Comparison with Alternative Models

A set of simple univariate quarterly time series models was developed for the 23 CPI price series in growth rates and the results of these models were compared against ERS unpublished, quarterly forecasts. Analysts use alternative models as a forecasting tool. The alternative models were estimated using the same information available to ERS staff when they made their forecasts.

The alternative models are estimated using the period 1984q1-1991q4 to initialize the models. One- to four-quarter-ahead forecasts are constructed starting in 1992q1. Then, actual data from 1992q1 is added to the sample and the model is re-estimated. Based on the updated model, one- to four-quarter-ahead forecasts are constructed once again. The price series for 1992q2 are added to the sample and the process is repeated. This continues until 1996q2 when the forecast horizon is reduced to a length of three quarters. Then, using data up through 1996q3 a two-quarter-ahead forecast is constructed. Finally, using data from the fourth quarter of 1996, a one-quarter-ahead forecast is made. Thus, there are 25, 24, 23, and 22 one-, two-, three-, and four-quarter-ahead forecasts, respectively, to compare with the internal ERS forecasts for the period 1992q1 through 1997q1. One of the weaknesses of comparing the ERS forecasting period 1984q1 through 1991q4 with the alternative models was that the ERS forecasting methods for this time period are unknown. Also, from 1987 through 1990, higher inflation rates ranging from 4.2 to 5.8 percent led to larger index changes for All Food. This would have made forecasting during this period more challenging.

Our method of identification and estimation differs slightly from Box and Jenkins' original suggestions for identifying and estimating time series models. We used the menu-based Time Series Forecasting System in SAS and various other procedures in SAS/ETS (Economic Time Series) to identify an alternative time series model that best fits the observations from the second quarter of 1984 to the fourth quarter of 1991. Observations from first quarter of 1992 to the first quarter of 1997 are used to measure the model's forecast performance. Prior to selecting the alternative model, ETS identifies the appropriate transformation of the data. It first performs a test of the log transformation, and if the log transformation cannot be rejected, the logged data are analyzed. Next, it performs a Dickey-Fuller test of the presence of a unit root in either the level or natural log form of the data. Next, the procedure tests the statistical significance of seasonal dummies within an autoregressive model of large order. If the set of seasonal dummies cannot be rejected, each candidate model contains seasonal dummies. ETS's preliminary tests are consistent with the suggestions of Joutz, Maddala, and Trost. Once the data are transformed, the model selected to compete with ERS' forecast is the one with the smallest root-mean-squared error (RMSE) among the alternative univariate, ARIMA, and seasonal models.[11]

The MA(1) model with seasonal dummies minimized the RMSE for most price growth rate series. Exponential smoothing models had similar fits to the MA(1) with seasonal dummies models; we decided to use the moving average model. If y_t denotes the original time series, the MA(1) model is

$$y_t = \theta \in_{t-1} + \sum_{k=1}^{k=4} d_k S_k + \in_t \qquad (3)$$

where $\in_{t-1}$ is last quarter's forecast error, $\in_t$ is the current quarter's forecast error, S_k are quarterly seasonal dummies defined in the usual way, and θ and the d_k are parameters to be estimated.

The AR(P) model with seasonal dummies minimized the RMSE criterion for the remaining price data. This model is given by

$$y_t = \alpha_1 y_{t-1} + ... \alpha_p y_{t-p} + \sum_{k=1}^{k=4} d_k S_k + \in_t \qquad (4)$$

where the α=I are parameters to be estimated. Table 5 lists the univariate models selected for each series.

Comparison with Alternative Forecasts

This section compares the accuracy of ERS forecasts with the alternative univariate models. This is accomplished by generating out-of-sample forecast errors, computing four accuracy statistics, and performing a

[11]The particular parameter estimates of the identification stage are not reported. They are available from the authors upon request.

statistical test of forecast reliability. The forecasts evaluated are out-of-sample forecasts in the sense that the periods of the time series forecasted are separated from the periods used to estimate model parameters. In generating and evaluating rolling forecasts,[12] we are simulating a forecaster updating a model in light of new information. This procedure is used for each of the 23 CPI indices that are being forecast.

The rolling forecasts are generated as follows: The parameters from the alternative models are estimated using the initial base period of the second quarter of 1984 through the fourth quarter of 1991. Based on the parameter estimates, predictions are produced for 1992q1 through 1992q4. Next, the observation for 1992q1 is added, and the alternative model parameter estimates are recomputed using the 1984q2 to 1992q1 observations. The revised parameter estimates are used to forecast the 1992q2 through 1993q1 observations. This process is repeated until the 1996q2 observation, when forecasts are limited to three quarters ahead. In each succeeding update, the forecast horizons are reduced by one quarter until the 1996q4 observation is added to the sample, and a single forecast is generated for 1997q1. For our data, the rolling forecast procedure results in 21 one-quarter-ahead forecasts, 20 two-quarter-ahead forecasts, 19 three-quarter-ahead forecasts, and 18 four-quarter-ahead forecasts. Unfortunately, we do not have the same number of ERS forecasts. This is due to the process described in the section above. For each variable, there are 21 one-step-ahead forecasts, 18 two-step-ahead forecasts, 13 three-step-ahead forecasts and 8 one-year-ahead forecasts from ERS for comparison purposes.

The forecast error of a j-quarter forecast made in quarter t is defined by

$$e_{t+j} = y_{t+j} - \hat{y}_{t,t+j} \qquad (5)$$

where y_{t+j} is the actual observation, and $y_{t,t+j}$ is the j-quarter-ahead forecast at time *t*.

Summary measures of forecast performance are computed using the following three statistics: the mean error (ME),

$$ME_j = \frac{1}{n_j}\sum_t (y_{t+j} - \hat{y}_{t,t+j}) \qquad (6)$$

[12]See Joutz, Maddala, and Trost for more details of the rolling forecast procedure.

the mean absolute error (MAE),

$$MAE_j = \frac{1}{n_j}\sum_t \left| y_{t+j} - \hat{y}_{t,t+j} \right| \qquad (7)$$

and the root-mean-squared error (RMSE),

$$RMSE_j = \sqrt{\sum_t \frac{(y_{t+j} - \hat{y}_{t,t+j})^2}{n_j}} \qquad (8)$$

for j = 1, 2, 3, and 4 where n_j denotes the number of j-step-ahead forecasts. The mean error is an indicator of bias in the forecast. The mean error measures how close the average value of the forecast is to the average value of the observations. The mean absolute error and the root-mean-squared error measure forecast error dispersion. Both the MAE and the RMSE reflect the potential uncertainty of a forecast. The larger the MAE or the RMSE, the more dispersion there is in a forecast error. The summary measures are reported in tables 6a-6d for the one- through four-quarter-ahead forecasts generated by both ERS and the alternate univariate models.

Across all prices for the ERS and alternate univariate models, the mean forecast errors do not appear to be biased. The t-statistic for the ratio of the mean error to RMSE is never greater than unity. Three series are particularly difficult to predict as might be expected from the initial examination of the data. Eggs, Fresh Fruit, and Fresh Vegetables in particular have the largest MAE and RMSE. The implications from MAE and RMSE statistics are very similar with respect to the relative performance of the two approaches. The alternate univariate models generally produce lower RMSE than the ERS forecasts. It should be noted that for All Food, the aggregate of the individual food subcategories, ERS and the alternative model are comparable. The RMSE for the ERS model ranged from 0.47 to 0.59 percent, while the alternative model RMSE range was 0.41 to 0.45 percent. In the one-quarter-ahead case, the RMSE is lower for 17 of the 23 price series using the alternate models. Other Meats, Poultry, and Fats and Oils show that the alternate models reduce the RMSE to 50 percent of the ERS forecast error.

The alternate models have lower RMSE for 18 of the 23 price series at the two-quarter-ahead forecast horizon. The alternate model RMSE for Fresh Vegetables is more than twice the size of the ERS forecast: 26 percent versus 11 percent. This problem gets worse at the three-quarter and four-quarter-ahead forecasts. In the first part of the sample, that part for which the alternate models were estimated and selected, there appears to be a fair amount of seasonal variation in the Fresh Vegetables inflation series. The seasonal pattern appears to diminish as we move further into the 1990's. The alternate model is a MA(1) with seasonal dummy variables.

The alternate model produces forecast RMSE which are lower in 17 and 16 of the price series at the three- and four-quarter-ahead horizons, respectively. The Food Away from Home RMSE for the alternate model is one-half that of the ERS forecast at both horizons.

Another way of evaluating these forecasts is to determine whether they have produced "good" results in the sense that they are unbiased and have incorporated the information contained in past forecasts and forecast errors. The regression approach to forecast evaluation recommended by Mincer and Zarnowitz does this. Consider the following regression of the historical series at time t+h on the conditional forecast for time t+h made at time t and a constant.

$$y_{t+h} = \beta_0 + \beta_1 \hat{y}_{t,t+h} + e_{t+h}$$

We test the null hypothesis that the coefficient on the constant is zero and the slope coefficient is one jointly and the residuals are white noise. This test is referred to as the weak form efficiency test in the forecast evaluation literature.

Tables 7 and 8 present the results for the weak form efficiency tests at the one-quarter-ahead horizon using the ERS forecasts and the Alternate Model forecasts, respectively. There are 21 observations for each regression since the sample period is 1992q1 through 1997q1. The results are not particularly promising for either forecasting approach.

The tables are formatted in a similar manner. Coefficient estimates for the intercept and slope are provided in the second and third columns with t-statistics reported below them. The hypothesis or F-test is given in the fourth column; the p-value is provided below the test statistic. R-squared and the Durbin Watson statistic are found in the fourth and fifth columns.

The weak form efficiency hypothesis test is rejected at the 5-percent level for 17 of the 23 ERS price forecasts. It cannot be rejected at the 1-percent level for three series: Fish and Seafood, Fruits and Vegetables, and Processed Fruit. Six price series forecasts appear to pass the weak form test at 5-percent level of significance. They are Dairy Products, Fresh Vegetables, Processed Fruits and Vegetables, Processed Fruits, Sugars and Sweets, and Nonalcoholic Beverages. The Fresh Vegetables result is curious given the tremendous volatility and forecast errors reported earlier. Nevertheless, the ERS forecasts and forecasters appear to make good predictions of this CPI component in this environment.

The alternate univariate forecasts appear to be inefficient in only 11 of the 23 price forecasts. Four of the rejections of the weak-form test are at 5 percent, the remainder are at 1 percent. Among these is the Fresh Vegetables series. There are 12 price series where the null hypothesis cannot be rejected at the 5-percent level. Among the non-rejections is where the value is 3.94 and the standard error is 1.59.

There are seven price series for which both forecasting approaches cannot reject the null hypothesis of weak form efficiency. They are Fish and Seafood, Dairy Products, Fresh Vegetables, Processed Fruits and Vegetables, Processed Vegetables, Sugars and Sweets, and Nonalcoholic Beverages.

There is a third method for evaluating two sets of forecasts. Granger and Newbold (1977) propose a statistical test designed to compare the one-step-ahead forecast uncertainty of the two competing models. The test presumes that the forecasts are unbiased and the forecast errors from each model are serially uncorrelated. Since the forecast errors associated with j-step-ahead forecasts are generally serially correlated for j > 1, Granger and Newbold's test can only be applied to compare the uncertainty of one-step-ahead forecasts.[13]

Since the one-step-ahead forecasts from each model are presumably unbiased, and the RMSE of the forecast errors is a monotonically increasing function of the variance of the forecast errors, the variance of the forecast errors is a measure of forecast uncertainty.

[13]The optimal j-step-ahead forecast errors follow an MA(j-1) process, and hence, for greater than one-step-ahead forecasts, violate the uncorrelated residuals condition necessary for applying the Granger and Newbold test.

The null hypothesis is the variance, σ_1^2, of the one-step-ahead ERS forecast errors, e_1, equals the variance, σ_2^2, of the one-step-ahead Alternate forecast errors, e_2. The test assumes the vector (e_1,e_2) is randomly drawn from a bivariate normal distribution with parameters, σ_1^2 and σ_2^2, and correlation coefficient, ρ.[14] Under the null hypothesis, $\sigma_1^2 = \sigma_2^2$, the correlation between the variables $(e_1 - e_2)$ and $(e_1 + e_2)$ is zero. Consider the regression of $(e_1 - e_2)$ on $(e_1 + e_2)$:

$$(e_1 - e_2)_t = \alpha + \beta * (e_1 + e_2)_t + u_t \qquad (10)$$

The hull hypothesis, $\beta=0$ is equivalent to $\sigma_1^2 = \sigma_2^2$, which means the two models' one-step-ahead, unbiased forecasts are equally reliable. The statement $\beta \neq 0$ is equivalent to the statement $\sigma_1^2 \neq \sigma_2^2$, implying that the models' forecasts are not equally reliable. The statement $\beta>0$ is equivalent to the statement $\sigma_1^2 > \sigma_2^2$. (Alternate more reliable than ERS.) Finally, the statement $\beta<0$ is equivalent to the statement $\sigma_1^2 < \sigma_2^2$. (ERS more reliable than Alternate.)

Table 9 reports results from the Granger and Newbold test for minimum RMSE. The coefficient estimate for the slope term is provided in the second column with the associated p-value below. We use a 10-percent rule to test if one forecast methodology provides a significantly lower RMSE. The Alternate is the lowest for seven price series: Food Away from Home, Poultry, Processed Vegetables, Sugars and Sweets, Cereals and Bakery Products, and Other Prepared Foods. The ERS forecast appears to have a lower RMSE in three cases: Fruits and Vegetables, Fresh Fruits, and Processed Fruits. In the other cases, there is no significant difference between the forecast error variance.

Combining ERS Forecasts and Alternative Models

Forecast comparisons with respect to different loss functions are always interesting and likened to horse races. Typically, there is no clear and consistent winner for a particular variable or over all time periods. It has become a common forecasting practice to combine predictions generated by alternative methods. This can lead to improved forecasts since the hybrid forecast is using a larger information set. We can combine the forecasts in a linear combination and test if the Alternate forecast provides significant information towards a better forecast than using the ERS predictions alone.

$$y_{t+h} = \beta_0 + \beta_1 y^{ERS}_{t+h,t} + \beta_2 y^{Alt}_{t+h,t} + e_{t+h}$$

A restricted version of this model would be to have the slope coefficients, interpreted as weights, sum to unity.

$$y_{t+h} = \beta_0 + (1-\beta_2) y^{ERS}_{t+h,t} + \beta_2 y^{Alt}_{t+h,t} + e_{t+h}$$

Table 10 presents estimates for both of these models. The F-test for the restriction is in the last column. There are only six rejections out of 23 prices of the null hypothesis that the coefficients sum to unity: All Food, Food Away from Home, Food at Home, Other Meats, and Fruits and Vegetables, and Fresh Fruits. The fifth column contains the restricted coefficient for the Alternate forecast. The coefficient is significant in 17 of 23 cases; 14 of these are significant at 1 percent. This result suggests that the alternate forecasts can provide valuable information to ERS forecasters.

Does the linear forecast combination exercise produce forecasts which would have had significantly lower RMSE than the ERS forecasts? It appears that the competing forecasts can be profitably combined to yield a composite forecast which is superior to each of the individual forecasts.

Table 11 presents the one-quarter-ahead RMSE for the ERS and alternative forecasts with the implied RMSE from the forecast combinations where the weights are not constrained and where they are forced to sum to unity. The last four columns show the improvement in the RMSE on a percentage basis from combining the ERS and alternative forecast in an optimal least squares approach over the individual forecasts. Focus on the fourth and second columns from the right labeled ERS since they show whether or not the forecast information provided by the alternative models helps to improve the ERS forecasts.

The results from table 10 suggest that the ERS and alternate forecasts can be merged into a simple weighted average with the weights (constrained) summing to unity for 17 of the CPI components. These forecast combinations produce lower RMSE than the ERS forecasts by 20 percent or more for 18 and 16 components when the weights are unconstrained and

[14] The assumption the forecast errors are randomly drawn from a bivariate normal assumption rules out the possibility of serially correlated forecast errors.

constrained, respectively. There is significant value in using the alternative models as a means to improve the forecasts. The current ERS forecasting methodology or information set is not improved by the alternative models for Dairy Products, Fruits and Vegetables, Fresh Vegetables, and Processed Fruits. This result is consistent with the earlier evidence.

In practice, the weights assigned to different forecasts should not be considered to be fixed over time. They should be periodically reestimated. The forecast evaluation literature finds that the relative importance of individual forecasts can vary over time.

Limitations of the Alternative Time Series Model Forecasts

Although the best alternative time series model was selected, it is not reliable in forecasting turning points. Time Series Model predictions are based solely on the past behavior of the variable estimated and that variable alone. Some of the movement can be difficult to explain and if the past movement was due to factors that are not explainable such as the weather, changes in consumer tastes, or simply seasonal cycles in consumer spending, the model may not be able to relate to other economic variables. After careful review of each food category from 1992 through 1997, the alternative time series model consistently overestimated many of the food indexes, which indicated serial correlation in several food categories. Since 1992, changes in consumer tastes and preferences for certain foods plus the low general inflation index for the all items CPI, which ranged from 2.3 to 3.0 percent, may have contributed to the alternative time series model overestimation. Although the alternative time series model RMSE was generally lower than the ERS forecasts, the time series model did not capture some of the recent trend changes in several of the food CPI categories.

The time series model overestimated the actual index 5 out of 6 years for All Food, Fruits and Vegetables, and Other Foods; and 4 out of 6 years for Food Away from Home and Food at Home. In addition, the time series model overestimated the actual index 4 out of 6 years for Other Meats, Fish and Seafood, Dairy Products, Processed Fruits and Vegetables, Sugar and Sweets, Cereals and Bakery Products, and Nonalcoholic Beverages. Many of the food categories that were overestimated at least 4 out of 6 years are highly dependent on changes in the All Items inflation index, which has not increased at the rate that the time series model would have expected. When consumer tastes and preferences for selected foods changed and the All Items inflation index remained lower than expected, the time series model did not detect the changes from 1992 through 1997.