Eastern Economic Journal,Vol. VIII, No. 4, October–December 1982

# A Note on the Informational Contents of Alternative Forecasting Benchmarks

JACK NARAYAN and JAMES CICARELLI

According to Mincer and Zarnowitz (1969), the ultimate test of an econometric forecasting model is how well it performs relative to the optimal naive benchmark, defined as the mechanical extrapolation with the minimum mean square error (MSE). The justification for this criterion is clear: since the optimal naive benchmark is available to all forecasters, the accuracy of any forecasting model depends on the additional information it contains relative to this extrapolative model.

Until the popularization of the Box-Jenkins approach, simple linear autoregressive models usually served as crude proxies of the optimal naive benchmark. Against these standards, econometric forecasting models generally perform well (McNees (1974)), although there have been instances to the contrary (Pfaff (1977)). However, increased acceptance of univariate ARIMA models as more appropriate proxies of the optimal benchmark have diminished the comparative advantage of econometric forecasting over mechanical extrapolations. Occasionally, econometric models have outperformed or proven equivalent to ARIMA benchmarks (Levenback, Cleary, and Fryk (1974); Spivey and Wrobleski (1978); Trivedi (1975)), but by and large the reverse has been true (Nelson (1972); Granger and Newbold (1974); Narasimhan, Castellino, and Singporwalla (1974); Ibrahim and Otsuki (1976); Cicarelli and Narayan (1980)).

Department of Mathematics and Department of Economics respectively. SUNY-Oswego, Oswego, N.Y. 13126

The authors would like to thank C. W. J. Granger, C. R. Nelson, and an anonymous referee for their comments on an earlier draft of this paper.

The econometric community has offered a variety of explanations for this relatively poor performance (Chow (1978); McNees (1978); Zellner (1978); Zarnowitz (1978); Howrey (1978); Su (1978)), including the claim that the apparent superiority of ARIMA models stems, in part, from the use of ex post instead of ex ante benchmarks. The gist of this argument is easily summarized.

An ex post ARIMA benchmark is one constructed using all available data, even that which is to be forecast, for the purpose of comparison with an econometric model. By contrast, an ex ante ARIMA benchmark would be one whose comparison forecasts are based on a model developed using only observations up to the point of forecast even though subsequent data points are known. Despite occasional exceptions (Spivey and Wrobleski (1978), most ARIMA benchmarks are of the ex post variety, and thus involve the utilization of information unavailable when econometric forecasters issue their predictions. It is the additional information and not the inhererent superiority of the ARIMA model which is responsible for the comparative success of these benchmarks. If, the argument continues, ex ante and not ex post ARIMA models were used as benchmarks, the relative superiority of ARIMA models over econometric models would decline appreciably, if not vanish outright. The purpose of this paper is to test that hypothesis.

## The Analytical Framework

Asserting that the superiority of ARIMA models would dissipate if ex ante rather than ex post benchmarks were used is equivalent to saying that the latter contain more information than the former. So, instead of testing the above hypothesis directly by comparing the forecast performance of various econometric models to that of ex post and ex ante ARIMA benchmarks, an indirect approach involving the comparison of ex post and ex ante ARIMA models will be utilized. This indirect approach is superior to the direct approach on two counts which are both related to limitations of publically-available data. First, although econometric forecasters provide private subscribers with predictions on a number of variables, publically-available forecasts such as those in the Conference Board's *Statistical Bulletin* are restricted to a few widely publicized time series, notably current and real GNP and the rate of unemployment. The direct comparison approach would, of necessity, be limited and so would the results. Second, publically available forecasts of different econometric models cover varying periods of time, some of very short duration. Hence the data sets available for direct comparisons would not be identical, and in some instances would be insufficient to perform the test of significance employed in this paper. The indirect approach overcomes these limitations in that it allows for the comparison of a variety of time series over a long and identified period of time, producing results of a general nature.

The outcomes of the indirect comparison of ex post and ex ante ARIMA models represents a test of the stated hypothesis in the following sense. If, on the basis of conventional measures of forecasting performance, this comparison shows ex post benchmarks to be more accurate than ex ante ones, then the hypothesis will have some merit. However, if the comparison reveals no difference in informational contents, then econometric forecasters will have to look for other explanations for their poor performances.

A number of different measures such as the Theil inequality index (Theil (1966)) or the Mincer-Zarnowitz relative mean square error (Mincer and Zarnowitz (1969)) can be used to determine the relative accuracy of one forecasting model compared to another. These indexes, however, only measure qualitative superiority in that they indicate which predictor is individually better but not by how much nor whether the difference is statistically significant. Thus, even though one forecasting model is superior to another, the difference in accuracy may be statistically insignificant, i.e., the informational contents of the two models may be statistically identical. In this paper, both quantitative and qualitative measures of relative forecasting accuracy will be used to test the hypothesis that ex post ARIMA benchmarks contain more information than ex ante ARIMA benchmarks. The results will indicate which benchmark is superior and whether or not the difference in predictions is statistically significant.

## Qualitative Evaluation

Listed in Table I are ten quarterly time series drawn from *Business Conditions Digest*. Although less than exhaustive, this set of time series represents a cross section of the variables econometric forecasters track when attempting to predict the future of the economy or important sub-sections thereof, and thus is sufficiently rich for the purpose of evaluating ex post versus ex ante ARIMA benchmarks.

To determine the relative informational contents of alternative benchmarks, the last forty quarters of each series in Table I were forecast using an ex post and ex ante ARIMA model, then those forecasts were compared. For all ten series, the ex post benchmarks were constructed using conventional Box-Jenkins techniques (Nelson (1973), Chp. 5) applied to a minimum of 92 data points for each series (basically the period from the early 1950s through 1977). In those instances where two or more ARIMA models appeared as equivalent predictors of a series, that model with the minimum MSE over the forty forecast quarters was selected as the optimal ex post ARIMA benchmark.



TABLE 1. Quantitative and Qualitative Measures of Relative Informational Contents of Alternative Forecasting Benchmarks

| *Quarterly Time Series* | *Dimensions of ARIMA Models* | *Modified Theil Index* | *Correlation Coefficient (with t-statistic)* |
|---|---|---|---|
| Bank Rates on Short Term Business Loans | (0, 1, 1) | .9912 | −.057 (−.35) |
| Changes in the Money Supply (M1) | (0, 1, 1) | .9422 | −.183 (−1.15) |
| Federal Funds Rate | (0, 1, 1) | .9797 | −.223 (-1.41) |
| Index of Help-Wanted Ads in Newspapers (1967 = 100) | (2, 1, 1) | .9435 | −.364 (−2.41)* |
| Index of Stock Prices; 500 Common Stocks (1967 = 100) | (0, 1, 1) | .9756 | −.170 (−1.12) |
| New Housing Starts | (2, 1, 0) | .9578 | −.372 (−2.83)* |
| Rate of Capacity Utilization; Manufacturing | (0, 1, 1) | .9868 | −.111 (−.68) |
| Real GNP | (1, 1, 0) | .9462 | −.399 (−2.75)* |
| Unemployment Rate | (0, 1, 1) | 1.0649 | .148 (0.97) |
| Value of Manufacturers New Orders; Durable Goods Industries | (1, 1, 1) | .9738 | −.261 (−1.67) |

*Statistically significant at the 5% level.

The initial ex ante models were constructed using Box-Jenkins applied to the same set of data points in a series except for the most recent forty quarters. Ideally, the ex ante models should be estimated with preliminary data available to forecasters for each successive forecast. As a practical matter, this is not possible because of the enormous increase in computations involved. Instead, the ex ante models were estimated using revised figures. Once the initial estimates were made, the models were continually revised as successive quarters became part of a series history. While the dimensions of the ex post and ex ante benchmarks were found to be the same for all series, the parameters of the ex ante models varied over time while those of the ex post models were, of course, constant.

As noted earlier, a number of measures can be utilized when evaluating the performance of one forecasting model with that of another. Since the MSE is a preferred indicator of accuracy because (a) it reflects the magnitude of forecasting error and (b) squaring means larger errors carry proportionally greater weight than smaller ones, a modified Theil index

$$+\sqrt{\frac{\text{MSE ex post ARIMA benchmark}}{\text{MSE ex ante ARIMA benchmark}}}$$

was used to evaluate the relative (qualitative) performance of ex post and ex ante benchmarks. If this ratio is less than, equal to, or greater than unity, then the ex post ARIMA benchmarks is superior, equivalent, or inferior to the ex ante ARIMA benchmark.

**Table I**

Table I indicates that the ex post benchmarks generally contain more information than the ex ante benchmarks. In only one instance—the Unemployment Rate—is the reverse true. These results support the hypothesis being tested, and suggest that the relative accuracy of econometric models vis-a-vis univariate ARIMA models would improve if ex ante instead of ex post benchmarks were used. The question then becomes "by how much?" We expect the one-quarter ahead ex post ARIMA forecasts to approximate the last forty quarters better than the one-quarter ahead ex ante forecasts simply because the model generating the ex post estimates has been fitted to these observations by a least squares criterion. The real issue is the magnitude of the improvement and whether or not it is statistically significant.

**Quantitative Evaluation**

The matter boils down to a quantitative evaluation of forecasts. If the informational contents of ex post benchmarks is significantly greater than that of ex ante benchmarks in a statistical sense, then the gain in the relative accuracy of econometric models could be substantial. If, however, the difference is not statistically significant, the gain in accuracy would be minimal. Additionally, the latter result would imply that ex post rather than ex ante benchmarks are superior since they contain the same amount of information in a statistical sense but cost less to produce because their parameters need only be estimated once.

Given the problem at hand, it would be tempting to use an F-test to determine if the informational contents of ex post benchmarks differs significantly from that of ex ante benchmarks, but Granger and Newbold (1977 p. 281) note that this would be inappropriate and recommend the following alternative test. Let $e_t^p$ and $e_t^a$, where $t = 1, 2, \ldots, n$, be the forecast errors of the ex post and ex ante ARIMA benchmarks, respectively. Assume that $(e_t^p, e_t^a)$ constitutes a random sample from a bivariate normal distribution with mean zero, variances $\sigma_p^2$ and $\sigma_a^2$, and correlation coefficient $\rho$. In particular, then, it is assumed that the individual forecasts are unbiased and the forecasts errors are not autocorrelated. Consider the pair of random variables $(e^p + e^a)$ and $(e^p - e^a)$. Then

$$E[(e^p + e^a)(e^p - e^a)] = \sigma_p^2 - \sigma_a^2.$$

Thus the two error variances, and hence, given the assumption of unbiasedness, the two expected squared errors, will be equal if and only if this pair of random variables is uncorrelated. Under the stated assumption, the standard test for zero correlation based on the sample correlation coefficient $r$ can be used to test the equality of expected squared forecast errors. Lehmann (1959) shows this test is uniformly most powerful unbiased. The statistical significance of $r$ is based on the $t$-ratio given by $t = r(n-2)^{1/2}/(1-r^2)$.

The hypotheses under which the test is valid were verified. First, one step ahead forecasts from the ARIMA model are unbiased (Box and Jenkins, 1970). An application of a test for normality (Ryan and Joiner) showed that the individual error series was normally distributed. Finally tests for autocorrelation on the individual error series (Box and Jenkins, 1970) indicated no significant autocorrelation. In short the distribution of the residuals was characterized as zero-mean white noise.

The correlation coefficients for the time series considered in this study are shown in Table I along with the $t$-ratio for each coefficient. In three of the ten cases, the informational contents of the ex post and ex ante ARIMA benchmarks differed significantly, and in these three instances, the ex post benchmark is superior to its ex ante counter part. Thus, when the accuracy of the ex post benchmark was greater than that of the ex ante benchmark, it was significantly greater one-third of the time. While this does not confirm beyond a reasonable doubt the hypothesis advanced earlier, it certainly gives it considerable credibility.



**Concluding Remarks**

Although derived from a relatively small sample and not necessarily invariant with respect to time, the results of this study suggest that use of ARIMA benchmarks for the purpose of evaluating the performance of econometric forecasting models must be consistent, i.e., ex post and ex ante econometric forecasts should be evaluated by ex post and ex ante benchmarks, respectively. To utilize ex post benchmarks to evaluate ex ante econometric forecasts is to assume that the relative accuracy of an econometric model is independent of the nature of the benchmark against which it is being compared. Clearly, there is no reason *a priori* to presuppose this to be the case. The increased use of ARIMA benchmarks is motivated by the desire to tighten the evaluation procedure applied to econometric forecasting models. The idea of heightened stringency is a two-edge sword and those who would apply it to econometric models must first apply it to their own work.

**References**

Box, G. E. P., and Jenkins, G. M., (1970) *Times Series Analysis*, San Francisco: Holden-Day

Chow, G. C., (1978) "Are econometric methods useful for forecasting?" *Journal of Business*, 51(4): 565–68.

Cicarelli, J., and Narayan, J., (1980) "The performance of eleven economic forecasting models in the 1970's." *Business Economics*, 15(4): 12–16.

Granger, C. W. J., and Newbold, P., (1974) "Economic Forecasting: the Atheist's Viewpoint," in G. A. Renton, editor, *Modelling the Economy*, London, Heinemann.

——, and ——, (1977) *Forecasting Economic Time Series*, New York: Academic Press.

Howrey, E. P., (1978) "The Use of Preliminary Data in Econometric Forecasting." *Review of Economics and Statistics*, 60(2): 193–200.

Ibrahim, I. B., and Otsuki, T., (1976) "Forecasting GNP Components Using the Method of Box and Jenkins." *Southern Economic Journal* 42(3): 461–70.

Lehmann, E. L. (1959) *Testing Statistical Hypotheses*. New York: J. Wiley & Sons.

Levenback, H., Cleary, J. P., and Fryk, D. A., (1974) A comparison of ARIMA and econometric models for telephone demand. *Proceedings of the American Statistical Association: Business and Economics Statistics Section*, p. 448–50.

Mincer, J., and Zarnowitz, V., (1969) "The Evaluation of Economic Forecasts," in J. Mincer, editor, *Economic Forecasts and Expectations*, New York: National Bureau of Economic Research.

McNees, S. K., (1974) "How Accurate Are Economic Forecasts?" *New England Economic Review*, (Nov./Dec.): 2–19.

——, (1978) "Are Econometricians Useful: Folklore Versus Fact." *Journal of Business*, 51(4): 573–77.

Narasimhan, G., Castellino, V. F., and Singpurwalla, N. D., (1974) On the predictive performance of the BEA quarterly econometric model and a Box-Jenkins type ARIMA model. *Proceedings of the American Statistical Association: Business and Economic Statistics Section*, p. 501–4.

Nelson, C. R., (1972) The predictive performance of the FRB-MIT-PENN model of the U.S. economy. *American Economic Review*, 62(5): 902–17.

——, (1973) *Applied Time Series Analysis*, San Francisco: Holden-Day.

Pfaff, P., (1972) "Evaluation of Some Money Stock Forecasting Models." *Journal of Finance*, 32(5): 1639–46.

Ryan, Jr., T. A., and Joiner, B. C., Normal probability plots and tests for normality, *Technical Report;* Pennsylvania State University.

Spivey, W. A., and Wrobleski, W. J., (1978) "An Analysis of Forecast Performance." *Proceed-*

*ings* of the American Statistical Association: Business and Economic Statistics Section, p. 31–40.

Su, V., (1978) "An Error Analysis of Econometric and Noneconometric Forecasts." *American Economic Review, Papers and Proceedings,* 68(2): 306–12.

Theil, H. (1966) *Applied Economic Forecasting.* Amsterdam: North Holland.

Trivedi, P. K., (1975) "Time series versus structural models: a case study of Canadian manufacturing inventory behavior." *International Economic Review,* 16(3): 587–608.

Zarnowitz, V., (1978) On the accuracy and properties of recent macroeconomic forecasts. *American Economic Review, Papers and Proceedings,* 68(2): 313–19.

Zellner, A., (1978) "Folklore versus Fact in Forecasting With Econometric Models." *Journal of Business,* 51(4): 587–93.