Forecast Evaluation: A Likelihood Scoring Method

by

Matthew A. Diersen and Mark R. Manfredo

Forecast Evaluation: A Likelihood Scoring Method

Matthew A. Diersen and Mark R. Manfredo*

While many forecast evaluation techniques are available, most are designed for the end user of forecasts. Most statistical evaluation procedures rely on a particular loss function. Forecast evaluation procedures, such as mean squared error and mean absolute error, that have different underlying loss functions, may provide conflicting results. This paper develops a new approach of evaluating forecasts, a likelihood scoring method, that does not rely on a particular loss function. The method takes a Bayesian approach to forecast evaluation and uses information from forecast prediction intervals. This method is used to evaluate structural econometric and ARIMA forecasting models of quarterly hog price.

Introduction

What makes a good forecast? The easiest response is that the forecast must be accurate - that is, it must predict well. Most evaluation measures in use today are simply summary statistics of the forecast errors under the assumption of a particular loss function (i.e., mean squared error assumes a quadratic loss function; mean absolute error assumes a linear loss function). A shortcoming of common statistical evaluation procedures is the possibility of obtaining conflicting results from measures which depend on alternative loss functions. For example, Ferris (1998) compares simple hog forecasts using conventional evaluation methods. His results show an example of contradicting results between adjusted mean absolute percentage error (AMAPE) and root mean square percentage error (RMSPE) and others (See Ferris, 1998, pp. 147-8). Preference for a forecast model thus depends on the evaluation measure used and the loss function of the decision maker.

Forecasters are concerned that the evaluation measures match the loss function of the forecast user. Generally, forecast users can simply use the forecast model that optimizes the evaluation measure corresponding to their loss function, e.g., mean squared error. Because of this, forecasters often report many forecasting models and evaluation measures. The forecaster himself may not have a well defined loss function and/or may not know the forecast user's loss function, causing the forecaster to struggle over which model to use. Even without contradictory evaluation measures, the forecaster may struggle since the measures employed tend to give a synopsis of performance. For example, mean squared error (MSE) of zero represents a perfect forecasting model. However, because MSE is simply relative to zero, no benchmark level of MSE exists to tell a forecaster when the model is no good. Forecasters usually have good

* Diersen and Manfredo are doctoral candidates in the Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign.

knowledge of their models and it seems shortsighted to discard or keep a forecast based solely on MSE or other summary measures.

The objective of this research is to examine evaluation from a forecaster's perspective and specifically 1) to develop an additional tool that measures forecast performance using the information from the prediction interval and 2) to apply the new method to a set of actual forecasts. The proposed evaluation method uses information easily obtainable by forecasters – the prediction interval of a forecast. Each forecasted point is assigned a score based on the likelihood of it coming from a distribution that is consistent with the forecasting model.

The remainder of the paper begins with a brief literature review that provides background information and motivation for the development of the likelihood scoring technique. The likelihood scoring method is then formally developed. Quarterly hog prices are modeled using regression and ARIMA methods, and the models are used to generate out-of-sample forecasts. The forecasting models are then compared using the likelihood scoring and traditional techniques. Finally, implications and avenues for further research are discussed.

Literature Review

Allen (1994) provides a general survey of forecasting methods. There are as many evaluation techniques as models. Allen (1994) further summarizes evaluation measures commonly used in price analysis, such as turning point measures and accuracy measures. While turning points are useful when taking market positions, in this paper we focus solely on accuracy

Kennedy (1992) provides an overview of forecast error sources and an outline of the loss functions or cost of being wrong assumed by different evaluation measures. Specification, conditioning, sampling and random error may lead to inaccurate forecasts. Forecasters have to work backward from forecast errors to improve their models by eliminating errors. For example, mean squared error can be decomposed to find the bias component. If this component is large, then the forecasting model can be examined to correct the bias.

In addition to MSE and similar measures, a simple forecast evaluation technique is to count the number of actual outcomes that fall within given confidence intervals. Makridakis et al. (1987) use this approach to assess a large number of forecast models. They found that most confidence intervals used were too narrow. Chatfield (1993), however, points out that some of the interval formulas used by Makridakis et al. (1987) were inaccurate.

The counting method has been formalized by Christofferson (1998) who addresses the confidence intervals in a calibration framework. Forecasts are evaluated based on how well different confidence intervals encompass actual observations. Implementing this technique seems to require a large sample size as repeated observations at various confidence levels are needed.

Diebold, et al. (1997) move beyond confidence intervals to predictive densities. Their method begins by assuming an unknown forecast density. If the density is calibrated with actual outcomes then probability transforms of the forecast will follow a uniform distribution. This technique also seems to require a large sample size. While this technique says when a forecast density is accurate, it gives no direction about moving from an inaccurate to an accurate density. A simpler form of this technique is presented in Bessler and Kling (1989).

Development of Likelihood Scoring Technique

This section outlines the conceptual framework necessary for developing the likelihood scoring technique. The likelihood scoring technique proposed relies on information from forecast prediction intervals. In creating prediction intervals, estimates of the variance of a forecast are needed. This section reviews forecast error variance statistics for regression and time series models. A graphical approach to forecast evaluation, a precursor to the likelihood scoring technique, is then described followed by an explanation of prediction density in a Bayesian framework. Finally, an exact description of the likelihood scoring technique is provided.

Forecast Error Variance: Regression Models

Before computing forecast prediction intervals, the variance of a forecast is needed. Consider a simple regression forecasting model

(1) $$Y_{T+1} = \hat{\alpha} + \hat{\beta} X_{T+1}$$

where X_{T+1} is known. In classical regressions the variance of a forecast is estimated and conditioned on the current observed independent variable(s). The formula for the univariate case of the estimated forecast error variance (see Pindyck and Rubinfeld, 1991) is

(2) $$s_{Econo}^2 = s^2 \left[1 + \frac{1}{T} + \frac{\left(X_{T+1} - \overline{X}\right)^2}{\sum_{t=1}^{T} \left(X_t - \overline{X}\right)^2} \right]$$

where T is the number of in-sample observations, s^2 is the estimated variance of the regression defined as

(3) $$s^2 = \frac{1}{T-k} \sum \left(Y_t - \hat{Y}_t\right)^2$$

and k is the degrees of freedom. The forecast error variance, (2), accounts for the difference of the current independent variable from the mean, accounts for differences in sample size, and shows that there is always remaining variability. This univariate example is the most intuitive for motivating the likelihood scoring technique, however, equation (2) can also be extended to the multivariate framework such that

(4) $$s_{Econo}^2 = s^2 \left(1 + \widetilde{\mathrm{X}}(\mathrm{X}'\mathrm{X})^{-1} \widetilde{\mathrm{X}}'\right)$$

where $\tilde{\mathbf{X}}$ is the vector of the values of the independent variable in the period that the forecast is made, and $\mathbf{X}$ is the matrix of independent variables.

Forecast Error Variance: Time Series Models

The forecast error variance is somewhat different for time series models since they are dependent on the forecasting horizon.[1] The [estimated] forecast error variance for one-period ahead Box-Jenkins ARIMA models is defined as

(5) $$s^2_{ARIMA} = s^2 = \frac{1}{T-p-q}\sum\left(Y_t - \hat{Y}_t\right)^2$$

where p and q are the number of autoregressive and moving average parameters respectively.[2] For multiple-period ahead forecast horizons, the forecast error variance is

(6) $$s^2_{ARIMA} = E\left[e^2_T(l)\right] = \left(\hat{\psi}^2_0 + \hat{\psi}^2_1 + \cdots + \hat{\psi}^2_{L-1}\right)s^2$$

where $E[e^2_T(l)]$ is the expectation of the squared forecast errors (l) periods ahead and $\hat{\psi}^2_0 + \hat{\psi}^2_1 + \cdots + \hat{\psi}^2_{L-1}$ are weights that minimize the mean square forecast error. [3] Note that for one period ahead forecasts $\psi^2_0 = 1$.

Prediction Intervals

Prediction intervals are key to the understanding of the graphical approach and likelihood scoring technique to be presented. Under the assumption that the standardized error of a forecast follows a t distribution, with degrees of freedom corresponding to the estimated model, the prediction interval around a forecast is defined as

(7) $$P.I._f = \hat{Y} \pm t_{\alpha/2} s_f$$

where f refers to the forecasting model. The classical interpretation of (7) is that, in repeated sampling, $(1-\alpha)\%$ of computed prediction intervals contain the true value of Y. For regression models, prediction intervals are conditioned on the current independent variable(s).[4]

Forecast Evaluation: A Graphical Approach

Once the prediction intervals are estimated it is possible to graphically analyze forecasts. The graphical approach, while not necessarily new, is a precursor to the likelihood scoring

---

[1] For a review of estimation techniques for the forecast error variance of time-series models see Chatfield (1993).

[2] This form of the forecast error variance for time series models illustrates the unconditional case and the current independent variable(s) are not accounted for. Chatfield (1993) states that using the unconditional case could misstate variances. Conditional forms have been proposed, but are not widely used. Chatfield (1993) also criticizes approximations to the error variance for time-series models.

[3] See Pindyck and Rubenfield chapter 18 for derivation of the optimal ψ weights.

[4] The regression prediction interval widens out as the independent variables move away from the mean while for the ARIMA model the prediction interval will have a constant width.

technique. The procedure plots the forecasts and prediction intervals and attempts to identify observations with large errors that may amplify differences (and may cause contradictions) among traditional statistical evaluation measures (i.e., MSE-style criteria). Paying attention to prediction intervals allows the forecaster to examine each forecast relative to its out-of-sample realization instead of relying on summary measures to determine the superiority of a forecasting methodology. The best forecasts are identified as those that fall within their widest respective prediction intervals while poor forecasts are identified as those in which the actual outcome(s) fall outside of the widest prediction intervals. Visually examining the extent to which a realized value falls relative to its prediction interval also provides insight into the penalty nature of existing statistical evaluation methods which assume a specific loss function.

The Likelihood Scoring Technique

The likelihood scoring technique developed in this paper uses ideas from the Bayesian framework. Rather than having to continuously changing prediction intervals and counting outcomes inside and outside of the intervals, a less pedantic approach is taken. Start with a regression equation such as

(8) $$Y_{T+1} = \hat{\alpha} + \hat{\beta} X_{T+1} + e_{T+1}$$

where $e_{T+1} \sim N(0,s^2)$. Using $E(e_{T+1})=0$ and inserting X_{T+1} into (8) gives $\hat{Y}_{T+1}$, the expected value of the forecast. In a Bayesian framework, the estimated coefficients and the error term follow t distributions. This implies that other values of Y_{T+1} can also be characterized by a t distribution (See Hey, 1983 and Zellner, 1971). Because Y_{T+1} is a combination of t distributions,

(9) $$\frac{Y_{T+1} - \hat{Y}_{T+1}}{s_f} \sim t_{T-\kappa}$$

where s_f is from the corresponding error variance in (2) or (5) dependant on the forecasting model and k is the number of estimated parameters. Therefore, (9) provides a prediction density instead of a prediction interval and allows a forecaster to model forecasted points as coming from a t distribution. Hey (1983, p. 233) calls this prediction density a "posterior assessment" and stresses that it is a "complete characterization" of Y_{T+1}.

Using this framework, once a prediction density is known or assumed, actual outcomes can be assessed relative to the prediction density. In (9), note that the expected outcome and standard error are fixed when considering the density. The prediction density, as characterized by (9) is shown in figure 1. This figure shows one forecast period. Outcomes can be tested to determine the likelihood that actual outcomes come from the underlying forecast distribution.[5] Once a forecasted outcome, Y_{T+1}^{Actual}, is observed it can be compared to the distribution (as shown at point l on the horizontal axis in figure 1). If the forecast was perfect then the actual would match the mean forecast and lie at the center of the t distribution where $\hat{Y}_{T+1} = Y_{T+1}^{Actual}$. The actual outcome is then used to compute a t-score using (9). The probability density function,

---

This approach is conceptually similar to the veridical approach described in Bunn (1988).

p.d.f., for the t distribution evaluated at the t-score defines the likelihood score for that observation (shown at the point f(l) on the vertical axis in figure 1).

By comparing likelihood scores for different forecasting models (i.e., causal vs. time series) a "best" forecast model can be chosen. We suggest using a summation of scores over the out-of-sample period. The forecasting model which produces the highest likelihood score would be considered the best model. Intuitively, this best model would be the one to have most likely produced the observed outcomes. Unlike traditional statistical evaluation procedures, this approach is not dependent on a particular loss function. This approach is particularly useful since it accounts for different standard errors and past history or knowledge of the separate forecasting models. Such a perspective rewards a forecast when it occurs within its learned history instead of penalizing mistakes.

Empirical Procedure

The following section provides an empirical application of the likelihood scoring method. Quarterly hog prices are modeled using both a structural econometric model and a Box-Jenkins ARIMA model. Out-of-sample forecasts from these models are computed. Subsequently the likelihood score is shown for these two models. Forecast performance is then evaluated using the likelihood score in addition to traditional methods.

Data

The data used in this study are quarterly data spanning from 1976.1 to 1997.3.[6] The data come from various government publications. Quarterly hog prices are for 230 pound barrows and gilts for 7 markets as reported in Livestock, Dairy and Poultry Monthly (ERS). Other data used in the estimation of the structural model include commercial cattle slaughter (1,000 hd.) and federally inspected, ready-to-cook, young chicken slaughter (Poultry Slaughter, NASS). Ten state sow farrowings (1,000 hd) are taken from the quarterly Hogs and Pigs report (NASS), while the hog-corn ratio is computed as the ratio of quarterly hog price to the U.S. average price of number 2 yellow corn received by farmers (Agricultural Prices, NASS).

Structural Econometric Model

The econometric forecasting model is a single price equation based on a reduced form specification inspired by Brandt (1985) and Leuthold, et al. (1989). The future price of hogs, P_{t+1}, is dependent on beef slaughter (BS_t), chicken slaughter (CS_t), the hog-corn ratio (HC_t), sow farrowings lagged one period (SF_{t-1}) and two periods (SF_{t-2}), and quarterly dummy variables ($D2_t$,

---

6 Quarterly data are consistent with hog quarters as reported in the Hogs and Pigs report (NASS).

$D3_t$, and $D4_t$).[7] In-sample estimates are computed from 1976.1 to 1990.1. Table presents the in-sample estimates of the econometric forecasting model.

The signs are as expected for the slaughter variables (negative), for the hog-corn ratio (positive), and for farrowings lagged one period (negative). These parameter estimates are all statistically significant. The exceptions were sow farrowings lagged two periods and the second and fourth quarterly dummy variables. Sow farrowings lagged two periods was kept in the model since it alleviated some minor first order autocorrelation. The model is adequate in explaining the variability in hog prices as evidenced by an adjusted R^2 of 0.44. The Durbin-Watson is in the indeterminate range.

ARIMA Model

The time series model for the price of hogs is ARIMA $(3,1,0)x(0,0,1)_5$. A number of specifications were checked during the identification stage prior to the final model presented. The final model passed the Lung-Box Q-test for a white noise process and the residual autocorrelation and partial autocorrelation plots are clean of any identifiable effects. The parameter estimates of the ARIMA forecasting model are shown in table 2. The AR(2), AR(3) and seasonal MA terms are all statistically significant.

Forecasts

Both the econometric and ARIMA models are estimated over the time period from 1976.1 to 1990.1. The forecasting models are tested out-of-sample from 1990.2 to 1997.3 giving 30 out-of-sample observations. All forecasts are for one-period ahead. Two separate sets of out-of-sample forecasts are developed for both the econometric and ARIMA specifications to illustrate the forecast evaluation procedures put forth in this paper. One set of forecasts uses the parameter estimates shown in tables 1 and 2 for each period (fixed), while another set of forecasts are developed from updated models (updated). For the updated models, the parameter estimates for each model are updated at each out-of-sample period, but the specifications remain unchanged. Table 3 presents forecasts for both the econometric and ARIMA models using fixed parameter estimates and table 4 gives forecasts from the updated models.

Figures 2 through 5 illustrates the graphical approach described earlier. In figure 2, econometric forecasts (fixed estimates) with their corresponding 95% prediction intervals are plotted along with the actual values of hog price. In the early part of the sample the predictions look good while towards the end they fall apart. The prediction interval widens as the standard error becomes larger due to the independent variables being far from the mean. Even with the wider interval, the econometric model does a poor job of predicting those later prices. Figure 3 shows the same information for the ARIMA forecasts (fixed estimates). At first glance the

---

[7] Brandt (1985) and Leuthold et al. (1989) included per-capita income in their specifications. However, per-capita income was found to be highly correlated with chicken-slaughter over the entire sample period and was subsequently dropped.

ARIMA model looks more consistent over the entire period. The prediction intervals are fixed for this model and are narrower than the econometric intervals. Throughout the period there are actual outcomes that fall near or outside the prediction interval. Hence quite often the actual outcomes occur where the ARIMA model doesn't say they would occur. Measures such as MSE would not provide this type of insight and would not say when the model was falling apart.

Figures 4 and 5 show similar plots of forecasts, 95% prediction intervals and actual values for the updated models (both econometric and ARIMA). Notice the improved performance of the econometric model for observations in the latter part of the sample due to updating the parameter estimates. The prediction interval is narrower relative to the fixed estimate forecasts. The magnitude of the errors declines and more actual values fall within the prediction intervals. The ARIMA model is little changed. Although this graphical method is not new, it illustrates what other information is available to forecasters by showing where the realized values fall relative to the forecast distribution.

Likelihood Scores

Table 3 shows the likelihood scores for the (fixed) econometric and ARIMA models and provides information needed to compute the likelihood score. To illustrate the likelihood scoring technique, consider the first (1990.1) forecast in table 3. The actual price, P_{T+1}^{Actual}, is \$56.07, the econometric forecast, $\hat{P}_{T+1}$, is \$48.37 with a standard error, $\hat{s}_{Econo}$, of 5.16. Dividing by the standard error (5.160) normalizes the forecast error (48.37-56.07). Normalizing the forecast error gives a t-score of -1.493. Using the Bayesian approach and assuming a naïve prior, the t-score follows a t distribution (See equation 8). The likelihood score is the p.d.f. value of the t-score, 0.131. Intuitively, this is the likelihood that the actual price came from the econometric model. The maximum height of t distributions increases with the degrees of freedom, but is close to 0.4.

The likelihood scoring technique formalizes the graphical approach and helps forecasters identify how a particular model performs on its own and relative to other models. For instance, starting in 1996 it becomes readily apparent that the fixed econometric model no longer predicts prices well. The individual likelihood scores become very small for those forecasts. It is not just that fact that the forecast errors are large, but that the actual outcomes are outside of the realm the econometric model predicts. Likelihood scores for the updated models are shown in table 4. The scores are higher for more of the latter econometric forecasts. Updating the estimates gave the econometric model more information or history to consider. The ARIMA models were similar for both fixed and updated models. The individual scores are then summed to provide an overall likelihood score for the particular forecast. This total likelihood score can then be compared against more traditional forecast evaluation procedures.

The likelihood scoring technique is similar to more traditional approaches in some respects. Measures such as root mean square error (RMSE) and mean absolute error (MAE) both weight forecast errors. Likelihood scoring weights errors by standardizing by the error and then plugging that t-score into a corresponding t distribution. Table 5 gives the summed likelihood score, RMSE and MAE of the different forecasting models. The likelihood score is consistent with the results of more traditional measures for these forecasts. Note that the likelihood score rewards good performance and a higher score is preferred. For RMSE and MAE low values are better because these measures penalize errors. Between the fixed models, the ARIMA is better than the econometric forecasting model. The updated results show large improvement in the econometric model and little change in the ARIMA model performance. The discrepancy between the MAE numbers, the RMSE numbers, and the overall likelihood scores all become smaller. The likelihood scoring technique thus provides additional information in the forecast evaluation process, augmenting traditional measures. The likelihood scoring technique outlined is unique, however, since it in essence rewards good forecasts instead of just penalizing forecasts with large errors.

Summary and Implications

This paper presented and developed a new forecast evaluation technique, a likelihood scoring method. Similar to a graphical analysis, this technique utilizes information contained in forecast prediction intervals. Forecast error variance statistics for both regression models and ARIMA models were also presented. The foundations of the graphical and likelihood scoring techniques were established and demonstrated for both econometric and ARIMA forecasts of hog prices.

The forecast evaluation procedure put forth in this paper is unique in many ways. First, it is intuitive and easy to implement, incorporating readily available statistics. This method also encourages forecasters to pay more attention to the information in forecast prediction intervals as suggested by Makridakis, et al. (1987) and makes forecasters pay more attention to individual forecasts. The graphical and likelihood scoring procedures can stand alone or can be used in conjunction with more traditional forecast evaluation procedures such as mean square error. They can also work with a small number of out-of-sample observations. The likelihood scoring technique is particularly useful for forecast evaluation in cases where traditional measures, especially those with different loss functions, provide conflicting results. Most importantly, however, is the approach likelihood scoring takes; it rewards good forecasts instead of penalizing poor forecasts. This gives insight into the underlying performance of the forecasting model – instead of blindly summarizing errors.

Like the graphical approach, the likelihood scoring technique can be extended for use with a variety of forecasting procedures. For instance, a composite forecast could be evaluated (but its prediction density must account for any correlation between forecasts). Subjective forecasts, such as expert opinion forecasts, may also be evaluated using likelihood scoring methods assuming a triangular distribution if a high, average, and low forecasts are provided. Similarly for futures forecasts, the distribution derived from corresponding option premiums may be used.

References

Allen, P.G. "Economic Forecasting in Agriculture." *International Journal of Forecasting* 10(1994):81-135.

Bessler, David A. and John L. Kling. "The Forecast and Policy Analysis." *American Journal of Agricultural Economics* 71(1989):503-506.

Brandt, Jon A. "Forecasting and Hedging: An Illustration of Risk Reduction in the Hog Industry." *American Journal of Agricultural Economics* 67(1985):24-31.

Bunn, Derek W. "Combining Forecasts." *European Journal of Operational Research* 33(1988):223-229.

Chatfield, Chris. "Calculating Interval Forecasts." *Journal of Business & Economic Statistics* 11(1993):121-135.

Christofferson, Peter F. "Evaluating Interval Forecasts." *International Economic Review*, in press, 1998.

Diebold, Francis X., Todd A. Gunther, and Anthony S. Tay. "Evaluating Density Forecasts." Working paper. 23 August 1997. http://www.ssc.upenn.edu/~diebold/.

Ferris, John N. *Agricultural Prices and Commodity Market Analysis*. Boston, MA:WCB McGraw-Hill, 1988.

Hey, John D. *Data in Doubt*. Oxford: Martin Robertson, 1983.

Kennedy, Peter. *A Guide to Econometrics*. Cambridge, MA: The MIT Press, 1992.

Leuthold, Raymond M., Philip Garcia, Brian D. Adam, and Wayne I. Park. "An Examination of the Necessary and Sufficient Conditions for Market Efficiency: The Case of Hogs." *Applied Economics* 21(1989):193-204.

Makridakis, Spyros, Michele Hibon, Ed Lusk, and Moncef Belhadjali. "Confidence Intervals: An Empirical Investigation of the Series in the M-Competition." *International Journal of Forecasting* 3(1987):489-508.

Pindyck, Robert S. and Daniel L. Rubinfeld. *Econometric Models and Economic Forecasts*. New York, NY: McGraw-Hill, Inc., 1991.

Zellner, Arnold. *An Introduction to Bayesian Inference in Econometrics*. New York: John Wiley & Sons, Inc., 1971.

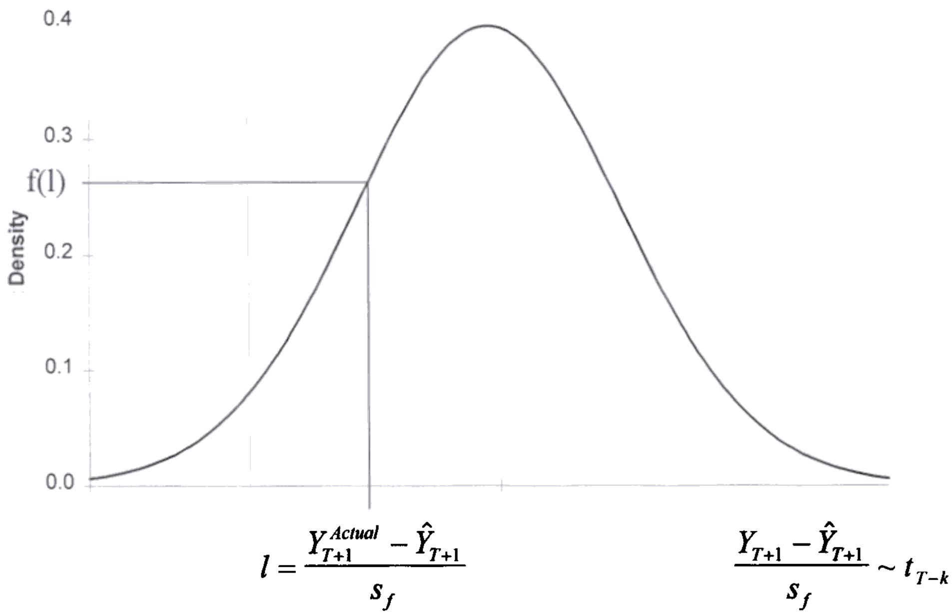

Figure 1. Illustration of likelihood score for a single forecasted point

Table 1. Structural Econometric Forecasting Model of Quarterly Hog Prices

$$P_{T+1} = \underset{(3.59)}{142.68} - \underset{(-2.29)}{0.00459\,BS_T} - \underset{(-1.95)}{0.0000139\,CS_T} + \underset{(2.08)}{0.294\,HC_T} - \underset{(-4.45)}{0.0323\,SF_{T-1}} + \underset{(1.66)}{0.0115\,SF_{T-2}}$$

$$- \underset{(-0.94)}{2.449\,D2_T} + \underset{(3.73)}{14.756\,D3_T} + \underset{(0.08)}{0.251\,D4_T}$$

Adjusted $R^2 = 0.44$ F-Statistic = 6.39 D.W. = 1.33 d.f. = 47

Note: Values in parentheses are t-statistics.

Table 2. ARIMA Forecasting Model of Quarterly Hog Prices

$$P_{T+1}-P_T = \underset{(0.60)}{0.107} - \underset{(-1.51)}{0.193\,(P_T-P_{T-1})} - \underset{(-2.94)}{0.344\,(P_{T-1}-P_{T-2})} - \underset{(-3.59)}{0.443\,(P_{T-2}-P_{T-3})} - \underset{(18.56)}{0.876\,\varepsilon_{T-4}}$$

Note: Values in parentheses are t-statistics.

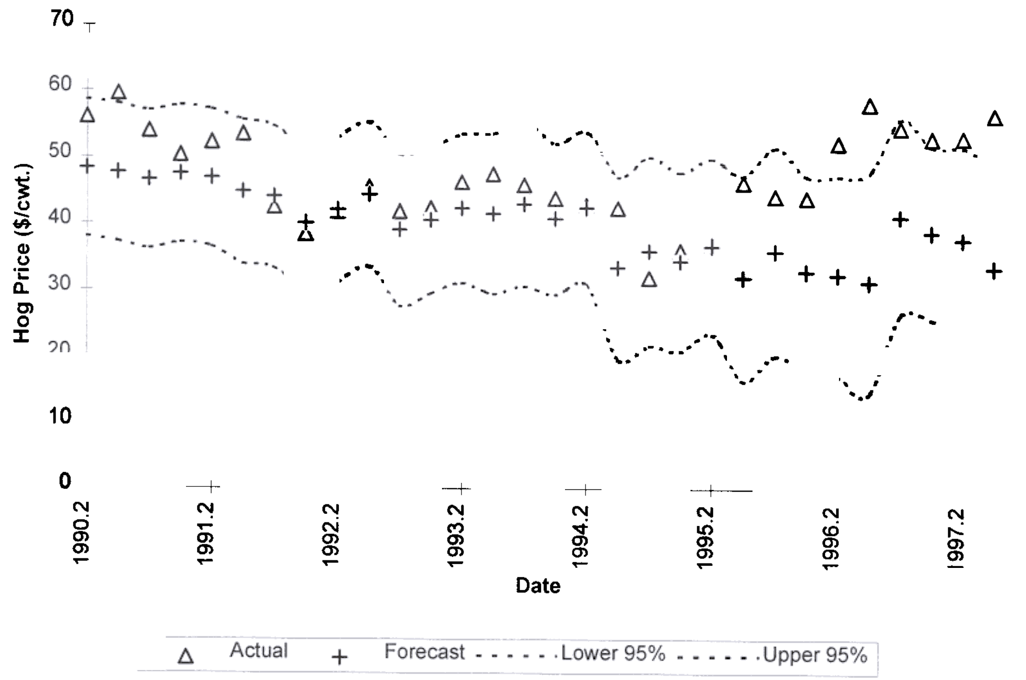

Figure 2. Forecasts from econometric model using fixed coefficients

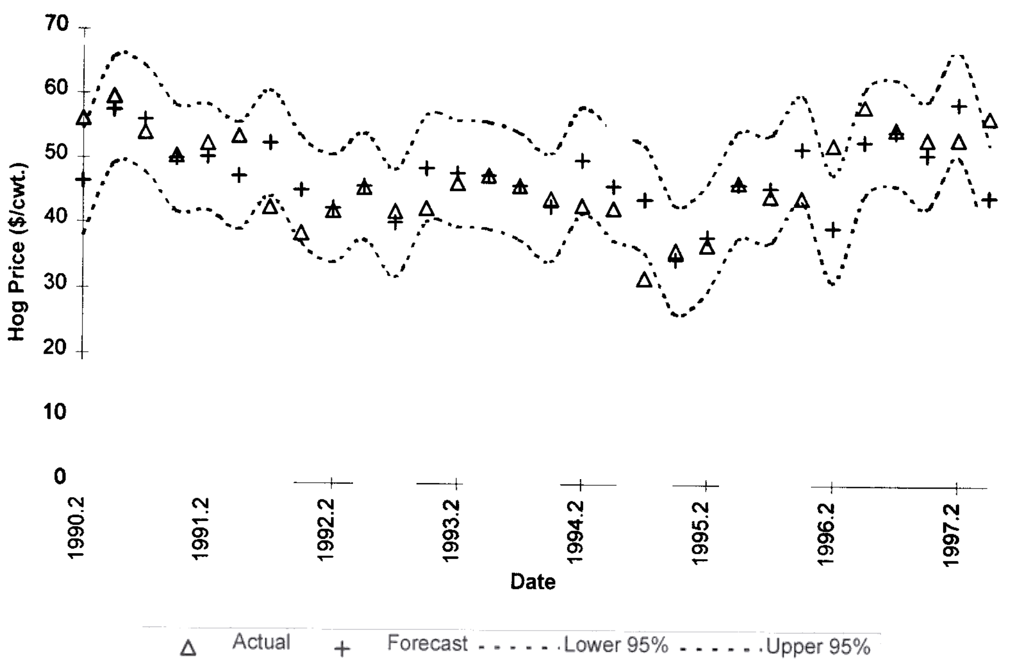

Figure 3. Forecasts from ARIMA model using fixed coefficients

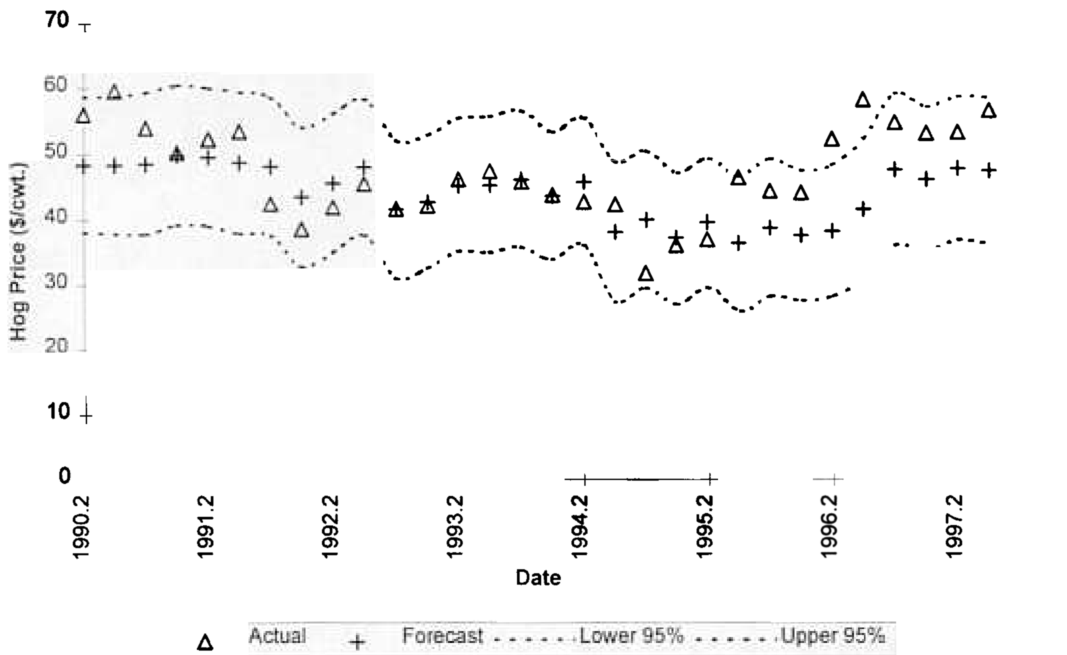

Figure 4. Forecasts from econometric model using quarterly updated coefficients

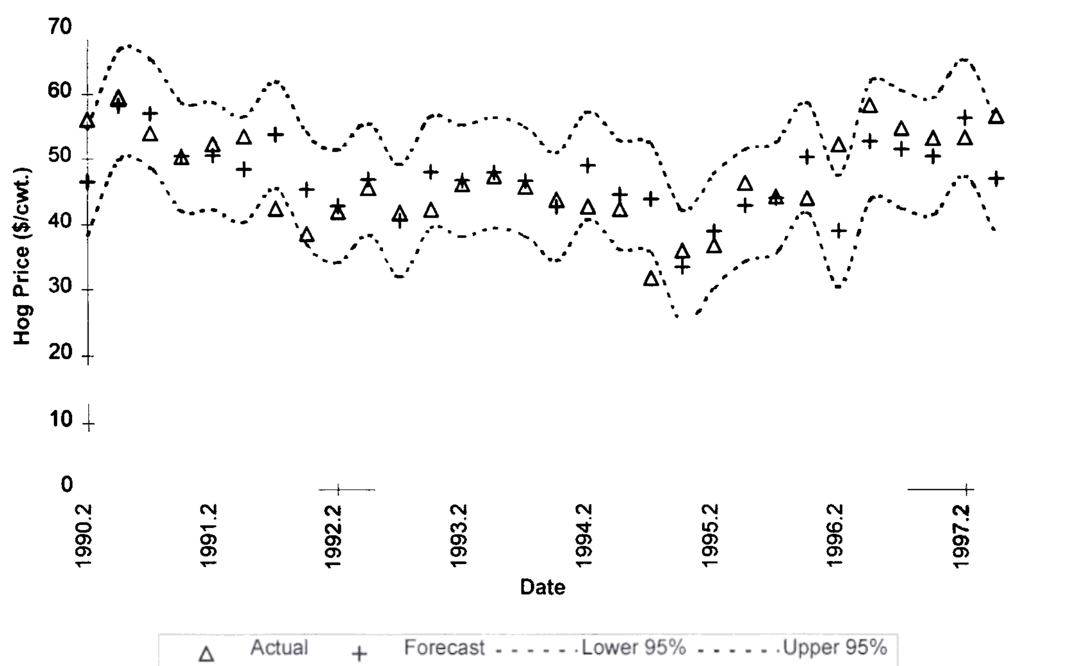

Figure 5. Forecasts from ARIMA model using quarterly updated coefficients

Table 3. Forecasts and Likelihood Scores from Models with Fixed Coefficients

		Econometric Model				ARIMA Model			
Quarter	Actual	Forecast	Std. Error	t-score	Likelihood Score	Forecast	Std. Error	t-score	Likelihood Score
	56.07	48.37	5.160	-1.493	0.131	46.41	4.13	-2.338	0.028
	59.56	47.76	5.184	-2.276	0.032	57.39	4.13	-0.526	0.345
	53.98	46.64	5.181	-1.418	0.145	55.96	4.13	0.478	0.353
	50.36	47.53	5.155	-0.549	0.340	50.01	4.13	-0.084	0.396
	52.35	46.96	5.167	-1.043	0.229	50.26	4.13	-0.507	0.348
	53.52	44.90	5.420	-1.590	0.113	47.33	4.13	-1.497	0.130
	42.50	44.14	5.357	0.306	0.378	52.42	4.13	2.401	0.025
	38.59	40.16	5.397	0.290	0.380	45.18	4.13	1.597	0.112
	41.99	42.18	5.475	0.036	0.397	42.42	4.13	0.106	0.395
	45.61	44.49	5.457	-0.205	0.388	45.83	4.13	0.053	0.396
	41.88	39.12	5.728	-0.482	0.353	40.17	4.13	-0.414	0.364
	42.39	40.61	5.543	-0.321	0.377	48.64	4.13	1.513	0.127
	46.32	42.38	5.649	-0.698	0.310	47.85	4.13	0.371	0.370
	47.52	41.58	6.067	-0.978	0.245	47.53	4.13	0.002	0.397
	45.92	43.01	6.203	-0.470	0.355	46.01	4.13	0.022	0.397
	43.91	40.83	5.730	-0.539	0.342	42.68	4.13	-0.298	0.379
	42.90	42.42	5.766	-0.083	0.395	49.90	4.13	1.695	0.095
	42.45	33.44	6.890	-1.307	0.168	45.94	4.13	0.846	0.276
	31.89	35.96	7.120	0.572	0.336	43.89	4.13	2.906	0.007
	36.06	34.39	6.760	-0.247	0.385	34.72	4.13	-0.325	0.376
	36.93	36.70	6.637	-0.035	0.397	38.12	4.13	0.288	0.381
	46.44	31.98	7.763	-1.863	0.072	46.25	4.13	-0.046	0.397
	44.43	35.98	7.850	-1.076	0.221	45.62	4.13	0.289	0.380
	44.17	32.88	7.300	-1.546	0.121	51.78	4.13	1.841	0.074
	52.38	32.38	7.556	-2.647	0.014	39.63	4.13	-3.087	0.005
	58.38	31.34	8.368	-3.232	0.003	52.92	4.13	-1.322	0.165
	54.88	41.29	7.428	-1.829	0.076	54.45	4.13	-0.105	0.395
	53.30	38.98	6.548	-2.187	0.039	51.01	4.13	-0.555	0.339
	53.43	37.89	7.069	-2.199	0.038	58.92	4.13	1.328	0.164
	56.76	33.66	7.924	-2.915	0.007	44.44	4.13	-2.983	0.006

Table 4. Forecasts and Likelihoods from Models with Updated Coefficients

		Econometric Model				ARIMA Model			
Quarter	Actual	Forecast	Std. Error	t-score	Likelihood Score	Forecast	Std. Error	t-score	Likelihood Score
	56.07	48.37	5.160	-1.493	0.131	46.57	4.131	-2.299	0.031
	59.56	48.33	5.236	-2.145	0.042	58.24	4.293	-0.308	0.378
	53.98	48.56	5.380	-1.007	0.238	57.04	4.254	0.719	0.306
	50.36	49.83	5.340	-0.099	0.395	50.47	4.218	0.025	0.397
	52.35	49.59	5.241	-0.527	0.345	50.60	4.197	-0.417	0.364
	53.52	48.76	5.392	-0.883	0.268	48.50	4.132	-1.215	0.189
	42.50	48.23	5.292	1.082	0.220	53.81	4.173	2.709	0.012
	38.59	43.53	5.309	0.930	0.257	45.44	4.380	1.563	0.118
	41.99	45.66	5.300	0.692	0.312	42.92	4.421	0.210	0.388
	45.61	48.10	5.187	0.479	0.353	46.95	4.359	0.308	0.379
	41.88	41.80	5.285	-0.015	0.397	40.73	4.370	-0.263	0.384
	42.39	42.88	5.104	0.096	0.395	48.13	4.335	1.325	0.165
	46.32	45.34	5.087	-0.193	0.390	46.82	4.340	0.116	0.395
	47.52	45.39	5.206	-0.409	0.365	48.10	4.310	0.135	0.394
	45.92	46.26	5.236	0.065	0.396	46.76	4.276	0.196	0.390
	43.91	43.77	4.909	-0.029	0.397	42.82	4.211	-0.260	0.384
	42.90	45.88	4.882	0.611	0.329	49.09	4.176	1.482	0.133
	42.45	38.28	5.316	-0.784	0.291	44.64	4.254	0.516	0.347
	31.89	40.18	5.205	1.592	0.112	44.02	4.236	2.864	0.008
	36.06	37.21	5.018	0.230	0.387	33.62	4.442	-0.549	0.341
	36.93	39.59	4.906	0.542	0.342	39.11	4.420	0.493	0.351
	46.44	36.38	5.118	-1.965	0.059	43.06	4.395	-0.770	0.295
	44.43	38.80	5.267	-1.069	0.224	44.22	4.333	-0.048	0.397
	44.17	37.66	5.001	-1.302	0.170	50.43	4.303	1.454	0.138
	52.38	38.30	5.027	-2.801	0.009	39.15	4.363	-3.032	0.005
	58.38	41.67	5.358	-3.119	0.004	52.85	4.583	-1.206	0.192
	54.88	47.70	5.765	-1.246	0.182	51.64	4.586	-0.706	0.309
	53.30	46.22	5.525	-1.281	0.175	50.52	4.566	-0.609	0.330
	53.43	47.85	5.461	-1.021	0.235	56.43	4.547	0.659	0.319
	56.76	47.53	5.584	-1.653	0.102	47.08	4.534	-2.135	0.042

Table 5. Comparison of Evaluation Measures

	Econometric Model			ARIMA Model		
	Likelihood Score	RMSE	MAE	Likelihood Score	RMSE	MAE
Fixed	6.786	10.417	7.730	7.623	5.715	4.106
Updated	7.523	6.589	5.157	7.879	5.543	4.117