1985

No. 1985-11

FORECASTING HOG PRICES USING TIME SERIES ANALYSIS OF RESIDUALS

Matthew T. Holt

and

Jon A. Brandt*

Swine -- Prices

Selected paper of the annual meeting of the American Agricultural Economics Association, Ames, Iowa, August 4-8, 1985.

*The authors are graduate research assistant and visiting associate professor, Department of Agricultural Economics, University of Missouri-Columbia.

ABSTRACT

FORECASTING HOG PRICES USING TIME SERIES ANALYSIS OF RESIDUALS

Matthew T. Holt

and

Jon A. Brandt

A time series analysis of the residuals (TSAR) of a single-equation econometric hog-price forecasting model is conducted. Post-sample forecasts from the integrated econometric-time series model were compared with forecasts from individual econometric and time series approaches. The TSAR forecasts offered some improvement over the individual methods.

FORECASTING HOG PRICES USING TIME SERIES ANALYSIS OF RESIDUALS

In recent years, a good deal of professional effort has been directed at developing quantitative models for forecasting economic variables. The types of quantitative models developed by agricultural economists can usually be classified under one of two broad categories: (1) the traditional econometric approach or; (2) the time series approach. Models from the former group are predicated on economic theory and may use a number of variables to explain price-quantity movements. Loosely speaking, the goal of econometric analysis is to identify and estimate the economic structure of the particular industry or commodity in question. On the other hand, models from the latter group typically have little basis in economic theory. The aim of time series analysis is to exploit the information contained in past values of the variable to be forecasted. In spite of this relative simplicity, it is often the case that univariate time series models will produce forecasts similar (if not superior) in quality to those generated by more sophisticated structural models. This has led economists to explore ways of combining the information contained in both econometric and time series models.

Recently, Ashley and Granger combined econometric analysis with time series models to forecast movements in several macro economic variables. The results indicated improvement in post sample forecasting performance was possible. The approach involved applying a time series analysis to the residuals (TSAR) of the structural equations of the econometric model. Few attempts to combine econometric analysis and time series models can be found in the agricultural economics literature. Ikerd attempted to model the residuals of a beef price spread equation (estimated with ordinary least squares) by using an auto-regressive integrated moving average (ARIMA) process but made no attempt to incorporate this additional information into post-sample predictions. Harris and Leuthold examined two methods to integrate time series and econometric approaches to forecast beef and hog prices. Their work is examined in greater detail below.

In this paper, a TSAR process is applied to the residuals of an econometric model to forecast quarterly hog prices. The forecasts from alternative TSAR models are compared with the original econometric forecasts and those from a univariate time series model of hog prices for forecasting accuracy. The objective of this paper is to determine if model builders can combine techniques to improve model performance and if more accurate information can be generated for decision makers.

The Forecasting Methods

Forecasting models have been classified as structural and nonstructural, mechanical and nonmechanical (Bessler and Brandt, 1979). Econometric models could be categorized as structural and mechanical whereas time series models could be considered nonstructural and mechanical. (Nonmechanical approaches tend to use opinion or judgment based on expertise.) Examples of the former technique for forecasting are numerous in the agricultural economics literature. While somewhat fewer studies using time series analysis are available, certainly within the last decade, this approach has received widespread interest, particularly for short-term forecasting. In this section, separate econometric and time series models are specified for forecasting live hog prices. Newbold (p. 24), however, argues that we should "not think of time series analysis as an alternative to building regression models for forecasting. Rather, modern time series techniques should be incorporated into the model building..."

Given that the two methods described above differ substantially in their approach, it is somewhat surprising that they typically produce results (forecasts) of comparable quality. (See Granger and Newbold for a review of such comparisons). Therefore, it seems reasonable to believe that both methods have something to contribute to our ability to anticipate the future and, as a result, a hybrid of the two approaches may be advantageous. Consequently, the discussion in this section ends with the development of a pair of models which apply time series analysis to the residuals of an the econometric model.

Econometric Methods

Econometric models are designed to explain or predict price and quantity patterns within a particular structure. Theoretical considerations regarding the development of a structural model must explicitly deal with factors affecting the supply and demand for a particular commodity. One of the objectives was to develop an econometric forecasting model for the price of hogs which is reasonably simple to understand, use and estimate. A single-equation unrestricted reduced-form model was specified to explain hog price behavior. This model is similar to one used by Bessler and Brandt. The model was initially estimated using ordinary least squares (OLS) over the 52-quarter period from 1966 through 1978. The model was then updated and re-estimated annually through 1983. The OLS estimation results for the initial fit period are reported in table 1.

In order to make the model truly predictive for one-step-ahead forecasting, only lagged values of relevant supply and demand variables were used to explain the changes in the dependent variable. (Harris and Leuthold used current exogenous variable values and assumed these were known for

TABLE 1. Alternative Model Specifications for Forecasting Quarterly Hog Prices.

1. Econometric (OLS)

$$PH_t = \underset{(-11.50)^a}{-182.0222} - \underset{(-7.53)}{10.6499}SF_{t-2} - \underset{(-4.86)}{6.7489}SF_{t-3} - \underset{(-3.69)}{14.9704}CS_{t-1} - \underset{(-5.60)}{178.3957}HATCH_{t-1}$$

$$+ \underset{(11.74)}{51.8099}\ln INC_{t-1} + e_{1t} \qquad R^2 = .90 \qquad D.W. = 1.35^b$$

2. ARIMA

$$(1 - B)PH_t = \underset{(1.78)}{.5447} + (1 - \underset{(-3.80)}{.5048}B^5)v_t \qquad R^2 = .89 \qquad Q^c_{(22)} = 23.12$$

3. TSAR 1

$$(1 - \underset{(-.67)}{.0836}B + \underset{(3.09)}{.3527}B^5 + \underset{(2.85)}{.3774}B^6)e_{1t} = u_{1t} \qquad R^2 = .44 \qquad Q^d_{(21)} = 10.14$$

4. TSAR2 (Econometric NLS)

$$PH_t = \underset{(12.41)}{-184.8830} - \underset{(-8.77)}{11.1922}SF_{t-2} - \underset{(-4.17)}{5.5063}SF_{t-3} - \underset{(-5.02)}{14.1374}CS_{t-1} - \underset{(-6.67)}{175.8830}HATCH_{t-1}$$

$$+ \underset{(18.19)}{51.4705}\ln INC_{t-1} + \underset{(.90)}{.1239}e_{2t-1} - \underset{(-2.49)}{.3300}e_{2t-5} - \underset{(-2.93)}{.4202}e_{2t-6} + u_{2t}$$

$$R^2 = .95 \qquad D.W. = 1.88^e$$

Notation:

PH - quarterly average price of barrows and gilts at seven terminal markets (\$/cwt.); SF - sows farrowing in ten states (million head); CS - commercial cattle slaughter (billion pounds); HATCH - eggs hatched for broilers (billion eggs); lnINC - the natural logarithm of disposable income (billion dollars); e - fitting errors corresponding to the econometric model; u - a white noise disturbance term corresponding to the stochastic process generating e; v - a white noise disturbance term for the stochastic process generating PH.

[a] t-ratios and asymptotic t-ratios are in parantheses below the estimated coefficients.

[b] The lower and upper bounds for the Durbin-Watson test in this case are $d_L = 1.39$ and $d_U = 1.72$ at the 5 percent significance level.

[c] The chi-squared statistic for 22 degrees of freedom at the 95 percent level is 33.92.

[d] The chi-squared statistic for 21 degrees of freedom at the 95 percent level is 32.67.

[e] The lower and upper bounds for the Durbin-Watson test in this case are $d_L = 1.22$ and $d_U = 1.92$ at the 5 percent significance level.

one-to-four quarters into the future. Such an assumption is likely to underestimate forecast error.)

The econometric model does a reasonable job of explaining movements in hog prices over the fit period (R^2=.90). All coefficients have theoretically correct signs and are statistically significant at normal significance levels, as indicated by the large size of the t-ratios. Even though the models were updated and re-estimated annually, there was relatively little change in the magnitude or significance of the coefficients, with the exception of the steady decline in magnitude and significance of the cattle slaughter variable. In fact, the only apparent problem with the OLS econometric model is the presence of first-order autocorrelation, as indicated by the low value of the Durbin-Watson test statistic. Ashley and Granger (p. 380) have noted that "a 'good' Durbin-Watson statistic is a necessary, but not sufficient, indication that autocorrelation is not present, since it tests only for first-order autocorrelation." Apparently, at least first-order and perhaps higher orders of autocorrelation require attention in this model. This will be explored more fully below.

Autoregressive Integrated Moving Average (ARIMA) Processes

An ARIMA model is identified, estimated, and checked (procedures described in Box and Jenkins) for the hog price series. The underlying assumption associated with this process is that the patterns of seasonal, cyclical, and trend movements which have generated the historical price data are expected to continue into the future. The model used in this analysis is a fifth-order moving average process of the differenced hog price data, the same as that estimated first by Bessler and Brandt and later by Harris and Leuthold. The initial maximum likelihood estimation results are present in table 1. The Ljung-Box Q statistic for 22 degrees of freedom is well below the critical chi-squared level indicating that the hypothesis of white noise in the residuals could not be rejected.

In a recent Monte Carlo study, Ansley and Newbold examined the finite sample properties of several estimators for ARIMA models. They found that in medium-sized samples (50 observations) of quarterly data, the maximum likelihood (ML) estimates were generally preferred to the least squares estimates. On the basis of these conclusions, all of the time series models in this study were estimated by ML procedures.

Time Series Analysis of Residuals (TSAR)

Because both econometric models based on structural characteristics and time series models which exploit well-defined historical patterns have been shown to do a reasonably good job in explaining and predicting variable movements over time, analysts have investigated the possibility of combining the approaches to improve the forecasting accuracy over either individual method. One popular approach has been to combine the individual forecasts into a composite (e.g., Bates and Granger, Brandt and Bessler (1981), Kulshreshtha et al). The results of these studies have been encouraging in that improved forecasting accuracy is possible.

An alternative to combining forecasts from different techniques is to integrate the two methods directly. One approach has been to develop multivariate time series models (Granger and Newbold) basing the selection of variables on industry structure and economic theory and lag lengths on time series analysis. For agricultural examples of vector autoregression analysis see Brandt and Bessler (1984); Harris and Leuthold; and Nerlove, Grether, and Carvalho. Such an approach has the advantage of allowing for the possibility of feedback within the system, however, it is computationally difficult and requires greater arbitrary input from the modeler than does the univariate case. Furthermore, multivariate time series models have generally been found to be relatively poor forecasting tools.

A second approach for integrating regression and time series models is with transfer function models. Pindyck and Rubinfeld (p. 594) state that "a transfer function model simply relates a dependent variable to lagged values of itself, current and lagged values of one or more independent variables, and an error term which is partially 'explained' by a time series model." Such a model is likely to generate more accurate forecasts than either an econometric model alone or an ARIMA model alone since it integrates a structural part of the variation in the dependent variable, as well as a time series part of the dependent variable that cannot be explained structurally.

Because the Durbin-Watson statistic of the estimated regression equation suggested that first-order autocorrelation was present during the fit period (and in fact this problem became worse as more observations were added to the model), an analysis of the residuals through a transfer function model seemed a potentially fruitful avenue of investigation to improve forecasting accuracy. Following the procedure outlined in the previous section, an ARIMA process was specified for the residuals of the econometric model over the initial fit period.

In the identification step, sample autocorrelations were calculated for the residual series. The value of the Ljung-Box Q statistic based on 24 degrees of freedom was almost twice as great as the critical value of the chi-square distribution at the 5 percent significance level. This leads us to reject the null hypothesis that the residuals of the econometric model are white noise. Further examination of the autocorrelation and partial autocorrelation functions revealed spikes at lags one, five, and six which were more than twice their standard errors. Consequently, a first-, fifth-, and sixth-order autoregressive process was specified for the error terms.

The results of the ML estimation of this TSAR model (identified as TSAR1) over the original 52-quarter fit period are also listed in table 1. With the exception of the first-order term, all coefficients were statistically significant. Although the R^2 value of this model is relatively low, the Ljung-Box Q statistic is well below the critical value at the 5 percent level of significance. Hence, there is no reason to suspect the adequacy of the estimated TSAR model.

As with the econometric and ARIMA models, the TSAR1 model was updated and re-estimated annually through 1983 and the same AR(1,5,6) process was maintained throughout. In latter estimations, the first-order coefficient gained in magnitude and significance while the sixth-order term declined in absolute value and became statistically insignificant. To make predictions with the TSAR1 model, it is necessary to first forecast the fitting error and then combine this with the price forecast generated by the econometric model.

Once a stationary error process has been identified, it is appropriate to perform feasible generalized least squares estimation on the original econometric model to obtain efficient parameter estimates. Although it is relatively simple to derive the appropriate data transformation matrix for low-order AR, MA, or ARMA(1,1) processes, tractable expressions for the transformation matrix of higher-order processes are not known to exist. As a result, the regression parameters and the parameters describing the error process are usually estimated simultaneously, either by nonlinear least squares or maximum likelihood estimation (Fomby, Hill, and Johnson; p. 221). Pierce has shown that both methods produce asymptotically efficient estimates assuming that the error terms are normally distributed.

The original econometric model was reparameterized to include the AR(1,5,6) error process and re-estimated using nonlinear least squares (NLS), hereafter referred to as the TSAR2 model. On the basis of asymptotic standard errors, all of the coefficients with the exception of the first-order autoregressive term were statistically significant (table 1). Furthermore, the TSAR2 estimates are reasonably close to those obtained when the econometric and TSAR1 models were fit separately.

Forecast Results and Performance Evaluation

Frequently in applied econometric and time series analysis, several specifications and functional forms are estimated before a final model is chosen. The sample standard errors are thus conditional upon the final model specification being correct. Typically the estimated standard errors will understate their true values, the result being that the power of statistical tests is overstated. This problem has led economists to search for alternative methods of model validation.

Ashley and Granger (and others) have stressed that model evaluation should primarily rely upon post-sample forecasting performance. Box and Tiao suggest that the only formal test of model adequacy is based on one-step-ahead prediction. In this study, one-step-ahead quarterly forecasts were made for each model over the 22-quarter period from 1979 through the second quarter of 1984. Several measures of predictive performance are utilized to evaluate the forecasting capability of each model.

Table 2 includes root mean squared errors (RMSE), mean absolute forecast errors (MAFE), and mean forecast errors (MFE) for each forecasting model. The ARIMA model had the highest RMSE followed by the econometric model. The forecasts from both the TSAR1 and the TSAR2 models resulted in RMSEs which were well below those of either the original econometric or ARIMA forecasts.

TABLE 2. Performance Measures of Post-Sample Hog Price Forecasts.[a]

	Econometric	ARIMA	TSAR1	TSAR2
Root Mean Squared Error	7.12	7.54	6.37	6.49
Mean Absolute Forecast Error	6.21	6.11	5.60	5.83
Mean Forecast Error	-4.33	-1.80	-4.03	-4.38

[a] The above evaluation techniques are well known and do not warrant further elaboration. For a review of these (and other) measures, see Pindyck and Rubinfeld, pp. 361-367.

TABLE 3. Turning Point Evaluation by Forecasting Approach.[a]

		Econometric		ARIMA		TSAR1		TSAR2	
		C	NC	C	NC	C	NC	C	NC
ACTUAL	C	5	6	4	7	6	5	4	7
	NC	3	6	7	2	4	5	4	5

[a] The measures indicate a change (C) or no change (NC) in the direction of the price movement. High performance is associated with large numbers on the positive diagonal.

The TSAR2 forecasts resulted in an RMSE which was only incrementally larger than the RMSE from the combined TSAR1 forecasts. An examination of the mean absolute forecast errors suggests performance similar to the RMSE evaluation, except that the econometric MAFE was slightly larger than the ARIMA MAFE.

The mean forecast error is simply the arithmetic mean of the prediction error series for a particular model and is representative of the forecast bias over the prediction period. The negative signs indicate that all models tended to "over predict." The ARIMA model had the smallest MFE while the TSAR models did little to reduce the forecasting bias associated with the original econometric model. In fact, the TSAR2 model had the largest MFE among all models over the post sample prediction period.

The ability of a forecasting model to track the actual movements of a series from period to period may also be important to decision makers. Table 3 evaluates each forecasting model's ability to anticipate turns or reversals in the direction of hog prices. A model which correctly anticipates turns or correctly forecasts no change in the actual series will be associated with large numbers on the main diagonal of the 2x2 matrix. As indicated in table 3, the reduced-form econometric and the TSAR1 models performed best, although each correctly anticipated only slightly more than 50 percent of the actual price movements. The ARIMA model performed poorest by a substantial margin under this criterion.

Conclusions and Implications

In practice, most applied econometricians will, at best, test and correct for an AR(1) error process. As illustrated here, however, there may often be much more information which can be exploited through a more rigorous investigation and analysis of the residual series for estimating and forecasting purposes. Integrating econometric and time series models can potentially result in forecasts which will be better than the forecasts of either an econometric or time series model alone. Perhaps the best reason for considering TSAR models is that with the exception of the MFE criterion, the TSAR forecasts were no worse than those of either individual method (e.g., econometric or time series). In short, there appears to be little to lose and potentially much to gain by combining econometric and time series analysis.

Another implication is that while nonlinear estimation of the structural and time series parameters produces asymptotically efficient estimates, there appears to be little gain in post-sample forecasting accuracy. This is not surprising in view of the fact that even though the OLS estimates are inefficient, they will, under ordinary circumstances, still be consistent. Consequently, if price forecasting is the ultimate goal, the relatively more expensive nonlinear or maximum likelihood estimates may not be worth the time and cost.

Finally, a word of warning is in order. Fitting time series models to error terms cannot compensate for a poorly specified econometric model in that it is not a "fix-all" device. In practice, there is no substitute for the judicious use of economic theory in the initial stages of model development. However, once economic logic has been exhausted or data limitations imposed, an analysis of the model's residuals could enhance both explanatory and forecasting power.

References

Ansley, Craig F., and Paul Newbold. "Finite Sample Properties of Estimators for Autoregressive Moving Average Models." J. Econometrics 13(1980): 159-83.

Ashley, Richard A., and Clive W. J. Granger. "Time Series Analysis of Residuals from the St. Louis Model." J. Macroeconomics 1(Fall 1979): 373-94.

Bates, J.M., and C.W.J. Granger. "The Combination of Forecasts." Operations Research Quarterly 20(1969): 451-68.

Bessler, David A., and Jon A. Brandt. Composite Forecasting of Livestock Prices: An Analysis of Combining Alternative Forecasting Methods. Purdue University Agr. Exp. Sta. Bul. 265, Dec. 1979.

Box, G.E.P, and G. Jenkins. Time Series Analysis: Forecasting and Control. Revised edition, San Francisco: Holden-Day, 1976.

Box, G.E.P., and G.C. Tiao. "Comparison of Forecast and Actuality." Applied Statistics 25(1976): 195-200.

Brandt, Jon A., and David A. Bessler. "Composite Forecasting: An Application with U.S. Hog Prices." Amer. J. Agr. Econ. 63(1981): 135-40.

Brandt, Jon A., and David A. Bessler. "Forecasting with Vector Autoregressions Versus a Univariate ARIMA Process: An Empirical Example with U.S. Hog Prices." No. Cent. J. Agr. Econ. 6(1984): 29-36.

Fomby, Thomas B., R. Carter Hill, and Stanley R. Johnson. Advanced Econometric Methods. New York: Springer-Verlag, 1984.

Granger, C.W.J., and P. Newbold. Forecasting Economic Time Series. New York: Academic Press, 1977.

Harris, Kim S., and Raymond M. Leuthold. "A Comparison of Alternative Forecasting Techniques for Livestock Prices: A Case Study." No. Cent. J. Agr. Econ. 7(1985):40-50.

Ikerd, John E. "Analysis of Retail Beef to Live Cattle Price Spreads." Proceedings of the NCR-134 Conference, Applied Commodity Price Analysis, Forecasting, and Market Risk Management, pp. 18-36, Des Moines, Iowa, April, 1983.

Kulshreshtha, S.N., J.D. Spriggs, and A. Akinfemiwa. A Comparison of Alternative Approaches to Forecasting Cattle Prices in Canada. Dept. of Agr. Econ., Univ. of Saskatchewan, Saskatoon, Saskatchewan, Canada. Tech. Bul. 82-01, 1982.

Nerlove, Marc, David M. Grether, and Jose L. Carvalho. Analysis of Econometric Time Series. New York: Academic Press, 1979.

Newbold, Paul. "ARIMA Model Building and the Time Series Analysis Approach to Forecasting." J. Forecasting 2(1983): 23-35.

Pierce, David A. "Least Squares Estimation in the Regression Model with Autoregressive-Moving Average Errors." Biometrika 58(1971): 299-312.

Pindyck, Robert S., and Daniel L. Rubinfeld. Econometric Models and Economic Forecasts. Second edition, New York: McGraw-Hill, 1981.