

Carl F. Christ

Carl F. Christ is a professor of economics at The Johns Hopkins University. Helpful comments on an earlier draft were made by Jonathan Ahlbrecht, Stephen Blough, Pedro de Lima and William Zame. Any remaining shortcomings are my responsibility.

Assessing Applied Econometric Results

IT IS A GREAT HONOR to be asked to participate in this conference to celebrate the work of Ted Balbach, who has long upheld the standard of relevant, independent, intelligible economic studies at the Federal Reserve Bank of St. Louis.

My invitation to this conference asked for a philosophical paper about good econometric practice. I have organized my views as follows. Part I of the paper defines the concept of an ideal econometric model and argues that to tell whether a model is ideal, we must test it against new data—data that were not available when the model was formulated. Such testing suggests that econometric models are not ideal, but are approximations to a changing reality. Part I closes with a list of desirable properties that we can realistically seek in econometric models. Part II is a loosely connected set of comments and criticisms about several econometric techniques. Part III discusses methods of evaluating econometric models by means of their forecasts and summarizes some results of such evaluations, as proposed in part I. Part IV resurrects an old, plain-vanilla equation relating monetary velocity to an interest rate and tests it with more recent data. The rather remarkable result is that it still does about as well today as it did nearly 40 years ago. Part V is a brief conclusion.

HOW TO RECOGNIZE AN IDEAL MODEL IF YOU MEET ONE

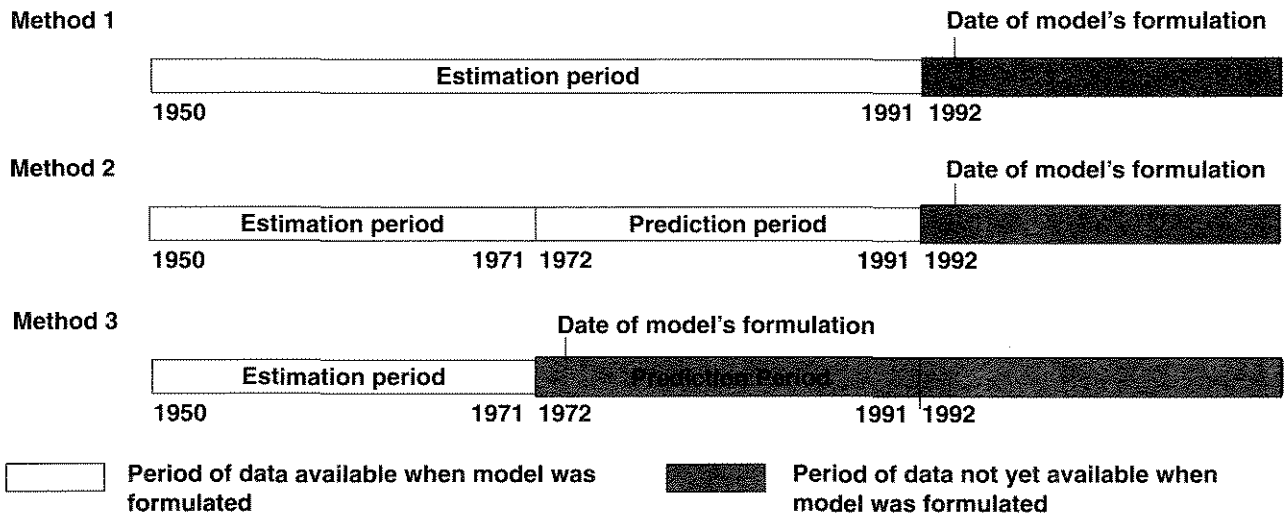
The Goal of Research and the Concept of an Ideal Model

The goal of economic research is to improve knowledge and understanding of the economy, either for their own sake, or for practical use. We want to know how to control what is controllable, how to adapt to what is uncontrollable, and how to tell which is which. The goal of economic research is analogous to the prayer of Alcoholics Anonymous (I do not suggest that economics is exactly like alcoholism)—“God grant me the serenity to accept the things I cannot change; the courage to change the things I can; and the wisdom to know the difference.”

The goal of applied econometrics is *quantitative* knowledge expressed in the form of mathematical equations.

I invite you to think of an ideal econometric model, by which I mean a set of equations, complete or incomplete, with numerically estimated parameters, that describes some interesting set of past data, closely but not perfectly, and that

Figure 1
Three Methods of Formulating and Estimating a Model and Checking Its Correspondence with 1950-1991 Data



will continue to describe all future data of that type.

The Need for Testing Against New Data

How can we tell whether we have found an ideal econometric model? We can certainly tell how well a model describes a given set of *past* data. (We will discuss what is meant by a good description later). Suppose we have a model in 1992, with estimated parameters, that closely describes past data for 1950–91. To tell whether it is the ideal model we seek, we must try it with future data. Suppose that after three years we try the model with data for 1992–94, and it describes them closely also. Still, in 1995 all we will be sure of is that it describes data closely for a past period, this time from 1950 through 1994. In principle we can never be sure we have found an ideal model because there will always be more future data to come, so we will never be able to say that a model is ideal. The longer the string of future data that a model describes closely, however, the more confidence we have in it.

Is this only a matter of the *amount* of data that the model describes, or is there something else involved? I argue that something else is involved.

Suppose again that in 1992 we have a model that closely describes an interesting data set for the past period 1950–91. Consider the following three methods, shown in figure 1, by which this model might have been obtained and by which its ability to describe data for 1950 through 1991 might have been assessed:

1. It was formulated in 1992, and fitted to data for the *entire* period 1950–91.
2. It was formulated in 1992, fitted to data for the *sub-period* 1950–71, and used to predict data from 1972 through 1991.
3. It was formulated in 1972, fitted to data for the *sub-period* 1950–71, and used to predict data from 1972 through 1991.

Methods 1 and 2 differ in that method 1 fits the model to *all* the available data, whereas method 2 fits it to the first part only and uses the result to predict the second part, from 1972 onward. 1972 is not a randomly chosen date. It was the year before the first oil crisis. Method 3 differs in that the model builder did not yet know about the oil crisis when formulating the model.

Now consider the following question: Given the goodness of fit of this model to data for the whole period 1950–91, does your confidence in the model depend on which of these three methods was

used to obtain it? I argue that it should. In particular, I argue that an equation obtained by a method similar to method 3, which involves testing against data that were not available to the model builder when the model was formulated, deserves more confidence than the same equation obtained by either of the other two methods.

The argument has to do with the goal of an econometric model—to describe not only past data, but also future data. It is easy to formulate a model that can describe a given set of past data perfectly but cannot describe future observations at all. Of course, such a research strategy should be avoided.

Here is a simple example. Imagine a pair of variables whose relationship we want to describe. Suppose we have two observations on the pair of variables. Then a line, whose equation is linear, will fit the data perfectly. Now suppose we obtain a third observation. It will almost certainly not lie on the line determined by the first two observations. But a parabola, whose equation is quadratic (of degree 2), will fit the three observations perfectly. Now suppose a fourth observation becomes available. It will almost certainly not lie on the parabola. But a sort of S-curve, whose equation is cubic (of degree 3), will fit the four observations perfectly. And so on. In general, a polynomial equation of degree n will fit a set of $n + 1$ observations on two variables perfectly, but a polynomial of higher degree will be required if the number of observations is increased. Methods of this type can describe any set of past data perfectly but almost certainly cannot describe any future data.

If a model is to describe future data, it needs to capture the enduring systematic features of the phenomena that are being modeled and it should avoid conforming to accidental features that will not endure. The trouble with the exact-fitting polynomial approach just discussed is that it does not try to distinguish between the enduring systematic and the temporary accidental features of reality. In the process of fitting past data perfectly, this approach neglects to fit enduring systematic features even approximately.

This relates to the choice among methods 1, 2 and 3 for finding a model that describes a body of data. When formulating a model, researchers typically pay attention to the behavior of available data, which perforce are past data. One tries

different equation forms and different variables to see which formulation best describes the data. This process has been called data mining. As a method of formulating tentative hypotheses, data mining is fine. But it involves the risk of being too clever, of fitting the available data too well and hence of choosing a hypothesis that conforms too much to the temporary accidental and too little to the enduring systematic features of the observed data. In this respect it is similar to the exact-fitting polynomial approach described earlier, though not as bad.

The best protection against having done too good a job of making a model describe past data is to test the model against new data that were not available when the model was formulated. This is what method 3 does, and that is why a model obtained by method 3 merits more confidence, other things equal.

Trygve Haavelmo once said to me, not entirely in jest, that what we economists should do is formulate our models, then go fishing for 50 years and let new data accumulate, and finally come back and confront our models with the new data.

Wesley Mitchell put the matter very well when he wrote the following:¹

The proposition may be ventured that a competent statistician, with sufficient clerical assistance and time at his command, can take almost any pair of time series for a given period and work them into forms which will yield coefficients of correlation exceeding $\pm .9$. It has long been known that a mathematician can fit a curve to any time series which will pass through every point of the data. Performances of the latter sort have no significance, however, unless the mathematically computed curve continues to agree with the data when projected beyond the period for which it is fitted. So work of the sort which Mr. Karsten and Professor Fisher have shown how to do must be judged, not by the coefficients of correlation obtained within the periods for which they have manipulated the data, but by the coefficients which they get in earlier or later periods to which their formulas may be applied.

Milton Friedman, in his review of Jan Tinbergen's pioneering model of the U.S. economy, referred to Mitchell's comment and expressed a similar idea somewhat differently:²

Tinbergen's results cannot be judged by ordinary tests of statistical significance. The reason is that

¹See Mitchell (1927).

²See Friedman (1940) and Tinbergen (1939).

the variables with which he winds up, the particular series measuring these variables, the leads and lags, and various other aspects of the equations besides the particular values of the parameters (which alone can be tested by the usual statistical technique) have been selected after an extensive process of trial and error *because* they yield high coefficients of correlation. Tinbergen is seldom satisfied with a correlation coefficient less than .98. But these attractive correlation coefficients create no presumption that the relationships they describe will hold in the future. The multiple regression equations which yield them are simply tautological reformulations of *selected* economic data. Taken at face value, Tinbergen's work "explains" the errors in his data no less than their real movements.

That last statement can be strengthened. Tinbergen's method, which has been the method of most model builders ever since, explains whatever temporary accidental components there may be in the data (regardless of whether they are measurement errors), as well as the enduring components.

Most macroeconomic models formulated before the 1973 oil crisis had no variables representing the prices and quantities of oil and energy. Most of these models were surprised by the oil crisis and its aftermath, and most of them made substantial forecast errors thereafter. Many models formulated after 1973 pay special attention to oil and energy. Of course many of those models provide better explanations of the post-oil-crisis data than do models that ignore oil and energy. But my point is different. A model that was formulated after the oil crisis was specifically designed to conform to data during and after the crisis, and if there are temporary accidental variations, the model will conform to them just as much as to the systematic variations. Hence the task of explaining data between the onset of the 1972 oil crisis and 1992 is easier for a model that was formulated in 1992 than for a model that was formulated before the crisis. Therefore if both models do equally well at describing data from 1950 to 1991, the one formulated before the crisis has passed a stricter test and merits more confidence.

What about the relative merits of methods 1 and 2? Sometimes method 2 is recommended; that is, it is recommended that researchers estimate a model using only the earlier part of the available data and use the later part as a test of the model's forecasting ability. When thinking about this proposal, consider a model that has

been formulated with access to all of the data. It does not make much difference whether part of the data is excluded from the estimation process and used as a test of that model, as in method 2, or whether it is included, as in method 1. Either way, we draw the same conclusions. If the model with a set of constant coefficients describes both parts of the data well, method 1 will yield a good fit for the whole period and method 2 will yield a good fit for the estimation period and small errors for the forecast period. If the model with a set of constant coefficients does not describe both parts of the data well, in method 1 the residuals, if examined carefully, will reveal the flaws, and in method 2 the residuals, the forecast errors or both will reveal the flaws. And with both methods 1 and 2 we have a risk that the model was formulated to conform too much to the temporary accidental features of the available data.

One noteworthy difference between methods 1 and 2 is that if the model's specification is correct, method 1 will yield more accurate estimates of the parameters because it uses a larger sample and thus has a smaller sampling error.

Econometric Models Are Approximations

When I began work in econometrics, I believed a premise that underlies much econometric work—namely, that a true model that governs the behavior of the economy actually exists, with both systematic and random components and with true parameter values. And I believed that ultimately it would be possible to discover that true model and estimate its parameter values. My hope was first to find several models that could tentatively be accepted as ideal and eventually to find more general models that would include particular ideal models as special cases. (One way to top your colleagues is to show that their models are special cases of yours. Nowadays this is called "encompassing.")

Experience suggests that we cannot expect to find ideal models of the sort just described. When an estimated econometric model that describes past data is extrapolated into the future for more than a year or two, it typically does not hold up well. To try to understand how this might happen, let us temporarily adopt the premise that there is a true model. Of course, we do not know the form or parameters of this true model. They may or may not be changing, but if they are changing according to some rule, then in principle it is

possible to incorporate that rule into a more general unchanging true model.

Suppose that an economist has specified a model, which may or may not be the same as the true model. If the form and parameters of the economist's model are changing according to some rule (not necessarily the same as the rule governing the true model), again in principle it is possible to incorporate that rule into a more general unchanging model.

Now consider the following possible ways in which the economist's model might describe past data quite well but fail to describe future data:

1. The form and parameter values of the economist's model may be correct for both the past period and the future period, but as the forecast horizon is lengthened, the forecasts get worse because the variance of the forecast is an increasing function of the length of the horizon. This will be discussed later.
2. The form of the economist's model may be correct for both the past period and the future period, but some or all of the true parameters may change during the future period.
3. The form of the economist's model may be correct for the past period but not for the future period because of a change in the form of the true model that is not matched in the economist's model.
4. The form of the economist's model may be incorrect for both periods but more nearly correct for the past period.

The last possibility is the most likely of the four in view of the fact that the economy has millions of different goods and services produced and consumed by millions of individuals, each with distinct character traits, desires, knowledge and beliefs.

These considerations lead to the conjecture that the aforementioned premise underlying econometrics is wrong—that there is no unchanging true model with true parameter values that governs the behavior of the economy now and in the future. Instead, every estimated econometric model is at best an approximation of a changing economy—an approximation that becomes worse as it is applied to events that occur further into the future from the period in which the model was formulated. In this case we should not be surprised at our failure to find an ideal general model as defined earlier. Instead, we should be

content with models that have at best only a temporary and approximate validity that deteriorates with time. We should sometimes also be content with models that describe only a restricted range of events—for example, events in a particular country, industry or population group.

Desiderata for an Econometric Model

If no ideal model exists, what characteristics can we realistically strive for in econometric models regarded as scientific hypotheses? The following set of desiderata are within reach:

1. The estimated model should provide a good description of some interesting set of past data. This means it should have small residuals relative to the variation of its variables—that is, high correlation coefficients. The standard errors of its parameter estimates should be small relative to those estimates, that is, its t-ratios should be large. If it is estimated for separate subsets of the available data, all those estimates should agree with each other. Finally, its residuals should appear random. (If the residuals appear to behave systematically, it is desirable to try to find variables to explain them.)
2. The model should be testable against data that were not used to estimate it and against data that were not available when it was specified.
3. The estimated model should be able to describe events occurring after it was formulated and estimated, at least for a few quarters or years.
4. The model should make sense in the light of our knowledge of the economy. This means in part that it should not generate negative values for variables that must be non-negative (such as interest rates) and that it should be consistent with theoretical propositions about the economy that we think are correct.
5. Other things equal, a simple model is preferable to a complex one.
6. Other things equal, a model that explains a wide variety of data is preferable to one that explains only a narrow range of data.
7. Other things equal, a model that incorporates other useful models as special cases is preferable to one that does not. (This is almost the same point as the previous one.)

In offering these desiderata, I assume that the purpose of a model is to state a hypothesis that describes an interesting set of available data and that may possibly describe new data as well. Of course, if the purpose is to test a theory that we are not sure about, the model should be constructed in such a way that estimates of its parameters will tell us something about the validity of that theory. The failure of such a model to satisfy these desiderata may tell us that the theory it embodies is false. This too is useful knowledge.

COMMENTS AND CRITICISMS ABOUT ECONOMETRIC TECHNIQUES

Theory vs. Empiricism

Two general approaches to formulating a model exist. One is to consult economic theory. The other is to look for regularities in the data. Either can be used as a starting point, but a combination of both is best. A model derived from elegant economic theory may be appealing, but unless at least some of its components or implications are consistent with real data, it is not a reliable hypothesis. A model obtained by pure data mining may be consistent with the body of data that was mined to get it, but it is not a reliable hypothesis if it is not consistent with at least some other data (recall what was said about this earlier), and it will not be understood if no theory to explain it exists.

The VAR Approach

Vector autoregression (VAR) is one way of looking for regularities in data. In VAR, a set of observable variables is chosen, a maximum lag length is chosen, and the current value of each variable is regressed on the lagged values of that variable and all other variables. No exogenous variables exist; all observable variables are treated as endogenous. Except for that, a VAR model is similar to the unrestricted reduced form of a conventional econometric model. Each equation contains only one current endogenous variable, each equation is just identified, and no use is made of any possible theoretical information about possible simultaneous structural equations that might contain more than one current endogenous variable. In fact, no use is made of any theoretical information at all, except in the choice of the list of variables to be included and the length of the lags. In macroeconomics it is

not practical to use many variables and lags in a VAR because the number of coefficients to be estimated in each equation is the product of the number of variables times the number of lags and because one cannot estimate an equation that has more coefficients than there are observations in the sample.

The ARIMA Approach

The Box-Jenkins type of time-series analysis is another way to seek regularities in data. Here each observable variable is expressed in terms of purely random disturbances. This can be done with one variable at a time or in a multivariate fashion. In the univariate case an expression involving current and lagged values of an observable variable is equated to an expression involving current and lagged values of an unobservable white-noise disturbance; that is, a serially independent random disturbance that has a mean of zero and constant variance. Such a formulation is called an autoregressive integrated moving average (ARIMA) process. The *autoregressive* part expresses the current value of the variable as a function of its lagged values. The *integrated* part refers to the possibility that the first (or higher-order) differences of the variable, rather than its levels, may be governed by the equation. Then the variable's levels can be obtained from its differences by undoing the differencing operation—that is, by integrating first differences once, integrating second differences twice, and so on. (If no integration is involved, the process is called ARMA instead of ARIMA.) The *moving average* part expresses the equation's disturbance as a moving average of current and lagged values of a white-noise disturbance. To express a variable in ARIMA form, it is necessary to choose three integers to characterize the process. One gives the order of the autoregression (that is, the number of lags to be included for the observable variable); one gives the order of the moving average (that is, the number of lags included for the white-noise disturbance); and one gives the order of integration (that is, the number of times the highest-order differences of the observable variable must be integrated to obtain its levels). The choice of the three integers (some of which may be zero) is made by examining the time series of data for the observable variable to see what choice best conforms to the data. After that choice has been made, the coefficients in the autoregression and moving average are estimated. The multivariate form of ARIMA modeling is a generalization of the

univariate form. And, of course, VAR modeling is a special case of multivariate ARIMA modeling.

VAR and ARIMA models can be useful if they lead to the discovery of regularities in the data. If enduring regularities in the data are discovered, we have something interesting to try to understand and explain. In my view, however, one disadvantage of both approaches is that they make almost no use of any knowledge of the subject matter being dealt with. To use univariate ARIMA on an economic variable, one need know nothing about economics. I think of univariate ARIMA as mindless data mining. To use multivariate ARIMA, one need only make a list of variables to be included and choose the required three integers. To use VAR, one need only make a list of the variables to be included and choose a maximum lag length. Knowledge of the subject the equations deal with can enter into the choice of variables to be included.

It may seem that the ARIMA approach and the conventional econometric model approach are antithetical and inconsistent with each other. Zellner and Palm (1974), however, have pointed out that if a conventional model's exogenous variables are generated by an ARIMA process, the model's endogenous variables are generated the same way.

General-to-Specific Modeling

General-to-specific modeling starts with an estimated equation that contains many variables and many lagged values of each. Its approach is to pare this general form down to a more specific form by omitting lags and variables that do not contribute to the explanatory power of the equation. Much can be said for this technique, but of course it will not lead to a correct result if the general form one starts with does not contain the variables and the lags that belong in an equation that is approximately correct.

The Error Correction Mechanism

The error correction mechanism (ECM) provides a way of expressing the rate at which a variable moves toward its desired or equilibrium value when it is away from that value. Economic theory is at its best when deriving desired or equilibrium values of variables, either static positions or dynamic paths. ECM has so far not been good at deriving the path followed by an economy that is out of equilibrium. Error correction models are appealing because they permit the nature of the equilibrium to be specified with the aid of the-

ory but permit the adjustment path to be determined largely by data.

Testing Residuals for Randomness

I have already discussed testing residuals for randomness. If an equation's residuals appear to follow any regular or systematic pattern, this is a signal that there may be some regular or systematic factor that has not been captured by the form and variables chosen for the equation. In such a case it is desirable to try to modify the equation's specification, either by including additional variables, by changing the form of the equation, or both, until the residuals lose their regular or systematic character and appear to be random.

Stationarity

It is often said that the residual of a properly specified equation should be stationary, that is, that its mean, variance and autocovariances should be constant through time. However, for an equation whose variables are growing over time, such as an aggregate consumption or money-demand equation, it would be unreasonable to expect the variance of the residual to be constant. That would mean that the correlation coefficients for the equation in successive decades (or other time intervals) would approach one. It would be more reasonable to expect the standard deviation of the residual to grow roughly in proportion to the dependent variable, to one of the independent variables, or to some combination of them.

The Lucas Critique

Robert Lucas (1976) warned that when an estimated econometric model is used to predict the effects of changes in government policy variables, the estimated coefficients may turn out wrong and hence the predictions may also turn out wrong. Under what conditions can this be expected to occur? Lucas says that this occurs when policymakers follow one policy rule during the estimation period and begin to follow a different policy rule during the prediction period. The reason for this, he argues, is that in many cases the parameters that were estimated are not constants that represent invariant economic relationships, but instead are variables that change in response to changes in policy rules. This is because they depend both on constant parameters and on varying expectations that private agents formulate by observing policymakers and trying to discover what policy rule is being followed. Jacob Marschak (1953) foreshadowed this idea when he cautioned that

predictions made from an estimated econometric model will not be valid if the structure of the model (that is, its mathematical form and its parameter values) changes between the estimation period and the prediction period. Therefore, to make successful predictions after a structural change, one must discover the nature of the structural change and allow for it.

I take this warning seriously. It need not concern us when policy variations whose effects we want to predict are similar to variations that occurred during the estimation period. But when a change in the policy rule occurs, private agents will eventually discover that their previous expectation formation process is no longer valid and will adopt a new one as quickly as they can. As they do so, some of the estimated parameters will change and make the previously obtained estimates unreliable.

Goodhart's Law

Lucas' warning is related to Goodhart's Law, which states that as soon as policymakers begin to act as if some previously observed relationship is reliable, it will no longer be reliable and will change.³ A striking example is the short-run, downward-sloping Phillips curve.

Are Policy Variables Exogenous?

Most econometric models treat at least some policy variables as exogenous. But public policy responds to events. Policy variables are not exogenous. The field of public choice studies the actions of policymakers, treating them as maximizers of their own utility subject to the constraints they face. Econometric model builders have so far not made much use of public choice economics.

BY THEIR FORECASTS YE SHALL KNOW THEM (MODELS, THAT IS)

Methods of Evaluating Models' Forecasts

A conventional econometric model contains disturbances and endogenous and exogenous variables. Typically some of the endogenous variables appear with a lag. Consider an annual model with data for all variables up to and including 1992.

Suppose that at the end of 1992 we wish to forecast the endogenous variables for 1993, one

year ahead. This is an *ex ante* forecast. For this we need estimates of the model's parameters, which can be computed from our available data. In addition, we need 1993 values for the lagged endogenous variables. These we already have because we have values for the years 1992 and earlier. Further, we need predicted 1993 values for the disturbances. We usually use zeros here because disturbances are assumed to be serially independent with zero means. (Some modelers, however, would use values related to the residuals for 1992 and possibly earlier years if the disturbances were thought to be serially correlated.) Finally, we need predicted 1993 values for the exogenous variables. These predictions must be obtained from some source outside the model.

Our predictions of the endogenous variables for 1993 will be conditional on our estimated model and on our predictions of the disturbances and exogenous variables. If we make errors in forecasting the endogenous variables, it may be because our estimated model is wrong, because our predictions of the disturbances or exogenous variables are wrong, or because of some combination of these.

It is possible—and desirable—to test the forecasting ability of an estimated model independently of the model user's ability to forecast exogenous variables. This is done with an *ex post* forecast. An *ex post* forecast for one period ahead, say for 1993, is made as follows: Wait until actual 1993 data for the exogenous variables are available, use them instead of predicted values of the exogenous variables to compute forecasts of the 1993 endogenous variables, and examine the errors of those forecasts.

When comparing forecasts from different models, bear in mind that the models may differ in their lists of exogenous variables and that this may affect the comparison. For example, a model that has hard-to-forecast exogenous variables is not going to be helpful for practical *ex ante* forecasting, even if it makes excellent *ex post* forecasts.

Errors of *ex ante* and *ex post* forecasts tell us different things. *Ex ante* forecasting errors tell us about the quality of true forecasts but do not allow us to separate the effects of incorrect estimated models from the effects of bad predictions of exogenous variables and disturbances. *Ex post* forecasting errors tell us how good an estimated model has been as a scientific hypothesis, which is

³See Goodhart (1981).

Table 1

Root Mean Square Percentage Errors of *Ex Post* Forecasts with No Subjective Adjustments of the Forecasts, from about 1965 to 1973, Averaged over Eight Models

Variable	Horizon		
	1 quarter	4 quarters	8 quarters
Nominal GNP	0.7	2.0	4.5
Real GNP	0.7	1.9	2.5
GNP Deflator	0.4	0.6	1.9

distinct from anyone's ability to forecast exogenous variables and disturbances. If you are interested in the quality of practical forecasting, you should evaluate *ex ante* forecasts. If you are interested in the quality of a model as a scientific theory, you should evaluate *ex post* forecasts. *Ex post* forecasts are usually more accurate than *ex ante* forecasts because the predictions of the exogenous variables that go into *ex ante* forecasts are usually at least somewhat wrong.

What if we want to make forecasts *two* years ahead, for 1994, based on data up to and including 1992? We need 1993 values for the endogenous variables to use as lagged endogenous values for our 1994 forecast; however, we do not have actual 1993 data. Hence we must make a one-year-ahead forecast for 1993 as before. Then we can make our 1994 forecast using our 1993 forecasts as the lagged values of the endogenous variables for 1994. Thus the errors of our 1994 forecast will depend partly on the errors of our 1993 forecast and partly on the values we use for the 1994 exogenous variables and disturbances. If we want to make forecasts for *n* years ahead instead of two years ahead, the situation is similar except that *n* steps are required instead of two. We can still consider either *ex ante* or *ex post* forecasts. As before, *ex post* forecasts use actual values of the exogenous variables.

When making *ex ante* forecasts, the typical econometric forecaster does not automatically adopt the forecasts generated by a model. Instead the forecaster compares these forecasts with his subjective judgment about the future of the economy, and if there are substantial discrepancies, he makes subjective adjustments to his model's forecasts. This is usually done with subjective adjustments to the predicted disturbances. Thus the accuracy of *ex ante* forecasts

typically depends not only on the adequacy of the estimated model, but also on the model builder's ability to forecast exogenous variables and to make subjective adjustments to the model's forecasts. Paul Samuelson once caricatured this situation at a meeting some years ago by likening the process that produces *ex ante* econometric forecasts to a black box inside which we find only Lawrence R. Klein!

Errors of Forecasts from Several Econometric Models

Most presentations of forecasting accuracy are based on *ex ante* rather than *ex post* forecasts, often with subjective adjustments, perhaps because of the interest in practical forecasting. I like to look at *ex post* forecast errors without adjustments because I am interested in econometric models as scientific hypotheses.

Fromm and Klein (1976) and Christ (1975) discuss root mean square errors (RMSEs) of *ex post* quarterly forecasts of real GNP, nominal GNP and the GNP deflator one quarter to eight quarters ahead by eight models with *no subjective adjustment* by the forecaster. The models were formulated by Brookings, the U.S. Bureau of Economic Analysis, Ray Fair, Leonall Andersen of the Federal Reserve Bank of St. Louis, T. C. Liu and others, the University of Michigan and the Wharton School (two versions). For GNP they show RMSEs rising from 0.7 percent to 2.5 or 4.5 percent of the actual value as the horizon increases from one quarter to eight quarters. For the GNP deflator they show RMSEs rising from 0.4 percent to 1.9 percent, as shown in table 1.

In a series of papers over the past several years, Stephen McNees (1986, 1988 and 1990) has reported on the accuracy of *subjectively*

adjusted ex ante quarterly forecasts of several macroeconomic models, for horizons of one to eight quarters ahead, and has compared them with two simple mechanical forecasting methods. One is the univariate ARIMA method of Charles Nelson (1984), which is called BMARK (for benchmark). The other is the Bayesian vector autoregression method of Robert Litterman (1986), which is called BVAR. The models discussed in McNees (1988) are those formulated by the U.S. Bureau of Economic Analysis, Chase Econometrics, Data Resources Inc., Georgia State University, Kent Institute, the University of Michigan, UCLA and Wharton.

McNees' results for *quarterly* forecasts may be summarized in the following five statements:

1. The models' forecast errors were usually smaller than those of BMARK.⁴
2. The models' forecast errors were usually slightly smaller than those of BVAR for nominal GNP and most other variables and slightly larger than those of BVAR for real GNP. Thus BVAR was usually better than BMARK for real GNP.⁵
3. Forecast errors for the *levels* of variables became worse as the forecast horizon lengthened from one quarter to eight quarters, roughly quadrupling for most variables and increasing tenfold for prices. However, forecast errors for the *growth rates* of many variables (but not for price variables) improved as the horizon lengthened. In other words, for many variables, the forecasts for growth rates averaged over several quarters were better than the forecasts for short-term fluctuations.⁶
4. Mean absolute errors (MAEs) of the models' forecasts of the level of nominal GNP were usually about 0.8 percent of the true level for forecasts one quarter ahead and increased gradually to about 2.2 percent for forecasts one year ahead and about 4 percent for forecasts two years ahead. Real GNP forecast errors were somewhat smaller. Errors for other variables were comparable. Price-level forecast errors were smaller for the one-quarter horizon but grew faster and were larger for the two-year horizon.⁷

5. When subjectively adjusted forecasts were compared with unadjusted forecasts, the adjustments were helpful in most cases, though sometimes they made the forecast worse. Usually the adjustments were larger than optimal.⁸

One-year-ahead *annual* forecasts of real GNP by the University of Michigan's Research Center in Quantitative Economics, by the Council of Economic Advisers and by private forecasters covered by the ASA/NBER survey all had MAEs of about 0.9 percent to 1.1 percent of the true level, and RMSEs of about 1.2 percent to 1.5 percent of the true level.⁹ (The relative sizes of the MAEs and RMSEs are roughly consistent with the fact that for a normal distribution, the RMSE is about 1.25 times the MAE.)

Implications of Worsening Ex Post Forecast Errors

Because the root mean square error of an econometric model's *ex post* forecasts roughly quadruples when the horizon increases from one quarter to eight quarters as in table 1, can we conclude that the model is no longer correct for the forecast period? The answer is possibly, but not certainly.

For a static model we could conclude this because the error of each forecast would involve disturbances only for the period being forecast, not for periods in the earlier part of the horizon. Hence there is no reason to expect great changes in the size of the forecasting error for a static model as the horizon increases. Small increases will occur because of errors in the estimates of the model's parameters if the values of the model's independent variables move further away from their estimation-period means as the horizon lengthens. This is because any errors in the estimates of equations' slopes will generate larger effects as the distance over which the slopes are projected increases.

But most econometric forecasting models contain lagged endogenous variables. Therefore, as noted previously, to forecast *n* periods ahead, we must first forecast the lagged endogenous-variable values that are needed for the *n*-periods-

⁴See McNees (1988 and 1990).

⁵See McNees (1990).

⁶See McNees (1988).

⁷See McNees (1988).

⁸See McNees (1990).

⁹See McNees (1988).

Table 2

**Standard Deviation of Error of N-Periods-Ahead Forecast
from the Equation $y_t = \alpha + \beta y_{t-1} + \varepsilon_t$
Relative to the Standard Deviation of ε_t , as a Function of N
and β , when α and β Are Known**

β	β^2	Horizon, N				
		1	2	4	8	∞
0.7070	0.50	1.00	1.22	1.37	1.41	1.41
0.8944	0.80	1.00	1.34	1.72	2.04	2.24
0.9486	0.90	1.00	1.38	1.85	2.39	3.16
1.0000	1.00	1.00	1.41	2.00	2.83	∞

Here and in what follows, I have expressed interest rates in units of percent per year, so a 5 percent rate is entered as 5, not as 0.05, and its inverse 0.20, not 20. The Appendix gives the definitions and data sources for variables in this and subsequent equations. Latané showed the unadjusted correlation coefficient r , but showed neither the standard deviation nor the t-ratio of the slope. I calculated the adjusted \bar{r}^2 and the t-ratio. The latter is the square root of $r^2 (df)/(1-r^2)$, where df , the number of degrees of freedom, equals 32.

This specification has some of the properties of a theoretical money demand equation—namely, a positive income elasticity (restricted to be constant and equal to one by construction) and a negative interest elasticity (restricted to have an absolute value less than one and not constant). But its least-squares estimate would almost certainly be biased or inconsistent, even if the form of the equation were correct, because the bond rate is almost certainly not exogenous and hence not independent of the equation's disturbances.

Nevertheless, this specification has continued to work fairly well for other periods. Nearly 30 years ago M1/GNP was described for 1892–1959 by a similar regression on the inverse of Moody's Aaa bond rate with almost the same coefficients, as follows:¹¹

$$(2) \text{ M1/GNP} = .131 + .716/\text{RAaa}, \bar{r}^2 = .76$$

(t-ratio) (14)

For 1959–91 the same specification describes the ratio of M1 to GNP with almost the same coefficients, as follows:

$$(3) \text{ M1/GNP} = .085 + .774/\text{RAaa}, \bar{r}^2 = .90$$

(t-ratio) (13) (17)

If GNP in equation (3) is replaced by the new output variable GDP for 1959–91, the result is almost identical, as follows:

$$(4) \text{ M1/GDP} = .086 + .771/\text{RAaa}, \bar{r}^2 = .91$$

(t-ratio) (13) (18)

David Dickey's discussion is based on the 1959–91 data that underlie equation (3).

For 1892–1991 a similar result is again obtained, as follows:

$$(5) \text{ M1/GNP} = .083 + .874/\text{RAaa}, \bar{r}^2 = .89$$

(t-ratio) (11) (28)

Table 3 shows the estimated equations (1)–(5) and several other estimated equations that will be described soon. Equations (1') and (2') are attempts to duplicate the results in equations (1) and (2) using the same data base that is used in equations (3), (5) and later equations. The Appendix gives data sources.

Figure 2 shows the graphs of M1/GNP and 1/RAaa over time. Figures 3 and 4 show the scatter diagrams for equations (3) and (5), respectively. (I should add that, of the four

¹¹See Christ (1963).

Table 3

Regressions of M1/GNP or M1/GDP on 1/RAaa and Other Variables*
(t-ratios are in parentheses)

Eq	Sample	Constant	Estimated Coefficient (and t-ratio)			R ²	DW
			1/RAaa	(M1/GNP) ₋₁	(1/RAaa) ₋₁		
PLAIN-VANILLA EQUATIONS							
1	1919-1952	.100	.795(10)			.75	
1'	1919-1952	.136(7)	.713(10)			.75	.56
2	1892-1959	.131	.716(14)			.76	
2'	1892-1959	.132(9)	.712(14)			.73	.62
3	1959-1991	.085(13)	.774(17)			.90	.38
4	1959-1991	.086(13)	.771(18)			.91	.36
5	1892-1991	.083(11)	.874(28)			.89	.48
AR(1) EQUATIONS**							
15	1960-1991	.013	.267(2.8)	.896(26)	-.239	.98	1.82
17	1893-1991	.020	.711(7)	.831(12)	-.591	.95	1.60
ECM EQUATIONS**							
18	1960-1991	.016(2.2)	.275(2.8)	.857(11)	-.212(-2.2)	.98	1.78
19	1893-1991	.016(2.1)	.593(5)	.807(12)	-.428(-3.6)	.95	1.59

*The dependent variable is M1/GNP in all equations except equation (4), where it is M1/GDP. Definitions and data sources for the variables M1, GNP, GDP and RAaa are given in the Appendix. As explained in the Appendix, a uniform set of data for M1, GNP and RAaa was used for equations (1'), (2'), (3), (5) and later equations; slightly different data were used for equations (1), (2) and (4). David Dickey's discussion is based on equation (3) and the data underlying it. Equations (1') and (2') are attempts to reproduce equations (1) and (2), respectively, using the same data that were used for equations (3), (5) and later equations.

**AR(1) means "first-order autoregressive." ECM means "error correction mechanism."

equations that can be obtained by regressing either the velocity of M1 or its inverse on either RAaa or its inverse, the form that is presented here fits the best.)

It is rather remarkable that this plain-vanilla specification continues to describe the relation between M1's velocity and the long-term Aaa bond rate with such similar regression and correlation coefficients for the four periods, especially in view of the changes in interest-rate regulation and in the definition of M1 that have occurred over the last century. However, the differences among the four estimated versions are not negligible, as seen in a comparison of the computed values of M1/GNP that they yield. For 1959-91 these computed values are shown in figure 5 together with the actual values of M1/GNP. Note that those computed from equations (1) and (2) using 1919-52 and 1892-1959 data are *ex post* forecasts, whereas those from equations (3) and (5) using 1959-91 and 1892-1991 data are within-sample calculated values. Figure 6 shows the values of M1/GNP obtained when equation (3) based on 1959-91 data is used to backcast M1/GNP for 1892-1958, and it also

shows the actual values and the calculated values from equation (5) using 1892-1991 data. The forecasting and backcasting errors are by no means negligible, but the general pattern of behavior of M1/GNP is reproduced.

The estimates of the plain-vanilla equation are rather stable across time, as indicated by figures 7 and 8 which show the behavior of the slope as the sample period is gradually lengthened by adding one year at a time. In figure 7 the sample period starts with 1959-63 and is extended a year at a time to 1959-91. In figure 8 the sample period starts with 1892-97 and is gradually extended to 1892-1991. In each figure the slope settles down quickly after jumping around at first and varies little as the sample is extended thereafter.

However, this simple specification does not by any means satisfy all of the desiderata listed previously. In particular, the 1959-91 Durbin-Watson statistic is a minuscule 0.38, and the 1892-1991 Durbin-Watson statistic of 0.48, is not much better, which suggests that the residuals have a strong positive serial correlation. This by itself would not create bias in the estimates if

Figure 2
M1/GNP and 1/RAaa, 1892-91

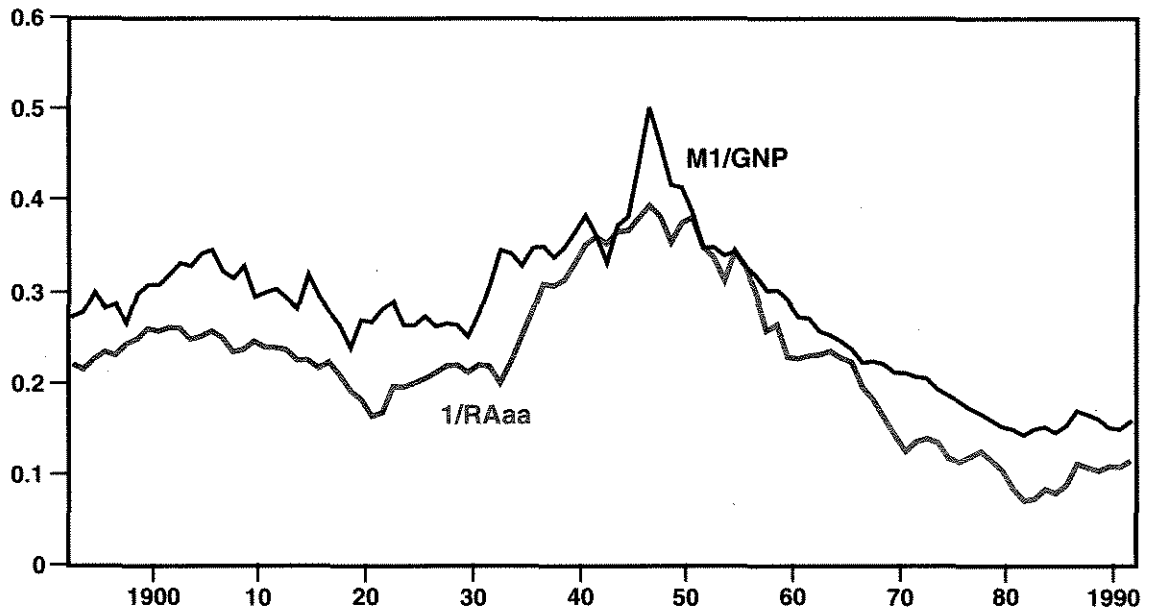


Figure 3
Regression of M1/GNP on 1/RAaa, 1959-91

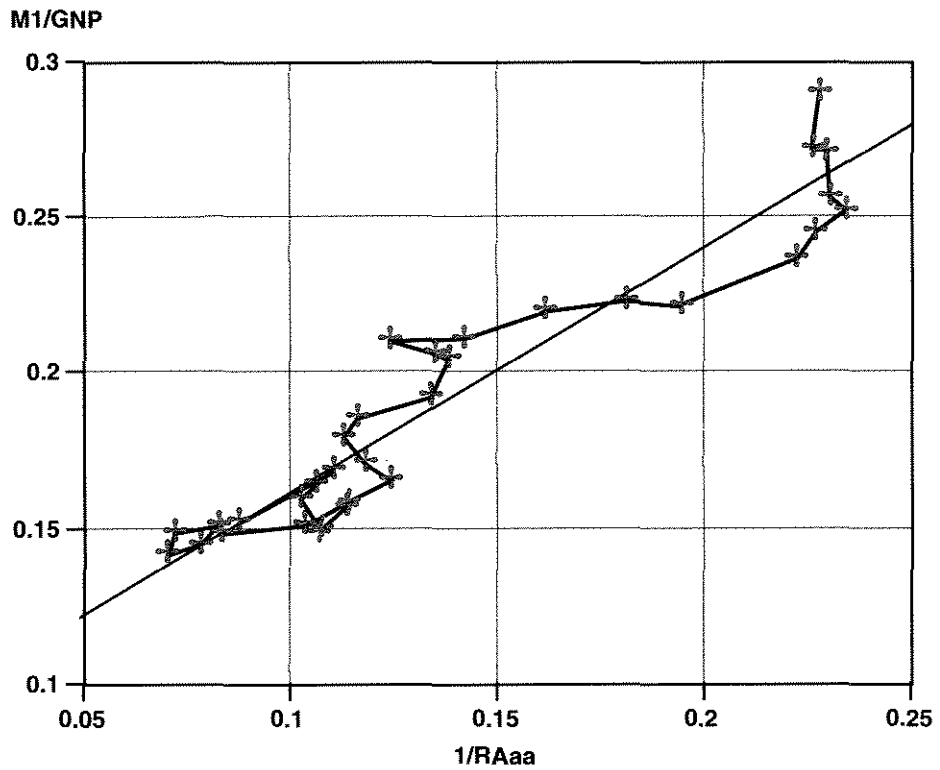


Figure 4
Regression of M1/GNP on 1/RAaa, 1892-91

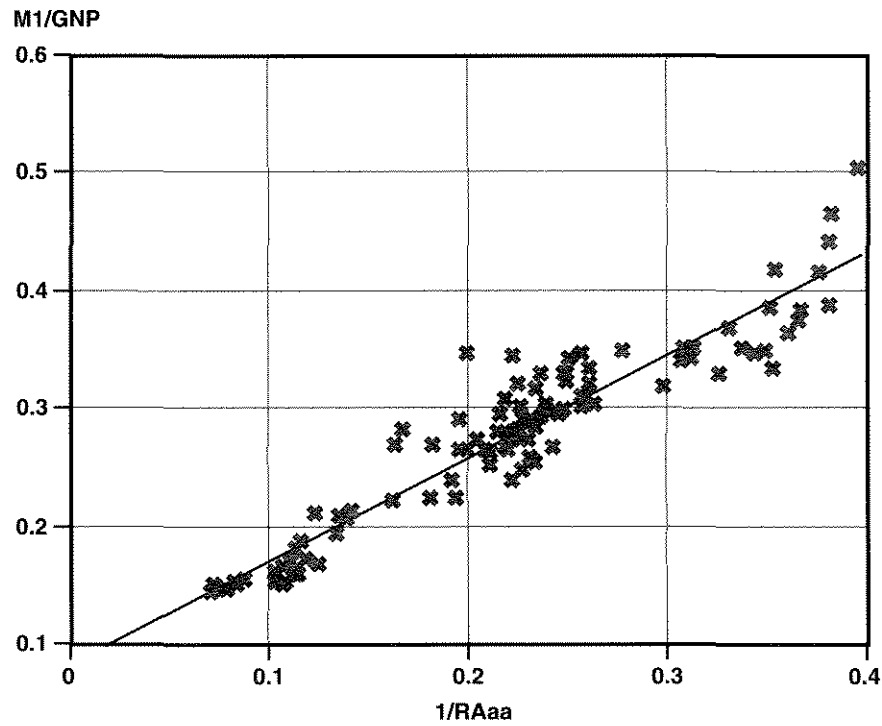


Figure 5
Actual, Computed and Forecast Values of M1/GNP from Regressions on 1/RAaa for Four Periods

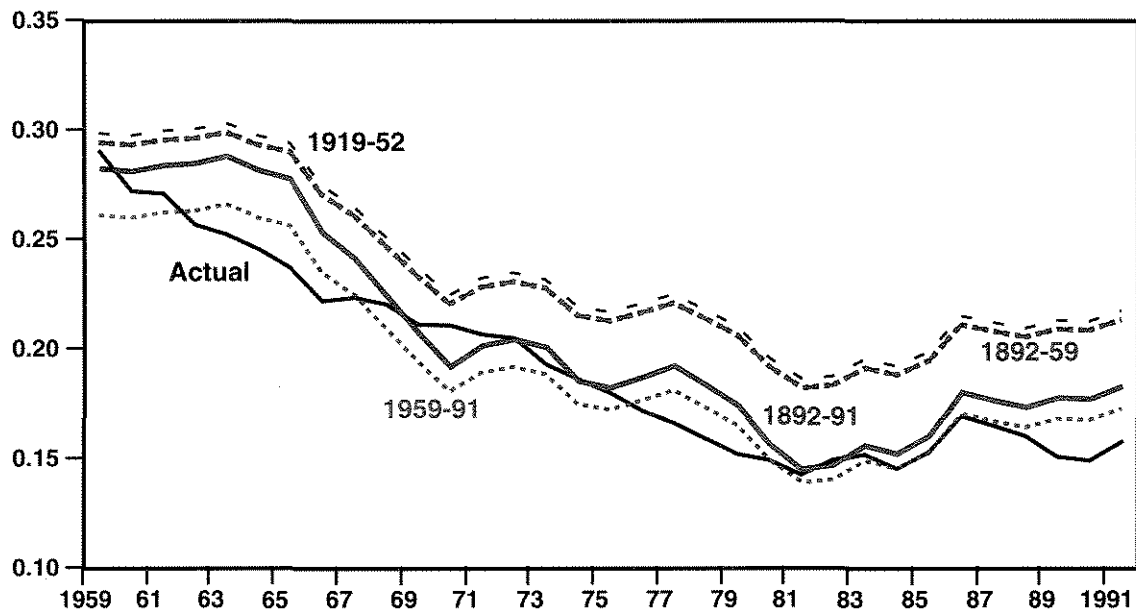


Figure 6
Actual, Computed and Backcast Values of M1/GNP from Regressions on 1/RAaa for Two Periods

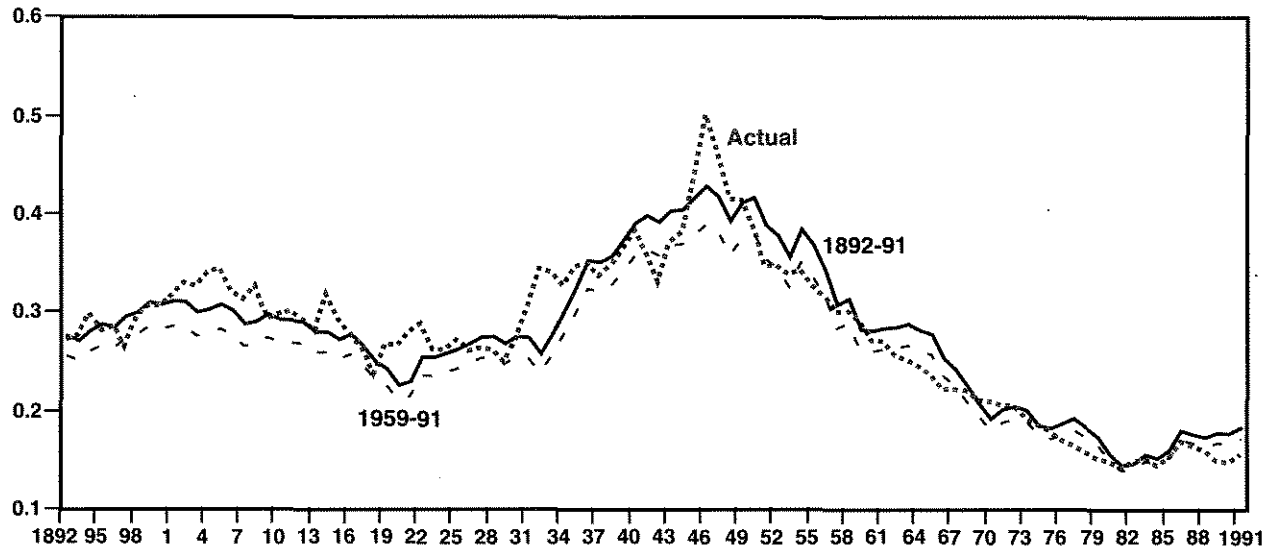


Figure 7
Estimates of Slope in Regression of M1/GNP on 1/RAaa for Samples Starting in 1959 and Ending in 1963...1991

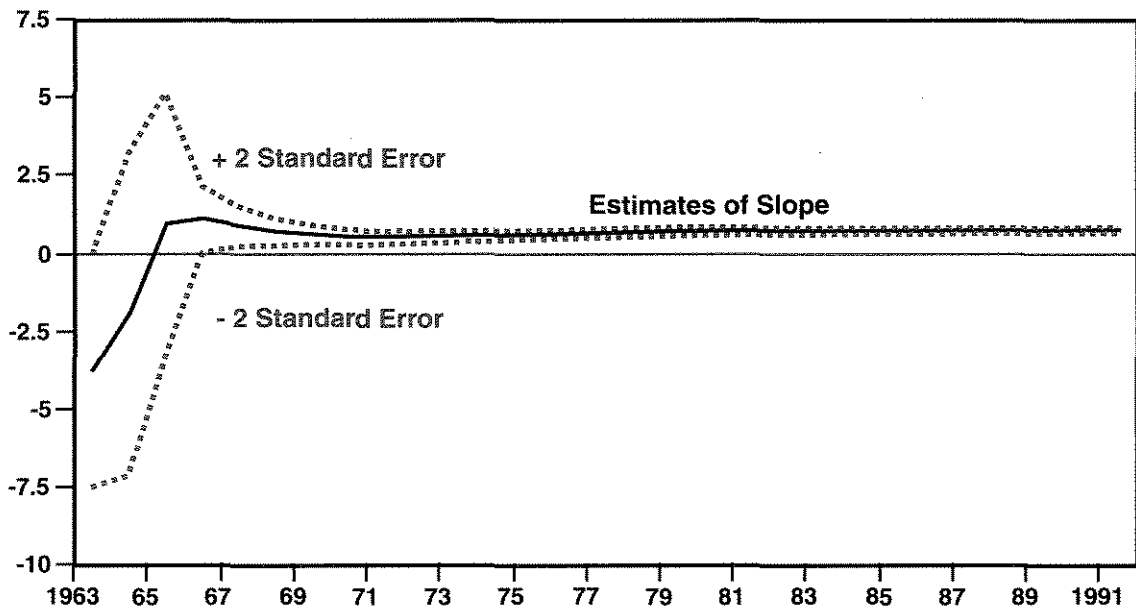
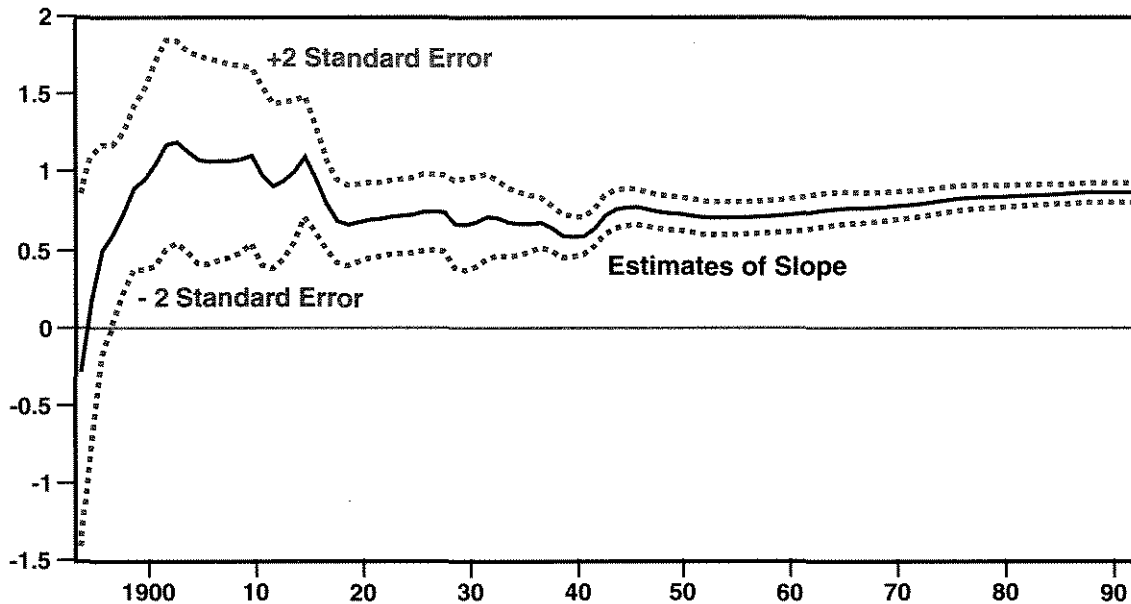


Figure 8
Estimates of Slope in Regression of M1/GNP on 1/RAaa for Samples Starting in 1892 and Ending in 1897...1991



the equation form were correct and if the disturbance were independent of the interest rate and had zero mean and constant variance. But it certainly suggests strongly that the equation has not captured all its relevant systematic factors. The graph of the residuals of the 1959-91 equation (3) against time is illuminating. It shows an almost perfect 12-year cycle of diminishing amplitude with peaks (positive residuals) in 1959 (or possibly earlier), 1970 and 1982 and troughs (negative residuals) in 1965, 1977 and 1990. It also suggests a negative time trend. The residuals of the 1892-1991 equation (5) show a roughly similar pattern. (See figures 9 and 10.)

The very low Durbin-Watson statistics suggest that the equation should be estimated either using the first differences of its variables, or better, using the levels of its variables with a first-order autoregressive [AR(1)] correction applied to its residuals. Estimation in levels with an AR(1) correction would be appropriate if the disturbance u in the original equation were equal to its own lagged value times a constant, ρ , plus a serially independent disturbance, ε , with constant variance, as follows:

$$(6) u_t = \rho u_{t-1} + \varepsilon_t$$

In this case, if the original equation is

$$(7) y_t = \alpha + \beta x_t + u_t = \alpha + \beta x_t + \rho u_{t-1} + \varepsilon_t$$

the AR(1) correction subtracts ρ times the lagged version of equation (7) from equation (7) itself and produces the following equation:

$$(8) y_t = \rho y_{t-1} + (1 - \rho) \alpha + \beta x_t - \beta \rho x_{t-1} + \varepsilon_t$$

This equation is nonlinear in the parameters because the coefficient of lagged x , $-\beta\rho$, is the negative of the product of the coefficients of x and lagged y . If that restriction is ignored and the coefficient of lagged x is denoted by γ , the equation becomes as follows:

$$(9) y_t = \rho y_{t-1} + (1 - \rho) \alpha + \beta x_t + \gamma x_{t-1} + \varepsilon_t$$

This equation can be given the following error correction interpretation. Suppose that the equilibrium value y^* of a dependent variable y is linear in an independent variable x , as follows:

$$(10) y_t^* = \alpha + \beta x_t$$

and that the change in y depends on both the change in the equilibrium value and an error

Figure 9

Residuals from Regression of M1/GNP on 1/RAaa, 1959-91

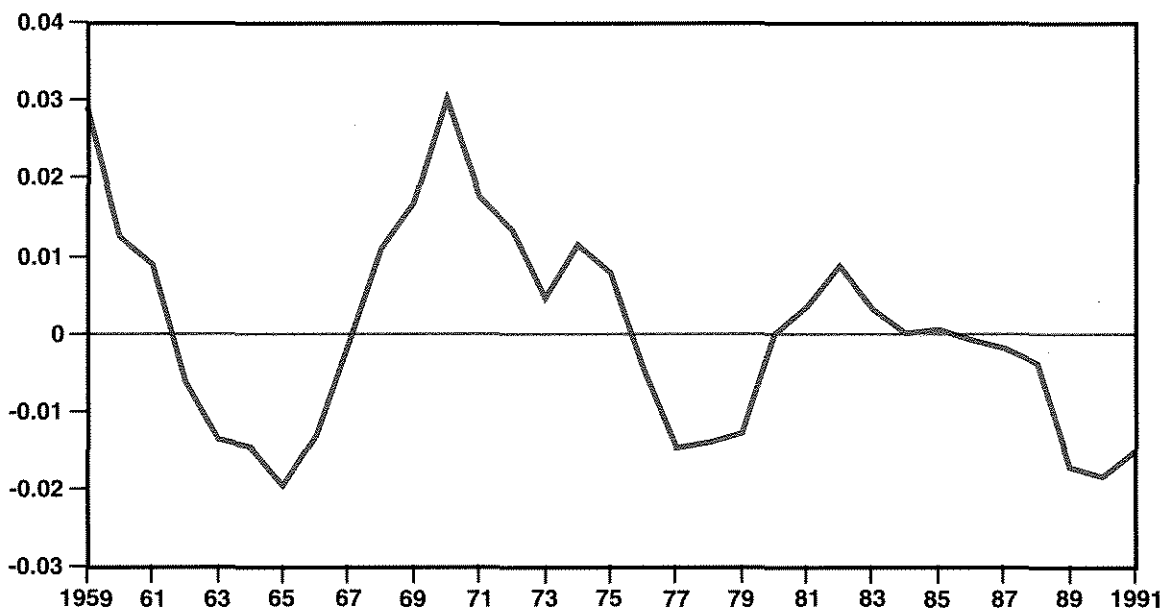
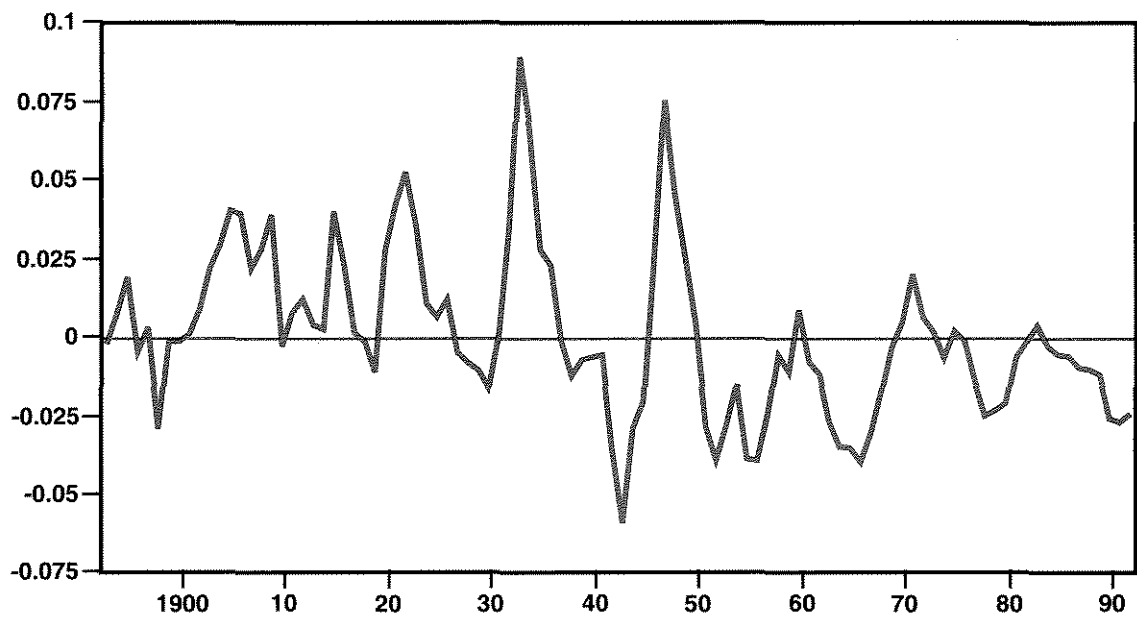


Figure 10

Residuals from Regression of M1/GNP on 1/RAaa, 1892-1991



correction term proportional to the gap between the lagged equilibrium and the lagged actual values, as follows:

$$(11) \Delta y_t = \theta \Delta y_t^* + (1 - \rho) (y_{t-1}^* - y_{t-1}) + \varepsilon_t$$

Substitution from equation (10) into equation (11) implies an equation with the same variables as the AR(1) equation (8) but with some different parameters, as follows:

$$(12) y_t = \rho y_{t-1} + (1 - \rho)\alpha + \theta \beta x_t + (1 - \rho - \theta)\beta x_{t-1} + \varepsilon_t$$

If the adjustment parameter θ in equation (12) were equal to one, then equation (12) would become the same equation as (8).

Estimates in first differences would be appropriate if the value of ρ in equation (6), (7) and (8) were one. In this case, equation (8) becomes a first-difference equation, as follows:

$$(13) \Delta y_t = \beta \Delta x_t + \varepsilon_t$$

The least-squares estimate of equation (8) in levels with the AR(1) correction for 1960–91 is as follows:

$$(14) \text{M1/GNP} = .896(\text{M1/GNP})_{-1} \\ = (1 - .896) .126 + .267(1/\text{RAaa} - .896/\text{RAaa}_{-1}) \\ \text{(t-ratio) (8) (2.8) (26)}$$

with an adjusted R squared of .98 and DW equal to 1.82. This is equivalent to the following equation:

$$(15) \text{M1/GNP} = .896(\text{M1/GNP})_{-1} \\ + .013 + .267/\text{RAaa} - .239/\text{RAaa}_{-1}$$

There is no evidence of a trend.

The least-squares estimate in levels with the AR(1) correction for 1893–1991 is as follows:

$$(16) \text{M1/GNP} = .831(\text{M1/GNP})_{-1} \\ = (1 - .831) .117 + .711(1/\text{RAaa} - .831/\text{RAaa}_{-1}) \\ \text{(t-ratio) (5) (7) (12)}$$

with an adjusted R squared of .95 and DW equal to 1.60. This is equivalent to the following equation:

$$(17) \text{M1/GNP} = .831(\text{M1/GNP})_{-1} \\ + .020 + .711/\text{RAaa} - .591/\text{RAaa}_{-1}$$

There is again no evidence of a trend.

Least-squares estimation of the ECM equation (12) for 1960–91 (without restricting θ to be one) yields the following equation:

$$(18) \text{M1/GNP} = .857(\text{M1/GNP})_{-1} \\ \text{(t-ratio) (11)} \\ + .016 + .275/\text{RAaa} - .212/\text{RAaa}_{-1} \\ \text{(2.2) (2.8) (-2.2)}$$

with an adjusted R squared of .98 and DW equal to 1.78. This is quite close to the AR(1) result in equation (15), which suggests that the adjustment coefficient θ in equation (12) is not very different from one. The hypothesis that in equation (18) the coefficient of lagged 1/RAaa is equal to the negative of the product of the coefficients of 1/RAaa and lagged M1/GNP, as required by equation (8) and as satisfied by equation (15), is strongly accepted by a Wald test (the p-value is .59).

Least-squares estimation of equation (12) for 1893–1991 (again without restricting θ to be one) yields the following equation:

$$(19) \text{M1/GNP} = .807(\text{M1/GNP})_{-1} \\ \text{(t-ratio) (12)} \\ + .016 + .593/\text{RAaa} - .428/\text{RAaa}_{-1} \\ \text{(2.1) (5) (-3.6)}$$

with an adjusted R squared of .95 and DW equal to 1.59. This is quite close to the AR(1) result in equation (17), which again suggests that the adjustment coefficient θ in equation (12) is not very different from one. The hypothesis that in equation (19) the coefficient of lagged 1/RAaa is equal to the negative of the product of the coefficients of 1/RAaa and lagged M1/GNP, as required by equation (8) and as satisfied by equation (17), is accepted by a Wald test (the p-value is .11).

Equations (15), (17), (18) and (19) are better than the plain-vanilla equations (3) and (5) in some respects, and worse in others. They have substantially higher adjusted R-squared values, much less serial correlation in their residuals, no evidence of a time trend, and significant coefficients. The ECM equations (18) and (19), however, are very unstable over time. In equation (18) the coefficient of 1/RAaa varies from about .6 for 1960–70, to .05 for 1960–78 and 1960–81, to .3 for 1960–86 and 1960–91. In equation (19) the coefficient of 1/RAaa varies almost as much but remains at about .7 or .6 for samples that include at least the years 1893–1950.

Table 4

**Regressions of $\Delta(M1/GNP)$ on $\Delta(1/RAaa)$ without a Constant*
(t-ratios are in parentheses)**

Eq	Sample	Coef of $\Delta(1/RAaa)$	\bar{R}^2	DW
24	1960-1991	.380(3.6)	.05	1.23
25	1893-1991	.494(4.1)	.15	1.76

*Definitions and data sources for the variables M1, GNP and RAaa are given in the appendix.

I conjecture that in the AR(1) equations (15) and (17) the coefficient of $1/RAaa$ is also unstable across time because the AR(1) and ECM equation estimates are quite similar.

By comparing equations (12) and (18), one can solve for the 1960-91 estimates of the four parameters ρ , α , β and θ , in that order, to obtain:

$$(20) \hat{\rho} = .857, \hat{\alpha} = .112, \hat{\beta} = .441 \text{ and } \hat{\theta} = .624$$

This implies that the equilibrium relation in equation (10) embedded in the ECM is as follows:

$$(21) (M1/GNP)^* = .112 + .441/RAaa$$

Similarly, by comparing equations (12) and (19) one can solve for the 1893-1991 estimates of the four parameters as follows:

$$(22) \hat{\rho} = .807, \hat{\alpha} = .083, \hat{\beta} = .855 \text{ and } \hat{\theta} = .694$$

This implies that the equilibrium relation in equation (10) embedded in the ECM is as follows:

$$(23) (M1/GNP)^* = .083 + .855/RAaa$$

The two equilibrium relations in equations (21) and (23) for the two periods 1960-91 and 1893-1991 are quite different, which is consistent with the instability of the ECM specification across time.

Now let us return to the first-difference equation (13). The least-squares estimate for 1960-91 is as follows:

$$(24) \Delta(M1/GNP) = .380\Delta(1/RAaa), \bar{r}^2 = .05$$

(t-ratio) (3.6)

with DW = 1.23. For 1893-1991 it is as follows:

$$(25) \Delta(M1/GNP) = .494\Delta(1/RAaa), \bar{r}^2 = .15$$

(t-ratio) (4.1)

with DW = 1.76. Table 4 shows the estimated equations (24) and (25). The estimates of this first-difference specification are not quite as stable across time as those of the specification in levels of the variables. This can be seen by comparing equations (24) and (25) and also from figures 11 and 12, which show the values of the estimates as the sample is increased one year at a time, starting respectively with 1960 and 1893. In each figure the estimates stabilize after an initial period of instability, but the values at which they settle differ by a factor of about .75.

If a constant term is included in equation (24), which implies a trend term in equation (3), the constant is small but significantly negative, the slope falls to about .3, and the adjusted R-squared and DW values improve slightly. The estimated slope, however, becomes wildly unstable across time. If a trend variable is included in equation (3), its coefficient is small but significantly negative, the interest-rate coefficient falls to .49 and remains highly significant, the adjusted R-squared and the DW values rise slightly, and again the estimated slope is wildly unstable across time.

If a constant term is included in equation (25), it is small and insignificantly negative, the rest of the equation is almost unchanged, and the slope becomes quite unstable through time, varying from .6 to zero and back to .6 again. If a trend is included in equation (5), its coefficient is small but significantly negative, the interest-rate coefficient is almost unchanged at .81, the adjusted R-squared value rises a bit, the DW value rises a bit, and the coefficient is again wildly unstable across time.

Figure 11
**Estimates of Slope in Regression of $\Delta(M1/GNP)$ on $\Delta(1/RAaa)$ for
 Samples Starting in 1960 and Ending in 1962...1991**

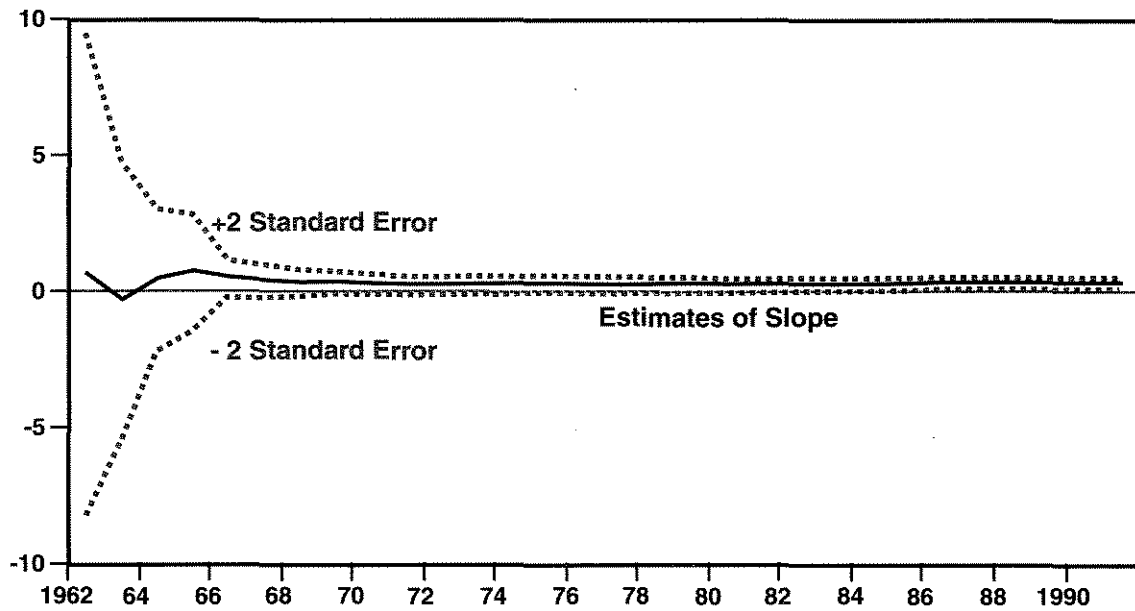
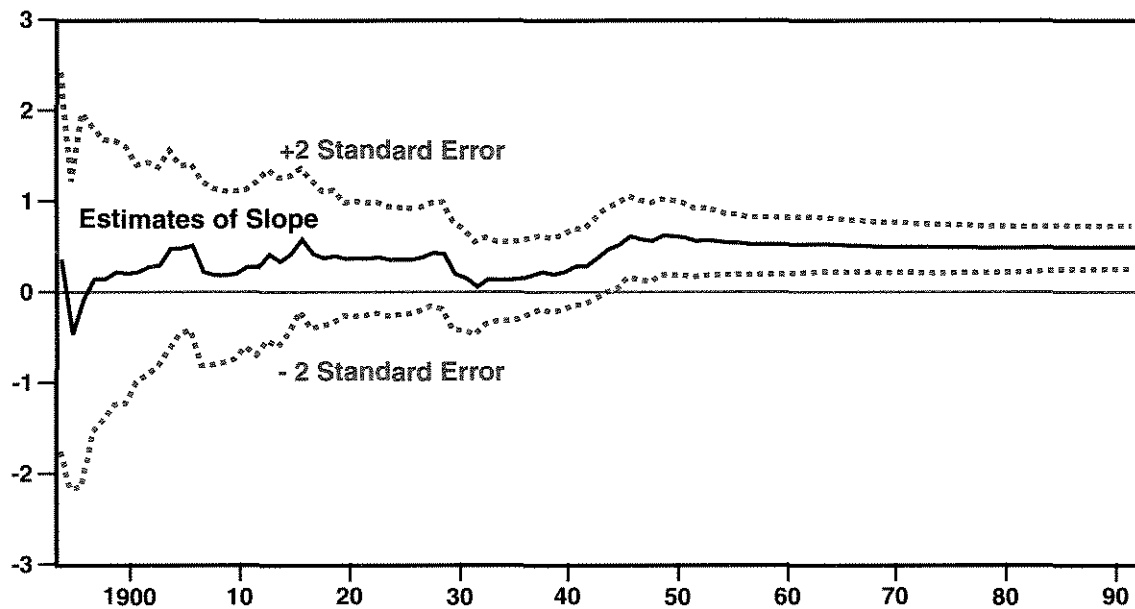


Figure 12
**Estimates of Slope in Regression of $\Delta(M1/GNP)$ on $\Delta(1/RAaa)$ for
 Samples Starting in 1893 and Ending in 1896...1991**



On the whole, the first-difference specification does not stand up well.

Where do matters stand? On the one hand, we have the plain-vanilla equation such as equation (3), which fits only moderately well and has severe serial correlation in its residuals but has an estimated slope that is rather stable across time. On the other hand, we have more complicated dynamic equations such as the ECM equation (18), which fit much better and have nice Durbin-Watson statistics but have estimated coefficients that vary greatly across time. Neither is quite satisfactory, but if the aim is to find an estimated equation that will describe the future as well as it does the past, I think I would now bet on the plain-vanilla specification, even though the relation of its estimated coefficients to structural parameters is unclear.

CONCLUSION

Econometrics has given us some results that appear to stand up well over time. The price and income elasticities of demand for farm products are less than one. The income elasticity of household demand for food is less than one. Houthakker (1957), in a paper commemorating the 100th anniversary of Engel's law, reports that for 17 countries and several different periods these income elasticities range between .43 and .73. Rapid inflation is associated with a high growth rate of the money stock. Some short-term macroeconomic forecasts, especially those of the Michigan model, are quite good.

But there have also been some nasty surprises about which econometrics gave us little or no warning in advance. The short-run downward-sloping Phillips curve met its demise in the 1970s. (Milton Friedman [1968] and Edmund Phelps [1968] predicted that it would.) The oil embargo of 1973 and its aftermath threw most models off. The slowdown of productivity growth beginning in the 1970s was unforeseen. The money demand equation, which appeared to fit well and be quite stable until the 1970s, has not fit so well since then.

How then should we approach econometrics, for science and for policy, in the future? As for science, we should formulate and estimate models as we usually do, relying both on economic theory and on ideas suggested by regularities observed in past data. But we should not fail to test those estimated models against new data

that were not available to influence the process of formulating them. As for policy, we should be cautious about using research findings to predict the effects of any large policy change of a type that has not been tried before.

REFERENCES

- Christ, Carl F. "Interest Rates and 'Portfolio Selection' among Liquid Assets in the U.S.," in Christ et al., *Measurement in Economics: Studies in Mathematical Economics and Econometrics in Memory of Yehuda Grunfeld* (Stanford University Press, 1963).
- _____. "Judging the Performance of Econometric Models of the U.S. Economy," *International Economic Review* (February 1975), pp. 54-74.
- Friedman, Milton. Review of "Business Cycles in the United States of America, 1919-1932" by Jan Tinbergen, *American Economic Review* (September 1940), pp. 657-60.
- _____. "The Role of Monetary Policy," *American Economic Review* (March 1968), pp. 1-17.
- Fromm, Gary, and Lawrence R. Klein. "The NBER/NSF Model Comparison Seminar: An Analysis of Results," *Annals of Economic and Social Measurement* (Winter 1976), pp. 1-28.
- Goodhart, Charles. "Problems of Monetary Management: The U.K. Experience," in A. S. Courakis, ed., *Inflation, Depression, and Economic Policy in the West* (Barnes and Noble Books, 1981).
- Houthakker, Hendrik. "An International Comparison of Household Expenditure Patterns. Commemorating the Centenary of Engel's Law," *Econometrica* (October 1957), pp. 532-51.
- Kendrick, John. *Productivity Trends in the United States* (Princeton University Press, 1961).
- Latané, Henry Allen. "Cash Balances and the Interest Rate—A Pragmatic Approach," *Review of Economics and Statistics* (November 1954), pp. 456-60.
- Litterman, Robert B. "Forecasting with Bayesian Vector Autoregressions—Five Years of Experience," *Journal of Business and Economic Statistics* (January 1986), pp. 25-38.
- Lucas, Robert E. Jr. "Econometric Policy Evaluation: A Critique," *The Phillips Curve and Labor Markets*, Carnegie-Rochester Conference Series on Public Policy, vol. 1, (North-Holland, 1976), pp. 19-46.
- Marschak, Jacob. "Economic Measurements for Policy and Prediction," in William C. Hood and Tjalling C. Koopmans, eds., *Studies in Econometric Method, Cowles Commission Monograph No. 14* (Wiley, 1953), pp. 1-26.
- McNees, Stephen K. "The Accuracy of Two Forecasting Techniques: Some Evidence and an Interpretation," *New England Economic Review* (March/April 1986), pp. 20-31.
- _____. "How Accurate Are Macroeconomic Forecasts?" *New England Economic Review* (July/August 1988), pp. 15-36.
- _____. "Man vs. Model? The Role of Judgment in Forecasting," *New England Economic Review* (July/August 1990), pp. 41-52.
- Mitchell, Wesley C. *Business Cycles: The Problem and Its Setting* (National Bureau of Economic Research, 1927).
- Nelson, Charles R. "A Benchmark for the Accuracy of Econometric Forecasts of GNP," *Business Economics* (April 1984), pp. 52-58.

Phelps, Edmund. "Money-Wage Dynamics and Labor-Market Equilibrium," *Journal of Political Economy* (Part II, July/August 1968), pp. 678-711.

Tinbergen, Jan. *Business Cycles in the United States of America, 1919-1932, Statistical Testing of Business Cycle*

Theories, vol. 2, (League of Nations, 1939).

Zellner, Arnold, and Franz Palm. "Time Series Analysis and Simultaneous Equation Econometric Models," *Journal of Econometrics* (May 1974), pp. 17-54.

Appendix On Data For Tables 3 and 4

A. Data for equations (1'), (2'), (3), (5), (14-19), and (24-25):

M1 = currency plus checkable deposits, billions of dollars

1892-1956, June 30 data: U.S. Bureau of the Census. *Historical Statistics of the U.S. from Colonial Times to 1957* (Government Printing Office, 1960), p. 646, series X-267.

1957-58, June 30 data: *Economic Report of the President*, 1959, p. 186.

1959-91, averages of daily data for December, seasonally adjusted: *Economic Report of the President*, 1992, p. 373.

Note: December data, seasonally adjusted, are close to June 30 data.

GNP = gross national product, billions of dollars per year

1892-1928: Kendrick (1961), pp. 296-7.

1929-59: *Economic Report of the President*, 1961, p. 127.

1960-88: *Economic Report of the President*, 1992, p. 320.

1989-91: *Survey of Current Business*, July 1992, p. 52.

RAaa = long-term high-grade bond rate, percent per year

1892-1918: Macaulay's unadjusted railroad bond rate, U.S. Bureau of the Census. *Historical Statistics of the United States from Colonial Times to 1957* (Government Printing Office, 1960), p. 656, series X-332.

1919-91: Moody's Aaa corporate bond rate:

1919-38: U.S. Bureau of the Census. *Historical Statistics of the United States from Colonial Times to*

1957 (Government Printing Office, 1960), p. 656, series X-333.

1939-91: *Economic Report of the President*, 1992, p. 378.

Note: For pre-1959 data I used sources that were available in 1960, in an attempt to make equation 2' reproduce the 1892-1959 equation 2, which originally appeared in Christ (1963). These same sources also yield equation 1', which is an approximate reproduction of the 1919-52 equation 1, from Latané (1954).

B. Data for 1959-91 for equation (4):

M1 = currency plus checkable deposits, billions of dollars: same as above.

GDP = gross domestic product, billions of dollars per year: *Economic Report of the President*, 1992, pp. 298 or 320.

RAaa = Moody's Aaa corporate bond rate, percent per year: same as above.

C. Data for 1919-52 for equation (1), as described in Latané (1954), p. 457:¹

M1: "demand deposits adjusted plus currency in circulation on the mid-year call date, (Federal Reserve Board Data)."

U.S. Bureau of the Census. *Historical Statistics of the United States from Colonial Times to 1957* (Government Printing Office, 1960). Series X-267

GNP: "Department of Commerce series from 1929 to date; 1919-28 Federal Reserve Board estimates on the same basis (National Industrial Conference Board, *Economic Almanac*, 1952, p. 201)."

¹Though Latané's work was published in 1954, research analysts at the Federal Reserve Bank of St. Louis used more recent data to replicate his work.

RAaa: "interest rate on high-grade long-term corporate obligations. The U.S. Treasury series giving the yields on corporate high-grade bonds as reported in the *Federal Reserve Bulletin* is used from 1936 to date. Before 1936 we use annual averages of Macaulay's high-grade railroad bond yields given in column 5, Table 10, of his *Bond Yields, Interest Rates, Stock Prices*," pp. A157-A161. Macaulay, Frederick R. *Bond Yields, Interest Rates, Stock Prices* (National Bureau of Economic Research, 1938).

D. Data for 1892-1959 for equation (2), as described in Christ (1963), pp. 217-18:²

M1: "currency outside banks" plus "demand deposits adjusted", "billions of dollars as of June 30."

U.S. Bureau of the Census. *Historical Statistics of the United States from Colonial Times to 1957* (Government Printing Office, 1960). Series X-267

U.S. Bureau of the Census. *Historical Statistics of the United States from Colonial Times to 1957; Continuation to 1962 and Revisions* (Government Printing Office, 1965). Series X-267

RAaa: "long-term interest rate (Moody's Aaa corporate bond rate, extrapolated before 1919 via Macaulay's railroad bond yield index)", "percent per year."

GNP: "gross national product, billions of dollars per year."

²Though Christ's work was published in 1963, research analysts at the Federal Reserve Bank of St. Louis used more recent data to replicate his work.