Forecasting With Statistical Models and a Case Study of Retail Sales

By Dan M. Bechter and Jack L. Rutner

Methods of economic forecasting have become increasingly elaborate. Highly refined statistical techniques are now being used to extract information from historical data and to project future values of economic variables. To a large extent, these advances in the science of economic forecasting have been made possible by progress in computer technology. But high-speed computers and sophisticated statistical techniques do not provide perfect forward vision. There is a lot of truth to the observation that economic forecasting is more art than science. It remains to be seen just how much the forecasting of economic variables can be improved by strengthening only the more scientific aspects of this activity.

This article has two purposes. The first is to review various approaches to economic forecasting, including a relatively new technique as well as traditional methods. The second is to report on a case study in which the performances of alternative ways of forecasting retail sales are compared.

FORECASTING MODELS

Many forecasters depend heavily on models to help in forecasting. A model consists of mathematical expressions, or equations, which describe relationships among economic variables. A forecaster's choice of a forecasting model is of key importance. A model that contains the wrong variables, or that incorrectly specifies relationships among variables, will be of little use in forecasting.

Economic Models

Economic theory usually provides a good guide to the selection of variables and the relationships for a model's equations, and a model based on theory is sometimes called an "economic model." For example, suppose a forecaster wants to predict retail sales. Since retail sales are closely associated with consumer spending, economic theory suggests that the dollar volume of retail sales during any period may be largely explained by the levels of personal income and personal wealth in that period. It is not realistic, however, to believe that changes in retail sales will always behave strictly in accordance with changes in income and wealth. Deviations will result from factors such as omitted variables (for example, unseasonable weather) and other considerations that are essentially random in their effects upon retail sales. The reasons for these deviations are not explained in economic models, but allowance is made for them by adding a

disturbance term, or error term, to the assumed relationship.

To illustrate, the relationship between retail sales and other variables could constitute an economic model that can be expressed mathematically as:[1]

$$S_t = a + bI_t + cW_t + u_t \tag{1}$$

where S_t = retail sales during period t
I_t = personal income during period t
W_t = personal wealth during period t
u_t = error term during period t
a, b, c = unknown constants.

The model in equation (1) states that the variable, retail sales, is determined by the variables, personal income and wealth; that the relationship is defined basically by the parameters a, b, and c; and that the relationship is inexact, requiring the inclusion of an error term. The variable, retail sales, is referred to as an endogenous variable because it is being explained and is to be forecast. Income and wealth are exogenous variables because they are being used to explain retail sales and are not to be forecast.

The unknown constants, or parameters, must be estimated by reference to data for some particular historical period. The estimation procedure usually used, linear regression, determines values for the parameters a, b, and c that give the best fit of retail sales to personal income and personal wealth over the estimation period selected. In its estimated form, the economic model can be expressed as:

$$\hat{S}_t = \hat{a} + \hat{b}I_t + \hat{c}W_t \tag{2}$$

where the symbol ^ denotes estimated values of the variables or parameters. In equation (2), retail sales in any period is expressed in terms of the actual values of personal income and personal wealth in that period, and numerical estimates of the parameters. The actual value of retail sales in any month will usually differ somewhat from its estimated value, and this difference is the value of that period's error term.

After the model is estimated, it may be used for forecasting. Forecasting with the estimated model is accomplished by solving the equation for the variable to be forecast after plugging in the appropriate period's values for the exogenous variables.

An estimated economic model of the type shown in (2) may not be particularly well suited for forecasting. Its principal drawback is that the values of the explanatory variables, I and W, would themselves have to be forecast before S could be forecast. One way around this problem is to choose a model in which current values of the variable to be explained depend on past, or lagged, values of the explanatory variables. Fit in this fashion, the estimated model might be:

$$\hat{S}_t = \hat{a} + \hat{b}I_{t-1} + \hat{c}W_{t-1}. \tag{3}$$

From relationship (3), it follows that next period's retail sales (S_{t+1}) can be forecast by

[1] In the absence of any good reason to believe otherwise, it is usual to assume that changes in the variable to be explained (in this example, retail sales) are proportional to changes in the explanatory variables (personal income and personal wealth). This assumption implies a linear relationship in which retail sales in any month are equal to some constant, plus some constant proportion of personal income in that month, plus some constant proportion of personal wealth in that month. More complicated models could, however, suggest a nonlinear relationship such as retail sales being a function of the items in (1) plus a sine curve of time and a constant. This formulation of (1) would require nonlinear estimation procedures.

using this period's personal income and personal wealth.

The use of lagged explanatory variables, besides being helpful in forecasting, also has some justification in theory. For example, retail sales may not react quickly to changes in current income because individuals may be slow in changing their spending behavior. In recognition of how some economic behavior may be better described by a weighted average of past values of certain variables, it is common for an equation in an economic model to include lags of different lengths for the same variable.

Economic models often consist of more than one equation. Indeed, some large models contain hundreds of relationships among variables. As an illustration, the single-equation economic model given by equation (1) might be expanded to a two-equation model in which personal income, as well as retail sales, are endogenous variables:

$$S_t = a + bI_t + cW_t + u_{1t} \quad \text{(4a)}$$

$$I_t = d + eN_t + u_{2t} \quad \text{(4b)}$$

where N_t = labor input, an exogenous variable, as well as personal wealth.

A system of equations such as (4a, b) is generally referred to as "structural" in that these equations describe how a particular segment of the economy operates according to a structure consistent with economic theory. In the structural model (4a, b), retail sales depend ultimately on wealth and labor input, the exogenous variables. Moreover, in general, for any structural model, the endogenous variables depend ultimately on the exogenous variables. When endogenous variables are expressed as depending only on exogenous variables, the model is referred to as a reduced form model.'

Time Series Model

A second type of forecasting model is constructed solely from the past values of the variable to be forecast. This type of model may be termed a "single-variable time series" model. A very naive application of this type of model is to forecast the value of a variable in the next period to be the same as it is in the current period. If the variable to be forecast has some trends and cycles in it, a better naive forecast may be achieved by forecasting next period's change in the value of a variable to be equal to the most recent change in its value. A somewhat more sophisticated, but still naive, single-variable time series model is the commonly used time-trend forecasting model, in which next period's value of the variable of interest is forecast to lie along a trend line, fitted by eye or by regression techniques to past values of the variable.

In recent years, significant advances have been made in the development of certain types of single-equation time series models known collectively as "autoregressive" models. Such forecasting models are purely self-determining: the variable to be forecast is related only to its past values, plus an error term. In its simplest

---

2 Economic forecasting models need not be relationships justified by economic theory. Besides economic models, there are other types of models that may be used for forecasting purposes. One such type is the "expectations" model, in which the explanatory variables are indicative of the intentions or mood of the people whose actions determine the value of the variable to be forecast. For example, if the forecaster is interested in next month's retail sales, he may choose indexes of consumer confidence and consumer buying plans for explanatory variables in his expectations model. Although the expectations approach provides an interesting alternative to economic theory in model building, it is not considered further here.

unrefined form, an autoregressive model for forecasting retail sales would be expressed as

$$S_t = a + bS_{t-1} + u_t, \quad (5)$$

where, as before, S_t represents retail sales in month t, a and b are parameters, and u_t is the error term.[3]

One of the most sophisticated forms of autoregressive models is the ARIMA model. The acronym ARIMA stands for "autoregressive integrated moving average," which describes the model. The first term, autoregressive, has already been defined to mean a model in which a variable is a function of only its past values except for deviations introduced by an error term. "Integrated" indicates that period-to-period changes in the level of the original variable are employed in the estimation procedure, rather than the level of the variable itself. "Moving average" means that a moving average procedure has been used to eliminate any intercorrelations of the error term to its own past or future values.

The elimination of intercorrelations among error terms from different periods is a key feature of ARIMA and other sophisticated models. When this intercorrelation is not eliminated, the model violates a requirement for obtaining valid parameter estimates: the requirement that the error term is a random disturbance to the model in each time period, unrelated to the error terms of other time periods. Invalid estimation procedures are likely to lead to forecasts that are inferior to those obtained from models that satisfy basic requirements of no interdependence among error terms.

[3] More complex autoregressive models would include the possibility that the current value of the variable is related to its value in many different preceding periods, not just to its value in the last period.

COMPARATIVE PERFORMANCES IN FORECASTING: ARIMA VS. ECONOMIC MODELS

Several studies have compared ARIMA's forecasting accuracy with the forecasting accuracy of economic models. In any such comparison, there are six steps involved. The first step is to select some variable or variables to forecast, such as gross national product (GNP), employment, or the variable to be examined in the second part of this article, retail sales. The second step is to select economic models to use in the comparison.

Selecting the economic model is by no means easy, since no very good economic model may exist, in which case it will have to be constructed and estimated. Or it may be that hundreds of economic models exist for forecasting the variable selected, in which case some choice will have to be made. No selection problem is presented in the case of the ARIMA model, of course, since it is defined solely with reference to past values of the variable to be forecast.

The third step is to choose estimation and forecast periods. Since forecasting accuracy cannot be determined without reference to actual values, the forecast period must be selected to be part of the past. To simulate actual forecasting, therefore, the estimation period used to arrive at parameter estimates of the forecasting models must end before the forecast period begins.

The fourth step is to statistically estimate the parameter values of the models, using the historical data selected. The forecasts themselves are the fifth step. As indicated earlier, forecasting with an estimated model involves using the parameter estimates and the values of the exogenous variables to solve for the variable being forecast. The sixth and final step requires choosing some measure of

forecasting accuracy, and then determining how well the ARIMA and the economic models perform, based on these measures.

All measures of forecast accuracy compare the values forecast by the models with those that actually were observed. The difference between the actual and the forecast values is the forecast error. Forecast errors are usually calculated for values of the forecast variable outside (beyond the last date) of the estimation period but, conceptually, a forecast error is closely related to an estimated value of an error term within the estimation period. Usually forecasts for several periods are made, so some summary statistics are needed. Among those commonly used are mean algebraic error (MALE), mean absolute error (MABE), and mean square error (MSQE). MALE is calculated by summing a model's forecast errors (differences between actual and forecast values) and taking the average. MABE is computed by summing the forecast errors without regard to sign (that is, summing the absolute values of these errors), then taking the average. MSQE is the average of the sum of the squared forecast errors.

Several researchers have compared the forecasting accuracy of ARIMA with that of economic models of the aggregate economy. Examples of macroeconomic models of the U.S. economy include those developed by the Bureau of Economic Analysis of the U.S. Department of Commerce, and by the Wharton School of Business of the University of Pennsylvania. Because of the macroeconomic nature of these models, the comparisons of their forecasting accuracy with that of ARIMA have involved forecasts of variables such as GNP, the GNP price deflator, and the national unemployment rate.

Ronald Cooper compared the forecasts of 33 endogenous variables from seven macroeconomic models with ARIMA forecasts of those same variables. The ARIMA model forecast 18 of 33 variables better than any of the economic models, although it should be noted one of the variables ARIMA did not forecast well was inflation. Charles Nelson compared the forecasts of 14 endogenous variables from the Federal Reserve-MIT-Pennsylvania (FMP) model with ARIMA forecasts, and found ARIMA forecast 9 of the 13 variables better than FMP, but again ARIMA did not forecast the rate of inflation well. In another study, J. Phillip Cooper and Charles Nelson obtained mixed results when they compared ARIMA forecasts of six variables to those generated by the St. Louis model (a model developed by the Federal Reserve Bank of St. Louis) and the FMP model. Nariman Behravesh found ARIMA's forecasts of inflation, not unexpectedly, to be decidedly inferior to forecasts of inflation generated by a lineal descendant of the FMP model.[4]

The principal conclusion that can be drawn from these model comparisons is that for some variables, single-equation ARIMA models forecast better than do macroeconomic models. But that is not necessarily surprising. Macroeconomic models are constructed with several objectives in mind, among which are forecasts

---

[4] The articles cited here are the following: Ronald L. Cooper, "The Predictive Performance of Quarterly Econometric Models of the United States," in *Econometric Models of Cyclical Behavior*, Vol. 2, Bert G. Hickman, ed., National Bureau of Economic Research *Studies in Income and Wealth*, No. 36, Columbia University Press, New York and London, 1972. Charles R. Nelson, "The Prediction Performance of the FRB-MIT Model of the U.S. Economy," *American Economic Review*, pp. 902-17, Vol. 72, No. 5, December 1972. J. Phillip Cooper and Charles R. Nelson, "The Ex Ante Prediction Performance of the St. Louis and FRB-MIT-PENN Econometric Models and Some Results on Composite Predictions," *Journal of Money, Credit and Banking*, pp. 2-32, Vol. 7, No. 1, February 1975. Nariman Behravesh, "Forecasting Inflation: Does the Method Make a Difference?" Federal Reserve Bank of Philadelphia *Monthly Review*, September/October 1976.

of many, not just one variable, with special attention to forecasting turning points in the business cycle, as well as to showing the effects of fiscal and monetary policies on various sectors in the economy. To keep the size of a macroeconomic model within reasonable limits, the model builder may be forced to sacrifice the forecasting accuracy of individual variables for some broader goal. Then, too, not all equations from macroeconomic models are econometrically sound, especially with regard to the attention they give to intercorrelations among error terms through time.

An appropriate test of ARIMA's forecasting accuracy with that of an economic model would seem to call for the choice of a variable to be forecast, and then the construction of an economic model designed with forecasting that variable as its only purpose. It was with this objective that a case study, described in the next section, was undertaken.

ARIMA VS. ECONOMIC MODELS: FORECASTING RETAIL SALES

This section compares the forecasting accuracy of ARIMA with that of two economic models designed expressly for forecasting one variable: retail sales. The comparisons also include a mixed model, with both autoregressive and economic features. The forecasting abilities of all three of these sophisticated techniques—the ARIMA, the economic. and the mixed models—are also compared with the forecasting ability of a naive time trend model.

Retail sales is an appropriate variable to use in comparing the forecasting accuracies of various models. Data on retail sales are important economic indicators, watched closely by analysts of business conditions. This is especially true of the monthly reports, which are based on larger samples than those used in compiling the weekly figures. Because of the economic importance placed on month-to-month percentage changes in retail sales, and because monthly observations provide a long enough series to adequately estimate ARIMA and economic models and to compare their forecasts, monthly percentage changes in retail sales (hereafter abbreviated as $\dot{S}$)[5] was selected as the forecast variable.[6]

Having selected $\dot{S}$ as the variable to be forecast, the next step was to choose the models whose forecasts were to be compared. The ARIMA model presented no problem, since it is defined once the forecast variable is selected. In choosing from among various possibilities for alternative economic models, it was decided that only single-equation models containing no more than two explanatory variables would be considered. Since one of the appealing features of the ARIMA model is its single-equation simplicity, it seemed appropriate to use a simple single-equation economic model for comparison, unless the findings indicated that fairly complex economic models were required to improve upon the forecasts of ARIMA.[7]

---

5 A dot above a symbolic character will denote its rate of growth.

6 Another reason for choosing $\dot{S}$ is that it varies a great deal, even after seasonal adjustment. An easy-to-forecast variable, such as one that remains constant or grows at a constant rate, provides little. chalienge to even the naive models. The real test of soohisticated models comes when the naive methods do not forecast very well.

After the analytic work on this article was completed, the Bureau of Census published the results of extensive changes in the monthly surveys of retail trade. The results reported here, therefore, are based on the "final" monthly retail sales data available before this latest revision.

7 The use of a single-equation model is analogous to estimating a reduced form in which all the explanatory variables in the model can be viewed as exogenous. A single-equation model rather than a multi-equation model was used to maintain control of the major source of problems with many models—the intercorrelations among error terms from one period to the next.

The two explanatory or exogenous variables chosen for inclusion in one set of economic models were personal income ($\dot{I}$); and nonfinancial personal wealth ($\dot{W}$), as measured by an index of the price of common stocks.' As indicated early in this article, economic theory argues for the use of both personal income and *personal wealth in a relationship explaining consumer spending, which is closely related to retail sales.[9] An alternative economic model employs the money supply (M) as the sole explanatory variable.[10] According to monetarist theory in economics, changes in the stock of money directly and indirectly result in an increase in the demand for commodities. Finally, past values of retail sales were included in alternative models that mixed autoregressive and economic components.

Before forecasts of S could be made, the various statistical models had to be estimated with historical data. The basic estimation period used for this purpose began in January 1947 and ended with December 1974, the month prior to the forecast period. The fitted models were then used to make forecasts for each of 30 consecutive months of retail sales, beginning in January 1975, and ending in June 1977. These forecasts were made in one-month-ahead fashion. That is, the forecast of each month's retail sales was made using the actual values of explanatory variables for preceding months.

With forecast values in hand, the forecast errors were easily obtained by subtracting the actual values of monthly retail sales from the forecast values. Table 1 summarizes the results for five models, using one measure of forecasting accuracy, the mean absolute error. The first column in Table 1 gives the 30-month mean absolute error—the average absolute value of the forecast error--over the entire 2%-year forecast period. The next five columns, which show the MABE for 6-month intervals, indicate if the forecasting accuracy of the models degenerated the further the forecast month was from the end of the estimation period.

The principal conclusion that one can draw from the empirical results summarized in Table 1 is that, based on the MABE's calculated for this experiment, ARIMA did not forecast retail sales any better than did the naive model, and not as well, on the average, as did the economic models. The mixed model had a better record over the entire 30-month forecast period than did any of the other three models.[11]

Another conclusion indicated by Table 1 is that none of the models forecast retail sales very well—a one percentage point absolute error in the forecast of the month-to-month percentage change in retail sales is very large,

---

8 There are more complete measures of personal wealth than wealth in common stocks, of course. But almost all of the variation in total personal wealth is due to fluctuations in the stock market; other components of personal wealth grow at fairly constant rates.

9 While not all retail sales are sales to consumers, and while much of consumer spending (mostly on services) is not included in retail sales, the correlation between consumer expenditures and retail sales is very high.

10 Two measures of the money supply were tried: the narrowly defined money supply (M1) consisting of currency plus demand deposits, and the more broadly defined money supply (M2) consisting of M1 plus time deposits at commercial banks (except large negotiable certificates of deposit). Since M2 performed better than M1, references in the text to the money supply are to M2.

11 Only the results for the best naive, the best two economic, and the best mixed models are shown. It should be noted, however, that the difference-in results in the economic and mixed models that used M instead of $\dot{I}$ (or vice versa) was small. It should also be noted that the residuals (estimated error terms) of each estimated economic model and mixed model were examined for serial correlation (evidence of correlation of error terms between time periods). If serial correlation was found to be present, it was eliminated by an appropriate filter.

Table 1
THE FORECAST ACCURACY OF FIVE MODELS OF RETAIL SALES
(Mean Absolute Error in Per Cent Per Month)

Type of Model	Functional Form: Variables Whose Past Values Explain Retail Sales ($\dot{S}$)	Forecast Period: 30 Months January 1975 Through June 1977	Six Months Ending: June 1975	Dec. 1975	June 1976	Dec. 1976	June 1977
Naive	Time	1.35	1.64	.80	1.16	1.48	1.67
ARIMA	$\dot{S}$	1.35	1.46	.76	1.35	1.53	1.63
Economic	I	1.34	1.56	.96	1.05	1.63	1.50
Economic	*M*	1.24	1.41	.69	1.17	1.17	1.76
Mixed	$\dot{S}$, $\dot{I}$, W	1.12	1.62	.91	.81	1.69	1.24

considering the fact that the average monthly rate of growth of retail sales over the forecast period was itself about 1 per cent. The breakdown into 6-month periods also suggests that when one model forecasts poorly relative to its average, the other models are likely to be forecasting relatively poorly also. This is probably due to some omitted variable or variables in all the models.

The 6-month breakdowns do not indicate a degeneration of forecasts by the models, for all the models forecast the **final** 6 months about as poorly as the first 6 months, after showing some improvement in between. It was felt, however, that most forecasters probably would reestimate their models periodically, so an experiment to simulate such reestimation was carried out. Each of the models was refit four times by successive additions of 6 months of data to the original estimation period. After each of the four reestimations of the models, monthly forecasts were computed for the remainder of the forecast period, which was reduced in length as the estimation period was extended. As before, forecast errors were calculated. With a few scattered exceptions, there was no indication that **refitting** the model by updating the estimation period improved the forecasting accuracy of any model.[12]

The additional reestimations and forecasts did serve to provide more comparisons of the forecasting abilities of the various models. One such comparison is summarized in Table 2. In the simulated forecasting experiment reported on in this table, the forecaster is assumed to refit his forecasting model every 6 months, from December 1974 through December 1976, then make one-month-ahead forecasts for the 6 months immediately after the end of the estimation period. The entries in Table 2 thus represent the forecasts for the 6-month period immediately following the reestimation of the model.

---

12 It must be admitted, however, that if shorter estimation periods had been used and if the oldest data were dropped when the newest data were added, the results may have been improved.

Table 2
THE FORECASTING ACCURACY OF MODELS OF RETAIL SALES, ESTIMATED WITH DATA UP TO THE BEGINNING OF 6-MONTH FORECAST PERIODS

Type of Model	Functional Form: Variables Whose Past Values Explain Retail Sales ($\dot{S}$)	Forecast Period: Average Of 6-Month Period Forecasts	Six Months Ending: June 1975	Dec. 1975	June 1976	Dec. 1976	June 1977
Naive	Time	1.30	1.64	.78	.85	1.59	1.65
ARIMA	$\dot{S}$	1.45	1.46	.93	1.72	1.52	1.64
Economic	I	1.34	1.56	.95	1.02	1.64	1.51
Economic	M	1.22	1.41	.68	1.14	1.31	1.75
Mixed	$\dot{S}$, $\dot{I}$, W	1.25	1.62	.87	.85	1.68	1.26

The conclusions from Table 2 are much the same as those from Table 1. Although ARIMA does better than one or two of the alternative models some of the time, most of the time ARIMA does not forecast as 'accurately as a very simple economic model.[13]

SUMMARY AND'CONCLUSIONS

There are 'various kinds of models that can be used to forecast economic variables. Among those developed in recent years is the ARIMA model, which has the appealing characteristic of being based on the simple notion that a variable's future value can be forecast with reference only to its current and past values. Several studies have compared the forecasting accuracy of the ARIMA model to that of economic models of the U.S. economy. On balance, these studies seemed to indicate that ARIMA forecasts single variables better than such models. It is quite another thing, however, to conclude that ARIMA can forecast better than an economic model designed with the forecast of a single variable as its sole purpose. The experiment reported on in this article does, in fact, indicate quite the contrary. In comparative forecasts of monthly percentage changes in retail sales, ARIMA forecasts were usually no better and often worse than forecasts generated by a simple single-equation economic model.

[13] Other measures of forecast accuracy (MALE and MSQE) were calculated, and they led to the same conclusions.