Forecasting With Econometric Methods: A Comment

Preston J. Miller

January 1978

Working Paper #: 104

PACS File #: 1520

The author is Associate Director of Research and Senior Economist, Federal Reserve Bank of Minneapolis.

The author is grateful to Thomas Sargent and Art Rolnick for their many helpful comments. The discussion of the failure of econometric models is essentially lifted from Sargent's "Is Keynesian Economics a Dead End?"

I have long smelled a rat with respect to econometric forecasting, and I am glad Professor Armstrong finally flushed one out. In addition to synthesizing the sciences of economics and animal psychology, Armstrong raises these two major points in his paper:

1. Forecasts from econometric models are not more accurate than forecasts from simple mechanical schemes or autoregressions.
2. Econometricians still believe they are, contrary to the evidence.

Armstrong's focus on forecasting accuracy is a bit misdirected, however, because it lets escape the biggest rat of all: the failure of econometric models to remain stable over time. Stability of econometric models is required both to measure forecasting accuracy and to perform policy simulations--the task for which econometric models were primarily intended.

I will comment first on Armstrong's two major points concerning forecasting accuracy of econometric models and then offer an explanation of their failure to remain invariant over time. The implications I draw for the direction of future research differ considerably from Armstrong's.

Most, if not all, of the studies Armstrong cites in support of the first point have two defects: they utilize questionable measures of accuracy, and the forecasts from econometric models which they analyze are really hybird mixtures generated from judgmental and econometric techniques.

In order to clarify the meaning of forecasting accuracy, it is useful to formally specify the underlying statistical problem. To begin, suppose we are interested in obtaining a forecast of the value of a particular variable in a particular time period; say GNP, one quarter hence. Let us suppose the loss function is the square of the forecast error. We have a number of alternative econometric and time series models which can be used to generate the forecast. The problem is to choose the model from those available to minimize the expected value of the loss function conditional on all current information. The expected value of the loss function for any given model is just the (square of the) model's standard error of forecast. Therefore, our statistical problem would seem to be to choose the model which minimizes the standard error of forecast.

There is more to it than that, however. The standard error of forecast is an estimated measure of the distribution of forecast errors over a historical period; in choosing among models, we also must consider the probability that the future forecast errors will be drawn from the same historical distributions. Here is where the issue of stability enters. If a model fails a test for stability over historical subperiods, we can reject the hypothesis that the distribution of forecast errors for the future period is the same as for the sample period. Therefore, all models should be tested first for stability over time, since the standard error of forecast is not a meaningful measure for models which fail such tests.[1/] Among models which pass tests for stability over time, we still might want to consider the number of periods over which the models pass the tests. We would attach a higher probability, for instance, to a model's holding up one more period if it had not failed a test for structural change over the previous one hundred periods rather than over just the previous ten.

[1/] Muench-Rolnick-Wallace-Weiler [5] is the only study in the literature of which I am aware that tests for stability over time of multiequation models. The test devised in this study is a multiequation version of a Chow test. Coefficients and residuals are randomly drawn from a jointly normal distribution with variances and covariances taken from sample estimates. Exogenous variable values are generated from autoregressions.

The studies Armstrong cites do not test models for stability over time, and they use root mean square error (RMSE) as the measure of forecast accuracy. For a model which fails tests for stability over time, the value of the RMSE statistic will be sensitive to the period over which it is computed, and its meaning is far from clear.

Even if the models are stable over time, the use of RMSE still does not permit a valid comparison of models which incorporate different sets of exogenous variables. First, the distribution of the RMSE is not known. The RMSE adds up residual errors over different periods when the values of exogenous variables are allowed to change. The distribution of residual errors in any period is a function of exogenous variable values, however, so that the RMSE essentially adds up apples and oranges.[2/] Second, when RMSE's of ex-ante forecasts (those using assumed values of exogenous variables) are compared across different models, we do not know whether the RMSE of one model is lower than another because the model is better or because the judgment of the model's econometrician was better in choosing future values of exogenous variables. Finally, when RMSE's of ex-post forecasts (those using actual values of exogenous variables) are compared across different models, we are attributing to the econometricians knowledge of more information than they actually had at the time the forecasts were made. This gives a relative advantage to models which attribute a large share of the determination of dependent variables to exogenous variable movements. In the extreme, the model X = C + I + G would do pretty well ex-post if C, I, and G were considered exogenous. Thus, to measure accuracy across models the forecasts of

---

[2/] Rolnick [6] page 439.

exogenous variables must also be explicitly modeled. The exogenous variables could be modeled as a vector autoregression, for example, and the hypothesis that these variables are exogenous as a block to the set of endogenous variables could be tested.[3/]

The second defect of the studies Armstrong cites in support of the first point is that most of the "econometric" model forecasts analyzed incorporate quite a bit of judgment. Most econometricians working with these models adjust the intercepts of equations by varying amounts over the periods of the forecast horizon. These adjustments are functions of the assumed values of exogenous variables and the outcome for endogenous variables that the econometricians expect. This last assertion is certainly consistent with the Haitovsky-Treyz finding that the ex-ante forecasts of the Wharton model, for instance, outperform the ex-post forecasts based on the same intercept adjustments.[4/] If the econometricians had known the actual values of exogenous variables beforehand, they would have made different intercept adjustments.

I think Armstrong's second point that respected econometricians believe econometric methods of forecasting are more accurate than time series methods is quite interesting, given all the evidence to the contrary. I have some doubt, however, that the beliefs of the sample of respondents are representative of those of the universe of econometricians. Not only is Armstrong's sample extremely small (21), but apparently no one from a school which emphasizes time series analysis (Chicago, Wisconsin,

---

[3/] See Sargent [10] pages 455-464 for a discussion of the theory of vector autoregression models and Sargent-Sims [11] for an application of this theory.

[4/] Haitovsky-Treyz [1] page 320.

Minnesota, etc.) was included. As Armstrong suggests, those who have a stake in econometric methods are likely to defend them, and those seem to be the ones sampled.

The reason I believe Armstrong let the big rat get away, though, is that econometric model builders might not be too upset with studies which show simpler methods of forecasting do better. They would agree that a reduced form model with few restrictions should forecast better than a structural model with over-identifying restrictions.[5/] They would argue, however, that the advantage of structural models is they can be used for policy control experiments. But can they? Here is where the issue of stability assumes fundamental importance. If the distribution of a model's coefficients shift in some unknown and unpredictable way every time a different policy is imposed, we can have no confidence in the predicted outcomes from the model's policy simulations.[6/] There are strong theoretical arguments for suspecting distributions of coefficients in econometric models to shift when policy changes, and there is an accumulating body of empirical evidence which rejects the invariance of such models over time periods when policy rules changed.[7/] A necessary condition for a model to be stable over time is that each equation be stable over time. Yet, individual equations in these models fail Chow tests dramatically. They always are being judgmentally adjusted or formally reestimated. This failure of econometric models to be stable is more troublesome than the finding that they are outperformed by simpler methods of forecasting.[8/]

[5/] Rolnick [6] page 439.

[6/] Lucas [3] and Wallace [14].

[7/] See, for example, Sargent [7] pages 76-77.

[8/] Koopmans emphasized the importance of stability in his debate with Friedman in [2].

Given enough observations, we know that the best linear least squares forecast of a given variable is a regression of it unconstrained against itself and all other variables over all lags. The problem is that we do not have enough observations, so that we quickly run out of degrees of freedom. To conserve on degrees of freedom we must generate restrictions across stochastic processes. Thus, the name of the game in economics and econometrics is to generate valid restrictions on the data. And it is here where econometric models have failed.

Econometric models have fared badly because they combine poor economic theory with poor econometric methods to derive invalid and mutually inconsistent restrictions. The main types of a priori restrictions imposed in these models are:

1. setting coefficients to zero,
2. sorting variables into exogenous and endogenous categories, and
3. limiting the orders of serial correlation and cross-serial correlation of the disturbance terms.

These restrictions should be deduced from theoretical and statistical modeling; they are not up for grabs to be determined by the model builder's judgment.

The a priori setting of coefficients to zero in econometric models contradicts some basic tenets of general equilibrium theory.[9/] General equilibrium models imply that individuals' excess demand functions should all depend on the same arguments: current and expected values of endowments, prices, and policy parameters. Moreover, as long as agents

---

[9/] See Miller [4] and Wallace [14] for a more extensive discussion of the internal inconsistencies found in large-scale macro models.

optimize, expectations should be rational in the sense that they are unbiased forecasts subject to available information. If we examine large scale econometric models, we find, for example, that consumption demand depends on one set of variables, money demand on another, and labor supply on yet another. No consistent model would ever imply these variable exclusion restrictions. We also find that the arguments of these demand and supply functions are likely to depend on flow variables such as income. We know, though, that agents are confronted with wages and prices, and income is determined by their labor-leisure decision. Finally, expectations are expressed in these models as simple, short lags of a few variables, but these expectations schemes are not consistent with the behavior of optimizing agents. Consumption demand functions in these models, for example, are claimed to be derived from permanent income theory. Yet because they model income expectations as a simple lag of past income, they cannot even distinguish between the demand effects of a permanent and a temporary tax change.

The *a priori* sorting of variables into exogenous and endogenous variables is unjustified on statistical grounds.[10/] The hypothesis of statistical exogeneity can be tested. Sims has shown in the bivariate case that this hypothesis implies restrictions on the cross correlations of the two time series in question.[11/] Skoog has extended Sims' results to multivariate models.[12/]

Finally, econometric models severely restrict the orders of serial correlation and cross-serial correlations of the error processes. Usually, the errors of any equation are assumed to exhibit at most

---

[10/] Sargent [9].

[11/] Sims [12].

[12/] Skoog [13].

first-order serial correlation, and they are assumed to be independent of past errors in other equations. Neither theory nor time series analysis suggests that these restrictions are likely to be satisfied.

The instability of econometric models cannot be attributed entirely to inappropriate restrictions; there must also be nonstationarities in real world stochastic processes. If the economy could be characterized by a stationary stochastic process, then no model would ever fail a test for stability over time.[13/] That is because a model which imposes an invalid restriction will do no worse in the post-sample period than it did in the estimation period. A major challenge of economics is to produce restrictions across stochastic processes which remain valid even when the processes are nonstationary. Current econometric models do not meet this challenge, and I doubt that naive autoregressions do either.

To meet the challenge I have posed requires better theory and better econometric techniques than are typically employed. Even with my strong indictment of econometric models, I would answer Armstrong's survey question by saying that although econometric methods do not provide more accurate forecasts than mechanical or naive time series techniques, they can. That is, they can if the econometric restrictions in the models are derived from theories of individual optimizing behavior in an explicitly defined dynamic stochastic setting. This is no longer a hope for the future. It is a current reality. Sargent has employed this methodology to a study of the labor market.[14/] I believe

---

[13/] This claim must be qualified because it holds only in large samples.

[14/] Sargent [8].

this to be the first successful attempt to derive restrictions across stochastic processes based on a model of optimizing behavior. I believe Sargent's approach is the right direction for future modelling efforts because it promises to produce restrictions which will hold up over time and under policy interventions.

Bibliography

[1] Haitovsky, Yoel, and George Treyz. "Forecasts With Quarterly Macroeconometric Models: Equation Adjustments, and Benchmark Predictions: The U.S. Experience." The Review of Economics and Statistics, Vol. 54, No. 3 (August 1972), 317-325.

[2] Koopmans, Tjalling C. "Comment." In Conference on Business Cycles. New York: National Bureau of Economic Research, Inc., 1951, 141-145.

[3] Lucas, Robert E., Jr. "Econometric Policy Evaluation: A Critique." In The Phillips Curve and Labor Markets, ed. by K. Brunner and A. Meltzer, Vol. 1 of the Carnegie-Rochester Conferences on Public Policy, a supplementary series to the Journal of Monetary Economics. Amsterdam: North-Holland (1976), 19-46.

[4] Miller, Preston J. "On Macroeconomic Theories and Models." A Prescription for Monetary Policy: Proceedings From A Seminar Series. Minneapolis: Federal Reserve Bank of Minneapolis (December 1976), 43-50.

[5] Muench, Thomas J., Arthur J. Rolnick, Neil Wallace, and William Weiler. "Tests for Structural Change and Prediction Intervals for the Reduced Forms of Two Structural Models of the U.S.: The FRB-MIT and Michigan Quarterly Models." Annals of Economic and Social Measurement, Vol. 3, No. 3 (July 1974), 491-519.

[6] Rolnick, Arthur J. Book review of Development of the Monetary Sector: Prediction and Policy Analysis in the FRB-MIT-Penn Model by J. Phillip Cooper. Lexington, Massachusetts: Lexington Books, 1974. In Journal of Political Economy, Vol. 83, No. 2 (April 1975), 438-441.

[7] Sargent, Thomas J. "Testing for Neutrality and Rationality." A Prescription for Monetary Policy: Proceedings From A Seminar Series. Minneapolis: Federal Reserve Bank of Minneapolis (December 1976), 65-86.

[8] ________. "Estimation of Dynamic Labor Demand Schedules Under Rational Expectations." Staff Report No. 27, Federal Reserve Bank of Minneapolis (September 1977).

[9] ________. "Is Keynesian Economics A Dead End?" Working Paper No. 101, Federal Reserve Bank of Minneapolis (October 1977).

[10] ________. Notes on Macroeconomic Theory. Book forthcoming, Academic Press.

[11] Sargent, Thomas J. and Christopher A. Sims. "Business Cycle Modeling Without Pretending to Have Too Much A Priori Economic Theory." New Methods in Business Cycle Research: Proceedings From A Conference. Minneapolis: Federal Reserve Bank of Minneapolis (October 1977), 45-110.

[12] Sims, Christopher A. "Money, Income, and Causality." American Economic Review, Vol. 62, No. 4 (September 1972), 540-552.

[13] Skoog, Gary R. "Causality Characterizations: Bivariate, Trivariate, and Multivariate Propositions." Staff Report No. 13, Federal Reserve Bank of Minneapolis (November 1976).

[14] Wallace, Neil. "Microeconomic Theories of Macoreconomic Phenomena and Their Implications for Monetary Policy." A Prescription For Monetary Policy: Proceedings From A Seminar Series. Minneapolis: Federal Reserve Bank of Minneapolis (December 1976), 87-98.