotomy of the shocks to the variables in their system; the focus instead is on isolating an action that can reasonably be termed "a shock to monetary policy." Bernanke and Mihov's evidence in favor of long-run neutrality is based on the impulse-response functions of their estimated VAR: These functions show that the long-run (120-month) response of output to the policy shock is not significantly different from zero, although positive. At the same time, they find short-term impacts of the policy shock, such as a liquidity effect, that are in accord with conventional wisdom.

Bernanke and Mihov (1998a) then turn to an analysis of the robustness of their results, by considering alternative identification schemes in a manner analogous to the King and Watson (1997) methodology. They find that the evidence on long-run neutrality is in a sense stronger when one is willing to accept an identification that produces a smaller liquidity effect. They also find that imposing long-run neutrality as an identifying restriction does not imply that one can reject their specification. Bernanke and Mihov conclude that these findings inspire confidence in their VAR model of monetary policy, since it is consistent with both a liquidity effect and long-run monetary neutrality.

CONCLUSIONS

This survey has covered a fair amount of territory. To avoid confusion about what the results actually say, this section includes a summary of the main findings organized by the nature of the proposition.

Long-Run Monetary Neutrality. In this survey, we did not find much evidence against the long-run neutrality of money. Fisher and Seater (1993) usefully reinterpreted some of the major time series studies on neutrality published in the 1970s and 1980s as consistent with long-run monetary neutrality, and uninformative regarding long-run monetary superneutrality

(Geweke's 1986 title notwithstanding!), once one takes proper account of the time series properties of the data. While Fisher and Seater (1993) found evidence against long-run neutrality with respect to real output for the United States during the last century, Boschen and Otrok (1994) pointed out that such a result did not hold once the Great Depression years were excluded from the sample. In another comment on this question, Haug and Lucas (1997) could not reject long-run neutrality in a century of Canadian data. Olekalns (1996) did find some evidence against long-run neutrality in a near-century of Australian data using a broad money measure, but the neutrality hypothesis could not be rejected using a narrower money measure. Coe and Nason (1999) find that long-run neutrality cannot be rejected for a century of U.S. data when the monetary base is the monetary variable, nor could they reject long-run neutrality for a century of U.K. data.[19]

Long-run neutrality received support from the studies focused exclusively on the postwar U.S. data. King and Watson (1997) searched over a wide range of identification schemes and found little evidence against long-run neutrality. Boschen and Mills (1995), studying a larger system of variables with cointegration techniques, but without the extensive robustness checking, also found little reason to doubt long-run neutrality. Bernanke and Mihov (1998a) argue that their model is consistent with long-run neutrality using the postwar U.S. data; like King and Watson (1997), they explore the robustness of their findings to an extensive range of alternative identification schemes.

The studies that used data from more than one country also found general support for the long-run monetary neutrality proposition. For instance, Weber (1994), using techniques similar to those of King and Watson (1997), generally supports long-run neutrality for the G7 economies during the postwar period across a wide variety of identification schemes, especially when money is measured using broader monetary aggregates. Weber's results also confirm the findings of King and Watson

[19] Coe and Nason also raise important questions concerning the statistical properties of the Fisher and Seater long-horizon regression test.

(1997) and Boschen and Mills (1995) for the postwar U.S. data. Serletis and Krause (1996) use the Backus and Kehoe (1992) data set for 10 industrialized countries, including the United States and Australia, covering more than a century. They found general support for long-run monetary neutrality, even in the United States and Australia, where Fisher and Seater (1993) —for the United States—and Olekalns (1996)—for Australia—had cast doubt. In a more extensive study, Serletis and Koustas (1998), use the same data set and apply the King and Watson (1997) technology of searching over plausible identification schemes. Here, wide support for long-run monetary neutrality is found across industrialized countries and plausible identification schemes, even though the data set is a very long-time series of the type that sometimes displayed evidence against long-run monetary neutrality in previous studies.

Long-Run Monetary Superneutrality. This survey has also shown that the evidence in favor of long-run monetary superneutrality is far more mixed. This is perhaps not too surprising since, as was stressed in the introduction, it is a relatively simple matter to write down neoclassical, market clearing, rational expectations theories in which superneutrality does not hold. In addition, since inflation is generally regarded as a distortionary force in macroeconomic systems, we might reasonably expect real variables to be altered in the face of permanent shocks to money growth and inflation.

Analyzing postwar U.S. data, King and Watson (1997) find that rejection of long-run monetary superneutrality with respect to real output is possible for a range of identification schemes they consider reasonable. Weber (1994) confirms this result using similar methodology across the G7 economies during the postwar period. Bullard and Keating (1995) analyze data from a number of countries worldwide during the postwar period. They consider permanent inflation shocks and the subsequent reaction of the level of real output. The results generally support superneutrality, but Bullard and Keating note that in the low inflation countries (such as the G7), point estimates tend to be positive and are sometimes statistically significant. Serletis and Koustas (1998) reject long-run monetary superneutrality for Italy over the last century, in a bivariate system with money and real output, over a range of identifying restrictions. Crosby and Otto (1998) generally find that permanent inflation shocks have little or no permanent effect on the level of the capital stock in a large sample of countries during the postwar period. When they do find statistically significant effects, permanently higher inflation is associated with a permanently higher capital stock. In Rapach's (1999a) study of a trivariate VAR using postwar data from 14 OECD countries, permanent inflation shocks generally were associated with permanently higher levels of real output and, more strikingly, with permanently lower real interest rates. Ahmed and Rogers (1998), using methodology that departs somewhat from the other studies in the survey, consider a century of U.S. data and conclude that permanent inflation shocks have permanent, positive effects on important real variables, including output, consumption, and investment. They also stress that these shocks do not explain a large portion of the forecast error variance in the data.

While the overall evidence on these questions is mixed, considering only lower inflation countries leads to the conclusion that permanently higher money growth or inflation is associated with permanently higher output and permanently lower real interest rates. As Ahmed and Rogers (1998) stress, this result is inconsistent with many —almost all?—current quantitative business cycle models, which generally predict that permanently higher inflation permanently lowers consumption and output. There is little support for such a prediction in the studies surveyed here. This is an important empirical puzzle that stands as a challenge for future research.

Related Propositions. We have also seen in this survey a smattering of evidence on other, related neutrality propositions.

King and Watson (1994, 1997) analyze the slope of a long-run Phillips curve in the postwar U.S. data, and find that a vertical curve is a reasonable approximation. Weber (1994) reports generally similar findings for the G7 economies.

King and Watson (1997) also study Fisher relations between interest rates and inflation, and conclude that nominal interest rates do not adjust one-for-one with permanent inflation shocks under a wide range of plausible identification schemes. Weber (1994) finds more mixed results for the G7 economies, but evidence presented by Rapach (1999a) and Koustas and Serletis (1999) is squarely on the side of less than one-for-one adjustment across industrialized nations.

Bullard and Keating (1995) comment on the effect of permanent inflation shocks on long-run economic growth. Because growth rates are generally stationary according to diagnostic tests, and inflation rates often are not, the methodology of Fisher and Seater (1993) and King and Watson (1997) suggests that permanent inflation shocks have no permanent effect on economic growth. Ahmed and Rogers (1998) include a comment in a similar vein.[20]

Areas for Further Research. Canova (1994), commenting on Weber (1994), stressed that the nature of the methodology of Fisher and Seater (1993) and King and Watson (1997) —however correct it may be from a logical point of view—places heavy reliance on the existence of (and on the number of) unit roots in the time series being studied. Canova comments that these tests of neutrality propositions depend in an important way on getting the inference on the unit root correct, and yet, tests for unit roots are known to have low power. Most authors, including Weber (1994), are well aware of this issue, and many use a battery of tests for a unit root in a series or other measures in an effort to be conservative about their conclusions in this regard. But Canova (1994, p. 121) notes, nevertheless, that this procedure "... conditions the results of economic hypotheses on shaky statistical ground. ..."

One problem is in the nature of the unit-root diagnostic tests. Since time series characterized by a unit root have such different properties from stationary time series, the researcher is forced into a declaration of a unit root or not. Once this declaration is made, the researcher can proceed with further analysis. This brings to mind a possible role for fractional integration in testing neutrality propositions. In fact, this possibility has been explored recently in a study by Bae and Jensen (1998). More work in this area may be fruitful in the future.

Canova (1994) also comments that Weber's (1994) results are based on bivariate VAR systems, as are many others reported here. He worries that the results may not be the same when larger systems are explored. Some papers surveyed here have taken steps in that direction, including Boschen and Mills (1995), Ahmed and Rogers (1998), Rapach (1999a), and Bernanke and Mihov (1998a). These studies generally have supported results from the bivariate VARs. However, much more could be done in multivariate systems than has been completed to date.

Even without turning to multivariate systems, one notes that much of the work surveyed has focused on real output, and that less work has been done on the long-run bivariate relationship between money or inflation and other important real variables. One exception is Crosby and Otto (1999), who take a step in this direction, using the capital stock instead of real output as their primary real variable.[21] A good deal more could be done by simply investigating the long-run bivariate relationships more systematically for variables other than real output.

All of the analyses surveyed consider one country at a time when testing neutrality propositions. One would like to know if a panel approach, implemented for a group of similar countries like the G7, would produce results similar to the ones reported in the studies surveyed here, or if important interactions between the countries are being left out. A simpler line would be to study multivariate systems for a single country that attempt to account for cross-border effects by including an international

[20] This idea is also related to work by Jones (1995).

[21] For another exception see Rapach (1999b).

variable. This would seem to be particularly important for some smaller, open economies sometimes included in these studies.[22]

While problems certainly remain, it seems that the 1990s have been quite fruitful in this area of empirical macroeconomics. Tests of neutrality propositions not subject to the critique of Lucas (1972) and Sargent (1971)—tests that have eluded economists during much of the postwar era—have been devised and executed for a variety of times and places. This body of work gives us economists what is perhaps our first glimpse at the evidence on long-run monetary neutrality and superneutrality, and allows assessment of the merits of these propositions separate from the logical force of theoretical arguments.

REFERENCES

Ahmed, Shaghil, and John H. Rogers. "Long-Term Evidence on the Tobin and Fisher Effects: A New Approach," International Finance Discussion Paper #566, Board of Governors of the Federal Reserve System (September 1996).

_____, and _____ . "Inflation and the Great Ratios: Long-Term Evidence From the U.S.," International Finance Discussion Paper #628, Board of Governors of the Federal Reserve System (October 1998).

Andersen, Leonall, and Denis Karnovsky. "The Appropriate Time Frame For Controlling Monetary Aggregates: The St. Louis Evidence," in *Controlling Monetary Aggregates II: The Implementation*, Federal Reserve Bank of Boston, 1972.

Backus, David K., and Patrick J. Kehoe. "International Evidence on the Historical Properties of Business Cycles," *American Economic Review* (September 1992), pp. 864-88.

Bae, Sangkun, and Mark J. Jensen. "Long-Run Neutrality in a Long-Memory Model," working paper, University of Missouri at Columbia (December 1998).

Barro, Robert J. "Inflation and Growth," this *Review* (May/June 1996), pp. 153-69.

Bernanke, Ben S., and Ilian Mihov. "The Liquidity Effect and Long-Run Neutrality," National Bureau of Economic Research Working Paper #6608 (June 1998a).

_____, and _____ . "Measuring Monetary Policy," *Quarterly Journal of Economics* (August 1998b), pp. 869-902.

Blanchard, Olivier J., and Danny Quah. "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review* (September 1989), pp. 655-73.

Boschen, John F., and Leonard O. Mills. "Tests of Long-Run Neutrality Using Permanent Monetary and Real Shocks," *Journal of Monetary Economics* (February 1995), pp. 25-44.

_____, and Christopher M. Otrok. "Long-Run Neutrality and Superneutrality in an ARIMA Framework: Comment," *American Economic Review* (December 1994), pp. 1470-73.

Bullard, James, and John W. Keating. "The Long-Run Relationship Between Inflation and Output in Postwar Economies," *Journal of Monetary Economics* (December 1995), pp. 477-96.

Canova, F. "Testing Long-Run Neutrality: Empirical Evidence For G7 Countries With Special Emphasis on Germany," *Carnegie-Rochester Conference Series on Public Policy* (1994), pp. 119-25.

Coe, Patrick J., and James M. Nason. "Long-Run Monetary Neutrality in Three Samples: The United Kingdom, the United States, and the Small," manuscript, University of Calgary and University of British Columbia (May 1999).

Crosby, Mark, and Glen Otto. "Inflation and the Capital Stock," *Journal of Money, Credit, and Banking*, forthcoming (1999).

Fischer, Stanley. "Why are Central Banks Pursuing Long-Run Price Stability?" in *Achieving Price Stability*, Federal Reserve Bank of Kansas City Symposium, 1996.

Fisher, Mark E. "The Time Series Implications of the Long-Run Neutrality and Superneutrality of Money," Ph.D. Dissertation, University of Chicago (1988).

_____, and John J. Seater. "Neutralities of Money," manuscript, Board of Governors of the Federal Reserve System and North Carolina State University (1989).

_____, and _____ . "Long-Run Neutrality and Superneutrality in an ARIMA Framework," *American Economic Review* (June 1993), pp. 402-15.

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

_____, and _____ . *Monetary Trends in the United States and the United Kingdom*, University of Chicago Press, 1982.

Geweke, John F. "The Superneutrality of Money in the United States: An Interpretation of the Evidence," *Econometrica* (January 1986), pp. 1-21.

Haug, Alfred A., and Robert F. Lucas. "Long-Run Neutrality and Superneutrality in an ARIMA Framework: Comment," *American Economic Review* (September 1997), pp. 756-59.

Hoffman, Dennis L., and Robert H. Rasche. *Aggregate Money Demand Functions*, Kluner Academic Publishers, 1996.

Jefferson, Philip N. "On the Neutrality of Inside and Outside Money," *Economica* (November 1997), pp. 567-86.

Jones, Charles I. "Time Series Tests of Endogenous Growth Models," *Quarterly Journal of Economics* (May 1995), pp. 495-525.

Kendrick, J. *Productivity Trends in the United States*, Princeton University Press, 1961.

[22] Nason and Rogers (1999) use methodology related to that discussed in this survey to study investment and current account dynamics for Canada.

King, Robert G., and Mark W. Watson. "Testing Long-Run Neutrality," working paper #4156, National Bureau of Economic Research (September 1992).

_____, and _____. "The Postwar U.S. Phillips Curve: A Revisionist Econometric History," *Carnegie-Rochester Conference Series on Public Policy* (1994), pp. 157-219.

_____, and _____. "Testing Long-Run Neutrality," Federal Reserve Bank of Richmond *Economic Quarterly* (Summer 1997), pp. 69-101.

Kormendi, Roger C., and Phillp G. Meguire. "Cross-Regime Evidence of Macroeconomic Rationality," *Journal of Political Economy* (October 1984), pp. 875-908.

Koustas, Zisimos. "Canadian Evidence on Long-Run Neutrality Propositions," *Journal of Macroeconomics* (Spring 1998), pp. 397-411.

_____, and Apostolos Serletis. "On the Fisher Effect," *Journal of Monetary Economics* (August 1999), pp. 105-30.

Lothian, James R. "Equilibrium Relationships Between Money and Other Economic Variables," *American Economic Review* (September 1985), pp. 828-35.

Lucas, Robert E., Jr. "Econometric Testing of the Natural Rate Hypothesis," in *The Econometrics of Price Determination,* Board of Governors of the Federal Reserve System, 1972.

_____. "Two Illustrations of the Quantity Theory of Money," *American Economic Review* (December 1980), pp. 1005-14.

_____. "Nobel Lecture: Monetary Neutrality," *Journal of Political Economy* (August 1996), pp. 661-82.

Marty, Alvin L. "What is the Neutrality of Money?" *Economics Letters* (April 1994), pp. 407-09.

McCandless, George T., and Warren E. Weber. "Some Monetary Facts," Federal Reserve Bank of Minneapolis *Quarterly Review* (Summer 1995), pp. 2-11.

Nason, James M., and John H. Rogers. "Investment and the Current Account in the Short Run and the Long Run," manuscript, University of British Columbia (September 1999).

Nelson, Charles R., and Charles I. Plosser. "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications," *Journal of Monetary Economics* (September 1982), pp. 139-62.

Olekalns, Nilss. "Some Further Evidence on the Long-Run Neutrality of Money," *Economics Letters* (March 1996), pp. 393-98.

Orphanides, Athanasios, and Robert M. Solow. "Money, Inflation and Growth," in *Handbook of Monetary Economics,* North-Holland, 1990.

Rapach, David E. "International Evidence on the Long-Run Superneutrality of Money," Working Paper, Trinity College (1999a).

_____. "Monetary Shocks and Real Exchange Rate Hysteresis: Evidence From the G-7 Countries," *Review of International Economics,* (April 1999b).

Sargent, Thomas J. "A Note on the 'Accelerationist' Controversy," *Journal of Money, Credit, and Banking* (August 1971), pp. 721-25.

_____. "The End of Four Big Inflations," in *Rational Expectations and Inflation,* Harper and Row, 1986.

Sbordone, Argia M., and Kenneth Kuttner. "Does Inflation Reduce Productivity?" Federal Reserve Bank of Chicago *Economic Perspectives* (November/December 1994), pp. 2-14.

Serletis, Apostolos, and Zisimos Koustas. "International Evidence on the Neutrality of Money," *Journal of Money, Credit, and Banking* (February 1998), pp. 1-25.

_____, and David P. Krause. "Empirical Evidence on the Long-Run Neutrality Hypothesis Using Low-Frequency International Data," *Economics Letters* (March 1996), 323-27.

Weber, A. "Testing Long-Run Neutrality: Empirical Evidence For G7 Countries With Special Emphasis on Germany," *Carnegie-Rochester Conference Series on Public Policy* (1994), pp. 67-117.

Zivot, Eric, and Donald W. K. Andrews. "Further Evidence on the Great Crash, the Oil Price Shock and the Unit Root Hypothesis," *Journal of Business and Economic Statistics* (July 1992), pp. 251-70.