Testing Macroeconometric Models

By RAY C. FAIR*

Interest in research topics in different fields fluctuates over time, and the field of macroeconomics is no exception. From Jan Tinbergen's (1939) model-building in the late 1930's through work in the 1960's, there was considerable interest in the construction of structural macroeconomic models. The dominant methodology of this period was what I will call the "Cowles Commission" approach. Structural econometric models were specified, estimated, and then analyzed and tested in various ways. One of the major macroeconometric efforts of the 1960's, building on the earlier work of Lawrence Klein (1950) and Klein and Arthur Goldberger (1955), was the Brookings model (James Duesenberry et al., 1965, 1969). This model was a joint effort of many individuals, and at its peak it contained nearly 400 equations. Although much was learned from this exercise, the model never achieved the success that was initially expected, and it was laid to rest around 1972.

Two important events in the 1970's contributed to the decline in popularity of the Cowles Commission approach. The first was the commercialization of macroeconometric models. This changed the focus of research on the models. Basic research gave way to the day-to-day needs of keeping the models up-to-date, of subjectively adjusting the forecasts to make them "reasonable," and of meeting the special needs of clients. The second event was Robert Lucas's (1976) critique, which argued that the models are not likely to be useful for policy purposes. The Lucas critique led to a line of research that culminated in real-business-cycle theories, which in turn generated a counter-response in the form of new Keynesian economics.

When Zvi Griliches asked me to organize a session entitled "What Happened to Macroeconometric Models?" he left ambiguous (at least to me) whether or not he felt that the premise of the session should be that macroeconometric models had died, with the task of the session being to examine why. My interest in structural macroeconomic model-building began when I was a graduate student at M.I.T. in the mid-1960's. This was a period when there was still interest in the Brookings-model project and when intensive work was being carried out on the MPS (M.I.T.-Penn-SSRC) model. Many hours were spent by many students in the basement of the Sloan building at M.I.T. working on various macroeconometric equations using an IBM 1620 computer (punch cards and all). This was also the beginning of the development of TSP (Time Series Processor), a computer program that provided an easy way of using various econometric techniques. The program was initiated by Robert Hall, and it soon attracted many others to help in its development. I played a minor role in this work.

Perhaps because of fond memories of my time in the basement of Sloan, I have never lost interest in structural models. I continue to believe that the Cowles Commission approach is the best way of trying to learn how the macroeconomy works, and I have continued to try to make progress using this approach. My view is thus that macroeconometric models have not died, even though there has been limited academic interest in these models in the last 20 years. I have argued elsewhere (Fair, 1992) that macroeconomics has not been well served by the real-business-cycle approach, which is not interested in testing models in a serious way, and by new Keynesian economics,

†*Discussants:* Olivier Blanchard, Massachusetts Institute of Technology; William Brainard, Yale University.

*Cowles Foundation, Yale University, Box 2125, Yale Station, New Haven, CT 06520

which has moved macroeconomics away from its econometric base.

This paper is a brief review of the progress that I feel has been made in the development of macroeconometric techniques in the last two decades. It is written for those who have paid little attention to the field for many years and would like a general idea of what has been going on. Because of space limitations, this is not an extensive review, and only a few references are given. More extensive discussions and lists of references are contained in Fair (1984, 1993). I argue that progress has been made in the last two decades in improving the ability of researchers to estimate, test, and analyze macroeconometric models. In particular, progress has been made in testing, and this is emphasized below. I hope in the next two decades that the Cowles Commission approach will attract more academic interest and that more attention will be given to testing and improving structural models.

I. Estimation, Stochastic Simulation, and Rational Expectations

The following notation is used. The (nonlinear) model is written as

$$(1) \qquad f_i(\mathbf{y}_t, \mathbf{x}_t, \boldsymbol{\alpha}_i) = u_{it} \qquad i = 1, \ldots, n \quad t = 1, \ldots, T$$

where $\mathbf{y}_t$ is an n-dimensional vector of endogenous variables, $\mathbf{x}_t$ is a vector of predetermined variables (including lagged endogenous variables), $\boldsymbol{\alpha}_i$ is a vector of unknown coefficients, and u_{it} is the error term for equation i for observation t. It is assumed that the first m equations are stochastic, with the remaining u_{it} $(i = m+1, \ldots, n)$ identically zero for all t. The T-dimensional vector $(u_{1t}, \ldots, u_{mt})$ will be denoted by $\mathbf{u}_t$, and $\boldsymbol{\Sigma}$ will denote the $m \times m$ covariance matrix of $\mathbf{u}_t$.

Advances in computational techniques and computer hardware have considerably lessened the computational burden of working with large-scale models. William Parke's (1982) algorithm opened up the possibility of estimating large-scale nonlinear models by three-stage least squares (3SLS) and full-information maximum likelihood (FIML), and with current personal computers like the 486's, estimation of a large-scale model by 3SLS or FIML requires at most a few hours of computer time. Estimation by two-stage least squares (2SLS) is almost instantaneous, and estimation using a robust estimator like two-stage least absolute deviations (2SLAD) is also very fast with the use of a computational trick.

The availability of fast, inexpensive computers has made stochastic simulation of macroeconometric models routine, and as discussed below, this has greatly expanded the kinds of research that can be done on these models. Stochastic simulation requires that an assumption be made about the distribution of $\mathbf{u}_t$. It is usually assumed that $\mathbf{u}_t$ is an independently and identically distributed multivariate normal $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, although other assumptions can be used. Given consistent estimates of $\boldsymbol{\alpha}_i$ for all i (denoted $\hat{\boldsymbol{\alpha}}_i$), the covariance matrix $\boldsymbol{\Sigma}$ can be estimated as $(1/T)\hat{\mathbf{U}}\hat{\mathbf{U}}'$, where $\hat{\mathbf{U}}$ is the $m \times T$ matrix of values of $\hat{u}_{it}$, where $\hat{u}_{it} = f_i(\mathbf{y}_t, \mathbf{x}_t, \hat{\boldsymbol{\alpha}}_i)$. Given the estimate of $\boldsymbol{\Sigma}$, error terms can be drawn from the $\mathcal{N}(\mathbf{0}, \hat{\boldsymbol{\Sigma}})$ distribution.

Coefficients can also be drawn in stochastic simulation work. Let $\hat{\boldsymbol{\alpha}}$ denote the vector of all the coefficient estimates in the model, and let $\hat{\mathbf{V}}$ denote the estimated covariance matrix of $\hat{\boldsymbol{\alpha}}$. ($\hat{\mathbf{V}}$ obviously depends on the estimation technique used.) Given $\hat{\mathbf{V}}$ and given, say, the normality assumption, coefficients can be drawn from the $\mathcal{N}(\hat{\boldsymbol{\alpha}}, \hat{\mathbf{V}})$ distribution. Exogenous-variable values can also be drawn for stochastic simulations once an assumption is made about the stochastic nature of the exogenous variables.

It is now possible to handle the rational-expectations (RE) assumption in macroeconometric models. Expected values of endogenous variables for future periods can appear as explanatory variables in the stochastic equations. If expectations are rational, they are based on the model and on information up to the beginning of the current period. In other words, under the RE

assumption, the expected values are the predicted values from the model—the expectations are "model consistent." Single-equation estimation of models with RE is possible using Lars Hansen's (1982) method-of-moments estimator, which is a modified version of 2SLS. Solution and FIML estimation are possible using techniques discussed in Fair and John Taylor (1983). Solution of models with RE is difficult because future predicted values affect current predicted values. An iterative technique is needed that iterates over solution *paths* of the endogenous variables. Even given this difficulty, however, most techniques are computationally feasible for models with RE, including stochastic simulation.

II. Testing Single Equations

Testing macroeconometric equations and models is very difficult, which is one of the main reasons why there is so much disagreement in macroeconomics about how the economy works. Lurking everywhere is the potential problem of "data mining"—finding an equation or model that fits well within the estimation period but is in fact a poor approximation of the data generating process. Another difficulty is that models can be based on different sets of exogenous variables, and controlling for these differences in making comparisons across models is not straightforward. Nevertheless, there are many tests available, both for single equations and for complete models.

First, it is possible to examine whether the asymptotic approximations of the distributions of the estimators that are used for hypothesis-testing are accurate. If some of the variables are not stationary, the asymptotic approximations may not be very good. In fact, much of the recent literature in time-series econometrics has been concerned with the consequences of nonstationary variables. The procedure for examining accuracy is to use stochastic simulation and reestimation to get a good approximation of the exact distribution of the estimates and then to compare this distribution to the asymptotic distribution. Take, say, the 2SLS estimates as the base coefficient values, and compute $\hat{\Sigma}$ using these estimates. From the $\mathcal{N}(\mathbf{0}, \hat{\Sigma})$ distribution, draw a vector of the m error terms for each of the T observations. Given these error terms and the 2SLS coefficient estimates, solve the model for the entire period 1 through T. This is a dynamic simulation (i.e., one in which the lagged values of the endogenous variables are updated as the solution proceeds). The predicted values from this solution form a new data set. Estimate the model by 2SLS using this data set, and record the set of estimates. This is one repetition. Repeat the draws, solution, and estimation for many repetitions, and record each set of estimates. If J repetitions are done (where J is a number like 500 or 1,000), one has J values of each coefficient estimate, which are likely to be a good approximation of the exact distribution. This distribution can then be compared to the asymptotic distribution.

Using this procedure, I have found that the estimates of the coefficients of lagged dependent variables are usually biased upward, something that has been known for simple equations since the late 1940's. It is possible to correct for this bias by obtaining "median unbiased estimates" using a modified version of the procedure discussed in Donald Andrews (1993). I have found after correcting for this bias that the exact-distribution approximations are close to the asymptotic distributions. In this sense nonstationarity does not appear to be a problem in macroeconometric models.

A straightforward way of testing the specification of an equation is to add variables to it and test their significance. For the 2SLS estimator, a chi-square test can be used. For example, a test of the dynamic specification of an equation is to add lagged values of the left-hand side and all right-hand-side variables and test whether they are significant. David Hendry et al. (1984) show that adding these lagged values is quite general in that it encompasses many different types of dynamic specifications. If the lagged values are not significant, this is strong support for the dynamic specification.

Another test of the structure of an equation is to add a time trend to the equation. Long before unit roots and cointegration became popular, model-builders worried about picking up spurious correlation from common trending variables. If adding a time trend substantially changes some of the coefficient estimates, this is cause for concern.

A third test is to estimate an equation under the assumption that its error term follows an autoregressive process of order n, where n is a number around 4. Many equations are estimated assuming a first-order process, and if adding a fourth-order process results in a significant increase in explanatory power, this is evidence that the serial-correlation properties of the error term have not been properly accounted for.

A fourth test is to add values *led* one or more times to the equation, estimate the equation using Hansen's (1982) method, and test whether the led values are significant. Again, a chi-square test is available for this purpose. If the led values are not statistically significant, this is evidence against the RE hypothesis. If the led values are significant, this suggests that expectations have not been adequately accounted for.

A fifth test is simply to add variables that might belong in the equation (according to some theories), and test for their significance. For example, I have found age-distribution variables to be significant in aggregate-consumption equations and have added these variables to the equations.

One of the most important issues to examine about an equation is whether its coefficients change over time (i.e., whether the structure is stable over time). A common test of structural stability is to pick a date at which the structure is hypothesized to have changed and then test the hypothesis that a change occurred at this date. The test is usually an F or chi-square test. Recently, however, Andrews and Werner Ploberger (1992; henceforth, AP) have proposed a test that does not require that the date of the structural change be chosen a priori, and I have found this test to be very useful. The hypothesis tested is that a structural change occurred between observations T_1 and T_2, where T_1 is close to 1 and T_2 is close to T. The AP test statistic is a weighted average of the chi-square values for each possible split in the sample period between T_1 and T_2. Asymptotic critical values for this statistic are provided in the AP paper.

If the AP value is significant, which means that the hypothesis of structural stability is rejected, it may be of interest to examine the individual chi-square values to see at what observation the largest value occurred. This is likely to give one a general idea of where the structural change occurred, even though the AP test itself does not pin down the exact date.

I have found that few macroeconomic equations pass all the above tests. If any equation does not pass a test, it is not always clear what should be done. If, for example, the hypothesis of structural stability is rejected, one possibility is to divide the sample period into two parts and estimate two separate equations. The resulting coefficient estimates, however, are not always sensible in terms of what one would expect from theory. Similarly, when the additional lagged values are significant, the equation with the additional lagged values does not always have what one would consider sensible dynamic properties. In other words, when an equation fails a test, the change in the equation that the test results suggest may not produce what seem to be sensible results. In many cases one may stay with the original equation even though it failed the test. My feeling (being optimistic) is that much of this difficulty is due to small-sample problems, which will lessen over time as sample sizes increase, but this is an important area for future research.

III. Testing Complete Models

When testing complete structural models, it is useful to have benchmark models to use for comparison purposes. Vector autoregressive (VAR) models provide useful benchmarks. If the interest is on GDP predictions, however, I have found "autoregressive components" (AC) models to be better benchmarks than VAR models in the sense of being more accurate. An AC model is one in which each component of GDP is

regressed on its own lagged values and lagged values of GDP. GDP is then determined from the GDP identity as the sum of the components. AC models do not have the problem, as VAR models do, of adding large numbers of parameters as the number of variables (components in the AC case) is increased.

Stochastic simulation allows one to compute forecast-error variances. Each repetition consists of draws of the structural error terms and (possibly) the coefficients. Given these draws, given the values of the exogenous variables, and given the initial conditions, the model is solved dynamically over the period of interest, say an eight-quarter period. This gives a solution value of each endogenous variable for each of the eight quarters. If this is done J times (where again J is a number like 500 or 1,000), one has J solution values for each variable and quarter. From these values one can compute means and variances for each variable and quarter.

One might think that forecast-error variances computed in this way could simply be compared across models to see which variances are smaller. There are, however, two additional problems. The first is controlling for different sets of exogenous variables across models (VAR and AC models, for example, have no exogenous variables). This can be done in a variety of ways. One is to estimate autoregressive equations for each exogenous variable and to add these equations to the model. The expanded model can then be stochastically simulated to get the variances. Another way is to estimate in some manner the forecast-error variance for each exogenous variable (perhaps using past errors made by forecasting services in forecasting the variable) and then to use these estimates and the normality assumption to draw exogenous-variable values for the stochastic simulation.

The second problem is the possibility of data mining. A model may have small estimated variances of the structural error terms and small estimated variances of the coefficient estimates (which lead to small forecast-error variances from the stochastic simulation) because it has managed spuriously to fit the sample well. A further step is needed to handle this problem, which is to compare variances computed from outside-sample forecast errors with variances computed from stochastic simulation. If this is done over a number of sample periods, it is possible to estimate adjustments to the forecast-error variances.

If both of these problems are taken care of, the final estimated forecast-error variances have accounted for the four main sources of uncertainty of a forecast—from the error terms, coefficient estimates, exogenous variables, and possible misspecification of the model (i.e., possible data mining)—and so they can be compared across models. More details are given in Fair (1980).

Another way to compare models, discussed in Fair and Robert Shiller (1990), is to regress the actual value of a variable on a constant and the predicted values of the variable from different models. If one model's forecast contains all the information in another model's forecast plus some, then its forecast should be significant in this regression, and the other model's forecast should not. If both forecasts contain independent information, then both should be significant. If neither forecast contains useful information, then neither should be significant. This test is related to the literature on encompassing tests and the literature on the optimal combination of forecasts.

Stochastic simulation can be used to calculate the probability of various events happening. Say one is interested in the probability that within an eight-quarter period at least two successive quarters have negative GDP growth. Draw a set of error terms for the period and solve the model using these draws. Record whether or not there were at least two successive quarters of negative growth for the solution values. This is one repetition. Do J repetitions, and calculate the percentage of the J repetitions in which the event occurred. This percentage is the estimated probability, an estimate that is consistent with the probability structure of the model.

This procedure can be used for testing purposes. It is possible for a given event to

compute a *series* of probability estimates and to compare these estimates to the actual outcomes (which are either 0 or 1). Various measures are available for computing the accuracy of the probabilities, and these measures can be compared across models to see which model's estimated probabilities best reflect the actual outcomes.

I have found that structural models generally do better than VAR and AC models in the above tests. Only limited work has been done, however, on comparing one structural model against another. Much might be learned in the future if more testing of structural models were done.

Note that in the process of testing models one is in effect testing the quantitative importance of the Lucas critique. If coefficients in a model change considerably when a policy variable changes and the model has not accounted for this, the model is misspecified and should not do well in tests.

IV. Analyzing Complete Models

A common way of analyzing macroeconometric models is to compute "multipliers." One or more exogenous variables are changed, and the effects of this change on the endogenous variables are computed. Stochastic simulation can be used to compute standard errors of these multipliers, and this is now a computationally routine matter. If J repetitions are made of a given change, one has J values of the change in each endogenous variable for each observation, and means and standard errors can be computed from these values. Computing standard errors of multipliers is useful because it allows one to gauge how much confidence to place on the results.

Stochastic simulation can be used to examine what a macroeconometric model says about the sources of economic fluctuations. One first computes a forecast-error variance drawing all the error terms and then computes the variance after taking one or more of the structural error terms as fixed. If the fixed error terms are uncorrelated with the other error terms, the difference between the two estimated variances is the amount of variation attributed to the fixed error terms. In practice the correlation of the error terms across equations is usually small, and so the assumption of no correlation is usually not very restrictive.

The optimal choice of monetary-policy instruments is another issue that can be examined using stochastic simulation. William Poole (1970) examined the optimal choice analytically in a stochastic *IS-LM* model, and stochastic simulation allows this to be done in larger models. Forecast-error variances of, say, GDP can be computed first fixing the short-term interest rate and second fixing the money supply, and then the variances can be compared.

Finally, optimal control problems are now fairly easy to solve using large-scale models. If one is willing to assume certainty equivalence even though the model is nonlinear, the optimal control problem can be set up as a standard unconstrained optimization problem, which can be solved by a number of numerical algorithms.

REFERENCES

Andrews, Donald W. K., "Exactly Median-Unbiased Estimation of First Order Autoregressive/Unit Root Models," *Econometrica*, January 1993, *61*, 139–65.

_____ **and Ploberger, Werner,** "Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative," Cowles Foundation Paper No. 1015, Yale University, February 1992.

Duesenberry, James S., Fromm, Gary, Klein, Lawrence R. and Kuh, Edwin, eds., *The Brookings Quarterly Econometric Model of the United States*, Chicago: Rand McNally, 1965.

_____, _____, _____ **and** _____, *The Brookings Model: Some Further Results*, Chicago: Rand McNally, 1969.

Fair, Ray C., "Estimating the Expected Predictive Accuracy of Econometric Models," *International Economic Review*, June 1980, *21*, 355–78.

_____, *Specification, Estimation, and Analysis of Macroeconometric Models*, Cambridge, MA: Harvard University Press,

1984.

_____, "The Cowles Commission Approach, Real Business Cycle Theories, and New-Keynesian Economics," in Michael T. Belongia and Michelle R. Garfinkel, eds., *The Business Cycle: Theories and Evidence*, Boston: Kluwer, 1992, pp. 133–47.

_____, "Testing Macroeconometric Models," unpublished manuscript, 1993.

_____ **and Shiller, Robert J.,** "Comparing Information in Forecasts from Econometric Models," *American Economic Review*, June 1990, *80*, 375–89.

_____ **and Taylor, John B.,** "Solution and Maximum Likelihood Estimation of Dynamic Rational Expectations Models," *Econometrica*, July 1983, *51*, 1169–85.

Hansen, Lars Peter, "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, July 1982, *50*, 1029–54.

Hendry, David F., Pagan, Adrian R. and Sargan, J. Denis, "Dynamic Specifications," in Z. Griliches and M. D. Intriligator, eds., *Handbook of Econometrics*, Amsterdam: North-Holland, 1984, pp. 1023–1100.

Klein, Lawrence R., *Economic Fluctuations in the United States, 1921–1941*, New York: Wiley, 1950.

_____ **and Goldberger, Arthur S.,** *An Econometric Model of the United States 1929–1952*, Amsterdam: North-Holland, 1955.

Lucas, Robert E., Jr., "Econometric Policy Evaluation: A. Critique," in K. Brunner and A. H. Meltzer, eds., *The Phillips Curve and Labor Markets*, Amsterdam: North-Holland, 1976, pp. 19–46.

Parke, William R., "An Algorithm for FIML and 3SLS Estimation of Large Nonlinear Models," *Econometrica*, January 1982, *50*, 81–95.

Poole, William, "Optimal Choice of Monetary Policy Instruments in a Simple Stochastic Macro Model," *Quarterly Journal of Economics*, May 1970, *84*, 197–216.

Tinbergen, Jan., *Statistical Testing of Business Cycle Theories*, Geneva: League of Nations, 1939.