Keep up the Good Work? An Evaluation of the USDA's Livestock Price Forecasts

Dwight R. Sanders

and

Mark R. Manfredo*

Paper presented at the NCR-134 Conference on Applied Commodity Price
Analysis, Forecasting, and Market Risk Management
St. Louis, Missouri, April 21-22, 2003

*Dwight R. Sanders (DwightS@siu.edu) is an Assistant Professor of Agribusiness Economics at Southern Illinois University, Carbondale, Illinois. Mark R. Manfredo (manfredo@asu.edu) is an Assistant Professor in the Morrison School of Agribusiness and Resource Management at Arizona State University.

Keep up the Good Work? An Evaluation of the USDA's Livestock Price Forecasts

Practitioner's Abstract

One step-ahead forecasts of quarterly live cattle, live hog, and broiler prices are evaluated under two general approaches: accuracy-based measures and the ability to categorize price movements directionally or within a forecasted range. Results suggest USDA price forecasts are not optimal. Broiler price forecasts are biased, and all the forecast series tend to repeat errors. While the USDA forecasts are more accurate that those of a univariate AR(4) time series model, the evidence suggests that live cattle forecasts could be improved with a composite forecast. However, the USDA correctly identifies the direction of price change in at least 70% of its forecasts. Prices fall within the USDA's forecasted range 48% of the time for broilers but only 35% for hogs. Finally, there is some evidence that the USDA's price forecasting accuracy has improved over time for broilers, but it has gotten marginally worse for hogs.

Keywords: forecast evaluation, forecast efficiency, USDA forecasts

Introduction

For agricultural producers and agribusinesses, prices directly impact the costs and revenues that drive a firm's profitability. For example, Tyson Foods, Inc. is involved in the production and processing of the three major proteins: chicken, beef, and pork. In their public announcements, Tyson clearly indicates that fluctuating protein prices directly impact their corporate earnings. Furthermore, Tyson's earnings projections necessarily rely on "forward-looking statements" about market prices (Tyson Foods, Inc.). To provide meaningful guidance to industry analysts, it is important that Tyson—and firms like them—understand and evaluate available price forecasts. Likewise, for smaller, private firms or producers, price forecasts are crucial for planning business operations and making investments.

Given the importance of prices in agricultural, it is not surprising that making and evaluating commodity price forecasts has long been an area of interest for economists (Green; Pettee). In particular, forecasts provided by public agencies such as the United States Department of Agriculture (USDA) or the National Agricultural Statistics Service (NASS) are of interest. Producers, agribusinesses, and financial institutions use these forecasts to make production, marketing, and lending decisions (NASS). The objective of the market outlook is to increase profits, utility, or social welfare through more efficient economic decisions (Freebairn). Accurate public information can result in improved decision-making by private forecasters, while public forecasts can reduce market price variation (Smyth). Conversely, systematic errors in forecasts could lead to a misallocation of scarce resources (Stein). Thus, it is important that industry participants understand the uncertainty surrounding USDA price forecasts as well as any systematic biases or inefficiencies (Aaron).

Most research examines the performance of USDA quantity or production forecasts in both crops (Irwin, Good, and Gomez) and livestock (Bailey and Brorsen). For example, Bailey and Brorsen document that USDA's annual beef and pork production forecasts are biased predictors over the entire 1982-1996 interval. Specifically, there is a tendency for the USDA to underestimate production over long horizons. Sanders and Manfredo find that the USDA's one quarter-ahead production forecasts are inefficient in that they are too extreme (not minimum variance). These findings lead one to question the efficiency of the USDA's price forecasts published in *World Agricultural Supply and Demand Estimates* (*WASDE*). It can be argued that these price forecasts are possibly more important than quantity forecasts. Certainly for small firms or producers, aggregate quantity forecasts do not directly impact their businesses. Rather, it is the resulting price that determines their costs or revenues. In this sense, the following research is an important extension to the existing literature.

Despite their importance for agribusiness decision-makers, USDA price forecasts have not been closely scrutinized. An exception is Elam and Holder who evaluate the USDA's price forecasts for rice. They find that the USDA's public forecasts compared favorably to those made by a univariate Box-Jenkins model. In related work, Kastens, Schroeder, and Plain document that the USDA's livestock price forecasts are not as accurate as those provided by extension economists. While these studies' comparative findings are important, they do not provide a complete picture of the USDA's forecasting performance. In this sense, USDA price forecasts lack a thorough evaluation. The following research fills the void in the literature by providing a comprehensive examination of the USDA's livestock price forecasts.

The analysis is comprehensive in the sense that it does not focus on a single aspect of forecast evaluation, such as traditional accuracy measures. Instead, the forecasts are assessed using multiple tests separated into two general areas: accuracy-based tests and classification-based tests. The accuracy-based tests include: 1) error measurements, 2) optimality as measured by bias and efficiency, 3) forecast encompassing with respect to time series forecasts, and 4) improvement through time. Accuracy-based tests are well established in the literature, and they usually hinge on a mean squared error loss function. The second area of evaluation is classification-based tests, which include: 1) directional accuracy or market timing, and 2) the probability that prices fall within a forecasted price range. Classification-based tests are binomial in nature and often rely on nonparametric statistical tests. USDA price forecasts have not previously been examined in this framework. Furthermore, in both of the classification-based approaches, statistics are calculated to test the USDA's performance versus a simple alternative. This analysis provides a methodological extension over prior research. Collectively, this array of testing procedures provides a complete picture of the USDA's ability to forecast livestock prices. The results may allow industry participants to more efficiently utilize the USDA's outlook information, increasing the efficiency and accuracy of their economic decisions.

Data

This study analyzes the performance of one-quarter ahead price forecasts for slaughter cattle, hogs, and broilers published by the USDA in the *WASDE* reports. The price forecasts are for 1100-1300 pound Nebraska slaughter cattle (direct trade), 51-52% lean hog carcasses (live equivalent, national base), and 12-city average wholesale broiler prices.[1] Price forecasts are published as an expected range. For instance, the USDA's slaughter cattle forecast for the first quarter of calendar year 2002 is $66.00-$68.00 per hundredweight. The mid-point of the forecasted range ($67.00) is used as the point forecast for the quarter.

The *WASDE* report is published on a monthly basis, and is released between the 8th and the 14th of each month. Because of this, the price forecasts for cattle, hogs, and broilers are collected from the January, April, July, and October reports for each calendar quarter. For example, the forecasted price for the first calendar quarter is collected from the January report. This data collection process results in a series of non-overlapping, independent, rolling-event forecasts; thus, it negates the problem of inconsistent OLS standard error estimates that stem from overlapping forecast horizons (Brown and Maital; Clements and Hendry p. 57). The actual (realized) price levels are taken from subsequent releases of *WASDE* reports to assure that they correctly match the prices that the USDA is attempting to forecast. The sample period is from the third quarter of 1982 (1982.3) to the third quarter of 2002 (2002.3), resulting in 81 quarterly observations of one step-ahead price forecasts and realized values.

As is well known, livestock prices demonstrate seasonal patterns as a result of natural fluctuations in production. For instance, hog prices tend to be higher in the summer, corresponding to the seasonal low in pork production. Therefore, the analyses focus on seasonal differences defined as the log-relative price change from the same quarter of the prior year. Define A_t as the actual price level in quarter t and F_t is the one-step ahead price forecast for quarter t. The change in actual prices is defined as $AP_t=\ln(A_t/A_{t-4})$, and the forecasted price change is $FP_t=\ln(F_t/A_{t-4})$. Thus, changes reflect the percent change in the quarterly average price from the prior year. This framework is consistent with that used by most industry analysts (e.g., Hurt; Kastens, Schroeder, and Plain).[2,3]

Methodology and Results

The objective of this research is to fully evaluate the USDA's price forecasts for cattle, hogs, and broilers. To do this, the USDA's forecasts are compared to those of a simple time series model (Granger). Past research has shown that simple ARIMA models perform comparably to more sophisticated VAR-style models (Brandt and Bessler). Therefore, $AP_t=\ln(A_t/A_{t-4})$ is modeled as an autoregressive process with four lags. The model was specified and estimated over the out-of-sample data from 1970.1 through 1982.2. The AR(4) model fit the data well and the residual autocorrelation and partial autocorrelation functions were not statistically significant out to eight lags. The AR(4) model represents a simple, low-cost alternative to the USDA's forecasts. The result is a series of 81 one-quarter ahead forecasts for AP_t from 1982.3 through 2002.3. These are used as a standard for comparison in the following tests.

Accuracy-Based Tests

The following tests all relate to the accuracy of the USDA's point forecast, which is assumed to be the midpoint of the forecasted range. In most instances, this is expressed either directly or indirectly as a mean squared error loss function. The following tests are based on well-established procedures in the literature. Therefore, the focus is on the results and the relative performance of the USDA's forecasts.

Measures of Forecast Error

Summary statistics for each series are presented in Table 1. Mean and standard deviation are measured as the percent price change from the prior year. None of the mean actual price changes are statistically different from zero (5% level, two-tailed t-test). Furthermore, hog prices are the most volatile with a standard deviation of 21.83%, more than double the volatility of cattle prices (8.31%). This is consistent with the relatively high volatility displayed by pork production (Sanders and Manfredo).

Traditional measures of forecast error are presented in Table 2. The statistics reported are root mean squared error (RMSE), mean absolute error (MAE), and Theil's U.[4] By all measures, the USDA forecasts are more accurate than the time series forecasts across the three sectors. Harvey, Leybourne, and Newbold (1997) suggest that differences in accuracy measures are best tested with their modified version of the Diebold-Mariano test. Given two time series of 1-step ahead forecast errors (e_{1t},e_{2t}), and a specified loss function g(e), then the null hypothesis of equal expected forecast performance is $E[g(e_{1t})-g(e_{2t})] = 0$. The modified Diebold-Mariano (MDM) test is based on the sample mean of $d_t = g(e_{1t})-g(e_{2t})$ with the test statistic having a t-distribution. The t-statistic from the MDM test is presented in the bottom two rows of Table 1. Using this test, the RMSE and MAE produced by the USDA's forecasts are statistically smaller (5% level) than those of the time series model.

Theil's U normalizes forecast errors by the volatility of the underlying series, so it provides some basis of comparison across the three markets. Theil's U has a lower bound of zero for perfect forecasts, and it takes a value of unity for naïve "no change" forecasts (Leuthold). As expected, both the USDA and time series forecasts offer superior performance over a "no change" forecast. Looking across the markets, Theil's U indicates that the most improvement over a "no change" forecast occurs in hogs. This suggests that much of the underlying volatility in hog prices is predictable.

Tests for Optimality—Bias and Efficiency

A forecast is optimal if it is unbiased and efficient (Diebold and Lopez). Granger and Newbold (p. 286) suggest that efficiency tests focus strictly on forecast errors, $e_t = AP_t - FP_t$, to avoid interpretive issues associated with the traditional linear regression based test (Holden and Peel). The following tests for bias and efficiency use the methodology demonstrated by Pons (2000).

The test for forecast bias relies on an OLS regression of forecast errors (e_t) on an intercept term (γ) such that:

$$e_t = \gamma + \mu_t. \qquad (1)$$

Given that optimal forecast errors should have a mean of zero (Diebold and Lopez), the null hypothesis of an unbiased forecast is $\gamma = 0$. This hypothesis is tested using a two-tailed t-test.

The estimation results for (1) are presented in Table 3.[5] The USDA forecasts are unbiased for cattle and hog prices, but they consistently overestimate broiler prices ($\gamma < 0$). The bias in broiler price forecasts is a statistically significant -2.42%. In contrast, the time series forecasts statistically underestimate all the price series ($\gamma > 0$). It is not clear why the time series forecasts demonstrate this bias. It could result from a misspecification of the autoregressive process, structural changes in the livestock industry that are difficult to capture in time series models, or the use of data from a period of rapid commodity inflation in the early 1970's.

To further explore the optimality conditions of these forecasts, tests for forecast efficiency are conducted. Forecasts are weakly efficient if e_t is orthogonal to both the forecast, as well as prior forecast errors (Nordhaus). Thus, weak efficiency is tested using the following regression framework (Pons, 2000):

$$e_t = \alpha_1 + \beta \, FP_t + \mu_t, \qquad (2)$$

and

$$e_t = \alpha_2 + \rho \, e_{t-1} + \mu_t. \qquad (3)$$

A condition for efficiency is that β=0 in (2) and ρ=0 in (3). These hypotheses are tested using a two-tailed t-test on the estimated parameters.

The results of estimating equation (2) are presented in Table 4. For the USDA forecasts, the null hypothesis of weak efficiency ($\beta = 0$) is not rejected at the 5% level for any of the price forecasts. So, the forecasts efficiently incorporate the utilized information set. This is in contrast to the inefficiency of USDA livestock production forecasts documented by Sanders and Manfredo. The time series forecasts also fail to reject the null of efficiency.

The results from estimating equation (3), presented in Table 5, show there is a consistent tendency across the three markets for the USDA to repeat like errors. That is, the estimated ρ is positive for all forecast series, and it is statistically significant (5% level) for cattle and broilers. So, past forecast errors have some tendency to be repeated. For instance, the estimated ρ for beef is 0.2465. Given this, if the previous quarter's forecast error is 5%, then the current quarter's forecast should be adjusted by subtracting 1.2325% (0.2465 x 0.05 = 0.012325). This positive serial correlation in the forecast errors ($\rho > 0$) could be caused by difficulty in modeling structural changes or slowly evolving price cycles in the livestock industry. Interestingly, the USDA's tendency to repeat price forecasting errors is consistent with the positive correlation in livestock production forecast errors reported by Sanders and Manfredo. There is no statistically significant error repetition in the time series models.

Forecast Encompassing

If a preferred forecast encompasses an alternative forecast, then the alternative forecast provides no useful information beyond that provided in the preferred forecast (Harvey, Leybourne, and Newbold, 1998). Therefore, there is no linear combination between the preferred and alternative forecast that could produce a mean squared error which is smaller than that produced by the preferred forecast (Mills and Pepper). Here, forecast encompassing is tested using the following OLS regression framework:

$$e_{1t} = \alpha_3 + \lambda(e_{1t} - e_{2t}) + \varepsilon_t. \qquad (4)$$

In equation (4), e_{1t} represents the forecast error series of the preferred forecasts, while e_{2t} is the forecast error series of the competing forecast. The estimated λ $(1-\lambda)$ is the weight placed on the competing (preferred) forecast in forming the optimal composite predictor. The null hypothesis that the covariance between e_{1t} and $(e_{1t} - e_{2t})$ is zero, $\lambda = 0$, is tested against the single tailed alternative, $\lambda > 0$ (Harvey, Leybourne, and Newbold, 1998).

Harvey, Leybourne, and Newbold (1998) show that the traditional regression based test in equation (4) is oversized in small samples when the forecast errors are not bivariate normal. They suggest that heavy tails are a common occurrence in the distribution of price forecast errors, and they recommend a modified Diebold and Mariano (1995) type test (MDM test). The Diebold and Mariano test statistic is computed as the ratio of the sample mean of the series, $d_t = (e_{1t} - e_{2t})e_{1t}$, divided by its sample standard error. Harvey, Leybourne, and Newbold (1998) modify the statistic through multiplying it by $n^{-1/2}[n+1-2h+n^{-1}h(h-1)]^{1/2}$, where n is the number of observations and h is the steps ahead for the forecasts. The MDM statistic is tested as a t-distribution with n-1 degrees of freedom under the null hypothesis that it equals zero versus a single-tailed alternative. The results of the MDM test are presented, along with those from traditional regression-based test, to verify the statistical results.[6]

USDA forecasts serve as the preferred models in the forecast encompassing tests. The OLS estimates of equation (4) are presented in Table 6. Using the regression-based test, the null hypothesis that the "preferred" USDA forecast encompasses the "competing" time series forecast is rejected at the 5% level only for cattle. The MDM test rejects the null that the USDA price forecasts encompass the time series forecasts at the 10% level for cattle. The MDM results generally confirm that regression-based tests of forecast encompassing may be over-sized. Collectively, the evidence suggests that USDA forecasts for hog and broiler prices appear to capture the information contained in time series forecasts, whereas USDA cattle forecasts do not. This implies that practitioners who utilize the USDA forecasts for cattle may want to supplement them with time series forecasts, and the USDA may want to incorporate time series techniques into their forecasting procedures.

Forecast Improvement

Forecast improvement over time is tested with a methodology similar to that used by Bailey and Brorsen.[7] In this test, the absolute value of the forecast errors are regressed on a time trend such that:

$$|e_t| = \theta_1 + \theta_2 \text{Trend}_t + \mu_t. \qquad (5)$$

If $\theta_2=0$, then there is no systematic increase or decrease in the absolute value of the forecast error, $|e_t|$, over time. Rejection of this null hypothesis would suggest that forecasts either improved ($\theta_2 < 0$) or worsened ($\theta_2 > 0$) over time. This hypothesis is tested using a two-tailed t-test with results presented in Table 7.

The θ coefficient estimate is statistically less than zero at the 5% level for broilers, indicating that the absolute forecast errors have become smaller. In contrast, USDA hog forecasts show a modest decline in accuracy through time (p-value =0.1328).[8] Because the AR(4) models parameters are estimated with greater precision as the estimation sample increases, one might expect the time series forecast errors to decline through time. However, this is not the case. Instead, it is interesting that the time series estimates of θ_2 have the same signs as those estimated for the USDA forecasts. That is, the time series forecast errors for hog prices also increased over the sample, and broiler price forecasting errors declined. This may suggest that the underlying cause of the changes in forecast performance resides in the structure of the industry and not with the forecasting method employed (see Barkema, Drabenstott, and Novack). These trends are visually apparent in the time series plots of $|e_t|$ in Figure 1.

The size of the forecast error is clearly important when an agent's loss function is represented by squared errors. However, sometimes simple binary classifications, such as the direction of price changes, are equally important to decision makers. This area of performance is assessed in the following section.

Classification-Based Tests

Henriksson and Merton demonstrate that it is only necessary for a forecast to have directional accuracy to provide value to a decision-maker. As pointed-out by McIntosh and Dorfman, the ability to predict the direction of price changes is "nontrivial" and often just as important as forecasting price levels. Therefore, in the following section we examine the directional accuracy of the forecasts using Henriksson and Merton's non-parametric approach. This is followed by an assessment of the USDA's ability to accurately categorize realized prices within a forecasted range.

Directional Accuracy

In many applications, it is important to know the direction of price change. That is, are current prices expected to move up or down in the next time period? For instance, a speculator in the livestock futures markets would certainly benefit from knowing if prices will be higher or lower in the next quarter. Likewise, a food service firm may need to know if prices in the upcoming quarter will be higher or lower than the previous year for planning and budgeting purposes. Therefore, evaluating forecasts in terms of directional accuracy is an important component of assessing their value, and complements the information provided by the accuracy-based tests.

McIntosh and Dorfman suggest the timing test proposed by Henriksson and Merton (HM) to qualitatively evaluate forecast performance. As demonstrated by Pesaran and Timmermann, Hendriksson and Merton's hypergeometric test is asymptotically equal to a chi-squared test for independence in a two-by-two contingency table (see Table 8). In Table 8, ΔF is the forecasted direction of change, ΔA is the actual direction of change, and n is the number of observations in each cell of the table. Perfect directional forecasting would be represented by $n_{21}=n_{12}=0$ or equivalently $n_{11}=N_1$ and $n_{22}=N_2$. Henriksson and Merton show that the null hypothesis of no timing ability is a test that the sum of the conditional probabilities of correct forecasts (n_{11}/FN_1 + n_{22}/FN_2) equals one and suggests a test based on the hypergeometric distribution of n_{22}. This is equivalent to a test of independence in a two-by-two contingency table (Cumby and Modest) and can be tested with a standard chi-squared test (Stekler and Schnader, 1991; Pons, 2001).

The definition of a forecasted price increase or decrease clearly depends on the base period of comparison. That is, a price change can be defined over successive time intervals (one quarter to the next) or year-over-year. Directional price changes from one quarter to the next might be important to a speculator or cash merchant, while year-over-year price changes might have greater importance to a corporate analyst whose budget is based on the previous year's prices. Therefore, it is useful to test the USDA's forecasts for directional accuracy in both of these cases (quarter-to-quarter and year-over-year price changes).

To test for quarter-to-quarter directional accuracy, we define the following variables: $\Delta F = 1$ if the price is forecasted to increase ($F_t > A_{t-1}$), and is zero otherwise, and, $\Delta A = 1$ if actual prices increase ($A_t > A_{t-1}$), and is zero otherwise. The resulting numbers are tabulated and entered into a two-by-two contingency table. A chi-squared statistic is used to test the null hypothesis of no directional forecasting ability—i.e., independence in the two-by-two contingency table. Table 9 shows the percent of directionally correct forecasts, the chi-squared statistic, and its p-value.

The USDA clearly demonstrates an ability to forecast quarter-to-quarter price direction. The USDA's price forecasts correctly predict price direction over 70% of the time for all three markets, and the results are statistically significant at the 1% level. The best directional forecasting is in live cattle with over 76% correct forecasts. Only in the case of broilers is the USDA's directional forecasts bested by those of the time series model. Overall, the USDA's price forecasts show a relatively strong ability to forecast the direction of quarter-to-quarter price changes.

It is important to determine if the USDA's directional forecasting ability is statistically better than that of the AR(4) alternative. To test this, we follow Chang, and use the normal distribution to approximate the mean and variance of Hendrikson and Merton's original test based on the hypergeometric distribution of n_{22} in Table 8. As shown by Chang, the parameters of the normal approximation are the mean and standard deviation of the hypergeometric distribution, $E(n_{22}) = FN_2\, N_2/N$ and $\sigma^2(n_{22}) = [FN_2\, N_2\, (N-N_2)(N-FN_2)]/[N^2(N-1)]$. Furthermore, the sampling distribution of the differences is also normally distributed and is given by $E(n^U_{22} - n^T_{22}) = E(n^U_{22}) - E(n^T_{22})$, and $\sigma^2(n^U_{22} - n^T_{22}) = \sigma^2(n^U_{22}) + \sigma^2(n^T_{22})$, where n^U_{22} and n^T_{22} are the number of correct forecasts for lower prices made by the USDA and the time series models, respectively.[9] The null hypothesis is that the two forecast series have equal timing ability. The z-score from the normal approximation is presented in the bottom of Table 9. It is clear that although the USDA does a relatively good job of forecasting price direction, its performance is not statistically better than the time series forecasts at conventional levels.

Similar to the quarter-to-quarter evaluation, year-over-year directional accuracy evaluation uses the following variables: $\Delta F = 1$ if the forecasted price is greater than that of a year ago ($F_t > A_{t-4}$), and zero otherwise, and, $\Delta A = 1$ if actual prices are above the prior year ($A_t > A_{t-4}$), and zero otherwise. Again, these numbers are entered into a two-by-two contingency table, and a chi-squared test for independence is performed with results presented in Table 10. Given the USDA's ability to forecast quarter-to-quarter price changes, it is not surprising that forecasts provide valuable information concerning year-over-year price changes. In fact, the USDA correctly forecasts year-over-year price changes over 80% of the time (statistically significant at the 1% level). USDA cattle, hog, and broiler forecasts are directionally correct more often than their time series counterparts. Again, however, the difference is not statistically significant (z-test, 10% level). Collectively, over all three livestock markets, the HM tests show that the USDA price forecasts provide valuable information to decision-makers interested in the direction of price changes, but there performance is not statistically distinguishable from that of the time series model.

Accuracy of Forecast Range

The USDA provides their price forecasts as a range. For instance, the forecast for broiler prices in 2002.3 was 57-59 dollars per hundredweight. Thus far, the analysis has focused on using the mid-point of this range as the USDA's point forecast. However, the range itself may provide some forecasting information. In this section, we ask a simple question: How often does the realized price fall within the USDA's forecasted range? The forecast ranges for cattle, hogs, and broilers are illustrated in Figure 2.[10]

To assess the value the USDA's forecasted price range, we calculate the percent of realized prices that fall within the forecasted range. Unlike the directional accuracy tests in the prior section, there is not a clear null hypothesis concerning how often the forecasted range should be correct by chance alone. Therefore, it is necessary to build a standard for comparison. A naïve range forecast is constructed using the previous quarter's price as the point forecast. Then, the range used by USDA is applied to this naïve point forecast. For example, the USDA's hog price range forecast for 1990.4 is 51-55 dollars per hundredweight. The previous quarter's (1990.3) actual price is 57.67 dollars per hundredweight. So, the naïve model's forecasted range for 1990.4 is 55.67-59.67 dollars per hundredweight. The proportion of times the realized price falls within the forecasted ranges is tabulated and presented in Table 11.

Actual prices fall within the USDA's forecasted price ranges 40.7%, 34.6%, and 48.1% of the time for cattle, hogs, and broilers, respectively (first row, Table 11). It is difficult to compare across markets due to their different levels of price volatility and the USDA's tendency to use the same range for each market. What is more important is how the USDA compares to the naïve alternative. As shown in the second row of Table 11, the naïve forecast's range is correct 29.6% for cattle, 34.6% for hogs, and 40.7% for broilers. The USDA's forecasted price range performs better than that of the naïve model for cattle and broilers but not for hogs. Again, it is important to test if the USDA's forecasts statistically outperform the naïve alternative. The test is the standard test for differences in sample proportions, which is normally distributed (Bender, Douglas, and Cramer, p.70).[11] The z-scores are presented in the bottom row of Table 11. In no market can the null hypothesis of equal sample proportions be rejected at the 10% level (two-tailed test). The results suggest that the range forecasts provided by the USDA are not statistically better at categorizing realized prices than those produced by the naïve alternative.

Collectively, the classification-based tests indicate that the USDA does a good job of categorizing prices. That is, it correctly forecasts quarter-to-quarter price changes at least 70% of the time and year-over-year price changes as least 80% of the time. Likewise, the USDA's forecasted price range contains the realized price at least 35% of the time. Although this appears to be a commendable performance, in this sample, it is not statistically different from that provided by the naïve alternatives.

Summary, Conclusions, and Discussion

This research examines the performance of the USDA's quarterly price forecasts for cattle, hogs, and broilers as reported in the *WASDE*. As a standard of comparison, forecasts are also generated from a univariate AR(4) time series model. This research takes a comprehensive approach to forecast evaluation. That is, the USDA forecasts are examined in the general areas of accuracy-based tests and classification-based tests.

Using accuracy-based tests, the USDA's price forecasts produce statistically smaller mean squared errors than those generated by the time series model. Also, the USDA's price forecasts are found to be unbiased for cattle and hogs, but broiler prices are systematically over estimated by 2.42%. In contrast, the time series model consistently underestimates prices in all three markets. The USDA forecasts are efficient in that they are neither too conservative nor too extreme; however, they are inefficient in that forecast errors tend to be repeated. That is, positive errors are followed by positive errors. This result is strongest in the cattle and broiler markets. The time series forecasts did not display a consistent inefficiency.

The third accuracy-based test conducted is for forecast encompassing. The USDA's hog and broiler price forecasts are conditionally efficient with respect to the time series forecasts, but there is evidence that the cattle forecasts do not encompass the time series forecasts. A forecaster may achieve greater accuracy, in a mean squared error framework, by combining the USDA cattle forecasts with those from a simple time series model.

The data suggests that broiler price became easier to forecast over time, while hog prices may be more difficult to predict. The time series models demonstrated the same increase or decrease in accuracy as shown by the USDA forecasts. This may suggest that regardless of the forecast method, hog prices became more difficult to forecast and broiler prices easier to forecast. Indeed, over the sample period, the hog industry documented the greatest structural changes, while the broiler industry was already fully integrated and the cattle sector has been slow to integrate (see Barkema, Drabenstott, and Novack). Structural shifts may partially explain the general decline in hog price forecast accuracy and increase in accuracy for broiler price forecasts.

Classification-based tests evaluate directional accuracy using the test initially proposed by Henriksson and Merton. Directional forecasts are evaluated both in terms of market direction versus the prior quarter and versus the prior year. The USDA forecasts demonstrate the ability to correctly forecast quarter-to-quarter price direction at least 70% of the time, and year-over-year price direction with at least 80% accuracy across the three markets. Despite these generally favorable results, the USDA's market timing ability is not statistically better than that of the time series model. Similarly, although the USDA's forecasted range captures realized prices at least 35% of the time, the performance is not statistically better than the naïve alternative.

Noteworthy, across the three markets examined—cattle, hogs, and broilers—the USDA's broiler forecasts seem to demonstrate the most inadequacies. That is, the broiler price forecasts were biased and suboptimal. Even so, they showed consistent improvement—through smaller absolute errors—over the sample period. Indeed, the vertically integrated structure of the broiler industry may be driving this result. The controlled nature of production, the concentration of producers, and the relatively short production cycle may make prices increasingly easy to forecast. Even with accuracy improving, the USDA's broiler forecasts are not statistically optimal, and they may want to consider changes in their forecasting procedures to correct the inefficiencies.

Collectively, the results imply that the USDA generally does an admirable job of forecasting livestock prices at a one quarter-horizon. Still, it may want to review the methods for producing quarterly livestock price forecasts. It may be possible for the agency to take steps to remove some of the documented biases and inefficiencies—i.e., composite forecasts between their current methodology and simple time series models could improve forecasting accuracy for cattle, or forecast ranges could be made more dynamic to reflect shifts in market volatility. Despite some of their shortcomings, the USDA forecasts likely provide value to industry participants. For instance, practitioners may use them to improve existing private forecasts. More importantly, the forecasts may provide value to market participants who lack the expertise, time, or resources to generate their own forecasts. Specifically, the USDA forecasts certainly outperform a naïve no-change forecast. Hence, they may provide welfare enhancement through reduced price uncertainty (Irwin, Good, and Gomez). As well, given the positive results of the directional accuracy tests, the USDA forecasts may prove useful to both traders and businesses alike who desire an indication of the direction of livestock price movements.

It is interesting that, aside from broilers, there is no improvement in the USDA's ability to forecast quarterly livestock prices from 1982 through 2002. This is despite marked advances in computing power and statistical methods. Over the twenty-year span, numerous academic articles have documented improved forecasting techniques in these markets (e.g., Goodwin, 1992). This raises an intriguing question as to whether or not the USDA and other forecasters are employing these new methodologies. If they are, do the methodological advances simply not provide improved performance in real-time forecasting? If they are not, why not? Are applied forecasters not receiving academic research results in a usable format? Or, are the new methods too costly to learn and implement? These questions are difficult to answer, but it may be a crucial next step in improving the relevance of forecasting research (Brorsen and Irwin).

Endnotes

[1] Data definitions did change over the sample period. Live cattle prices were defined as follows: choice slaughter steers, Omaha, 900-1100 pounds, 1982.3 to 1988.2; Omaha, 1000-1100 pounds, 1988.3 to 1991.1; Nebraska, direct, 1100-1300, 1991.2 to 2002.3. Live hog prices were defined as follows: barrows and gilts, seven market average, 1982.3 to 1992.2; Iowa-S. Minnesota, No. 1-3, 1992.3 to 1999.1; Iowa-S. Minnesota, live equivalent, 51-52% lean, 1999.2 1999.4; National Base, live equivalent, 51-52% lean, 2001.1-2002.3. Broiler prices changed from a nine-city average to the twelve-city average in 1983.2. Where data definitions changed, the USDA's predicted price changes are properly adjusted. In most instances, the new and old series used by the USDA closely correspond.

[2] The seasonally differenced prices, $AP_t=\ln(A_t/A_{t-4})$, and price forecasts, $FP_t=\ln(F_t/A_{t-4})$, are stationary series (augmented Dickey-Fuller tests). The results are available upon request.

[3] In the evaluation of forecast errors it does not matter if one uses year-over-year price changes, price changes over successive observations, or absolute errors. It is easily shown that $e_t = \ln(A_t/A_{t-4}) - \ln(F_t/A_{t-4}) = \ln(A_t/A_{t-1}) - \ln(F_t/A_{t-1}) = \ln(A_t) - \ln(F_t)$.

[4] For n observations, the RMSE = $(\sum e^2/n)^{0.5}$, MAE = $\sum|e|/n$, and Theil's U = $[(\sum e^2)/(\sum AP^2)]^{0.5}$.

[5] In this and all subsequent regression models, heteroskedasticity is tested using White's test and serial correlation using the Lagrange multiplier test. Heteroskedasticity is corrected using White's heteroskedastic consistent covariance estimator and serial correlation using the covariance estimator of Newey and West (Hamilton, p. 218).

[6] Note, the regression-based test in equation (4) is still necessary to estimate optimal weights assigned to the preferred (1-λ) and competing (λ) forecasts. In the event of non-normal errors, the parameter estimates are not biased. Instead, the estimated standard errors are inconsistent, which leads to the over-sizing of the encompassing tests (Harvey, Leybourne, and Newbold, 1998). The MDM statistic essentially provides an alternative test statistic for the null hypothesis that $\lambda=0$ in the regression-based test.

[7] Tests for structural change were conducted for the bias, efficiency, and encompassing tests (equations 1, 2, 3, and 4) using the Chow break point test. The third quarter of 1992 is used as the break point. The null hypothesis of no change in the parameter estimates between the two samples cannot be rejected at conventional levels.

[8] To test the robustness of these results, a Chow break-point test was also administered for equation (5) with the third quarter of 1992 serving as the breakpoint. There was not a statistically significant difference in the estimated parameters before and after 1992.3. Equation (5) was also estimated with quarterly intercept shifters to test for systematically higher or lower $|e_t|$ in particular quarters. The null hypothesis of equal parameter estimates for θ_1 across quarters could not be rejected with a standard F-test (5% level). There is no evidence that livestock price forecasting is more or less difficult in a particular quarter.

[9] This test explicitly assumes that the forecasts are generated independently. Given the time series model's reliance strictly on past prices and the USDA's use of all available information, this is not an unreasonable assumption. If this is not true, then the test's standard errors are too large and any bias is in favor of the null hypothesis.

[10] The price range for livestock forecasts seem to be based more on institutional procedure as opposed to varying market conditions. For instance, the price ranges for cattle, hogs, and broilers were predominately three or four dollars per hundredweight from 1982.3 through

1985.2. From 1985.3 through 1990.1 the range was typically four dollars per hundredweight. The range was expanded to six dollars per hundredweight from 1990.2 through 1994.2. After which, the range has consistently been two dollars per hundredweight.

[11] This test assumes that the forecasts are generated independently. A violation of this assumption biases the results toward a failure to reject the null.

References

Aaron, H.J. "Presidential Address—Seeing Through the Fog: Policymaking with Uncertain Forecasts." *Journal of Policy Analysis and Management*. 19(2000):193-206.

Bailey, D.V. and B.W. Brorsen. "Trends in the Accuracy of USDA Production Forecasts for Beef and Pork." *Journal of Agricultural and Resource Economics.* 23(1998):515-525.

Barkema, A., M. Drabenstott, and N. Novack. "The New U.S. Meat Industry." *Economic Review*. 86(2001):33-56.

Bender, F.E., L.W. Douglas, and A. Kramer. *Statistical Methods for Food and Agriculture*. AVI Publishing Company, Westport, Connecticut, 1982.

Brandt, J.A. and D.A. Bessler. "Forecasting with Vector Autoregressions Versus a Univariate Arima Process: An Empirical Example with U.S. Hog Prices." *North Central Journal of Agricultural Economics*. 6(1984):19-36.

Brorsen, B.W. and S.H. Irwin. "Improving the Relevance of Research on Price Forecasting and Marketing Strategies." *Agriculture and Resource Economics Review*. 25(1996):68-75.

Brown, B.W. and S. Maital. "What do Economists Know? An Empirical Study of Experts' Expectations." *Econometrica*. 49(1981):491-504.

Chang, E.C. "Returns to Speculators and the Theory of Normal Backwardation." *Journal of Finance*. 40(1985):193-208.

Clements, M.P. and D.F. Hendry. *Forecasting Economic Time Series*. Cambridge, United Kingdom: Cambridge University Press, 1998.

Cumby, R.E. and D.M. Modest. "Testing for Market Timing Ability: A Framework for Forecast Evaluation." *Journal of Financial Economics*. 19(1987):169-189.

Diebold, F.X. and J.A. Lopez. "Forecast Evaluation and Combination." *Handbook of Statistics 14: Statistical Methods in Finance*, G.S. Maddala and C.R. Rao, eds., Amsterdam: North-Holland, 1998.

Diebold, F.X. and R.S. Mariano. "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics*. 13(1995):253-263.

Elam, E.W. and S.H. Holder. "An Evaluation of the Rice Outlook and Situation Price Forecasts." *Southern Journal of Agricultural Economics*. (1985): 155-161.

Freebairn, J.W. "An Evaluation of Outlook Information for Australian Agricultural Commodities." *Review of Marketing and Agricultural Economics*. 46(1978):294-314.

Garcia, P., S.H. Irwin, R.M. Leuthold and L. Yang. "The Value of Public Information in Commodity Futures Markets." *Journal of Economic Behavior and Organization*. 32(1997):559-570.

Goodwin, B.K. "Forecasting Cattle Prices in the Presence of Structural Change." *Southern Journal of Agricultural Economics*. (1992):1122.

Granger, C.W.J. "Can We Improve the Perceived Quality of Economic Forecasts?" *Journal of Applied Econometrics*. 11(1996):455-473.

Granger, C.W.J. and P. Newbold. *Forecasting Economic Time Series*. Second Edition. New York: Academic Press, 1986.

Green, R.M. "Batting Averages in Agricultural Forecasting." *Journal of Farm Economics*. 8(1926):174-193.

Hamilton, J.D. *Time Series Analysis*. New Jersey: Princeton University Press, 1994.

Harvey, D.I., S.J. Leybourne, and P. Newbold. "Tests for Forecast Encompassing." *Journal of Business and Economic Statistics*. 16(1998):254-259.
Harvey, D., S. Leybourne, and P. Newbold. "Testing the Equality of Prediction Mean Squared Errors." *International Journal of Forecasting*. 13(1997): 281-291.
Holden, K. and D.A. Peel. "On Testing for Unbiasedness and Efficiency of Forecasts." *The Manchester School* 15(1990): 120-127.
Hurt, C. "Hog Producers Gearing Up For Expansion." *Livestock Price Outlook*. Purdue University, West Lafayette, Indiana and University of Illinois, Urbana, Illinois. October, 2000.
Irwin, S.H., D.L. Good, and J.K. Gomez. "The Value of USDA Outlook Information: An Investigation Using Event Study Analysis." Presented Paper, NCR Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management, St. Louis, Missouri, 2001 (http://agebb.missouri.edu/ncrext/ncr134/)
Henriksson, R.D. and R.C. Merton. "On Market Timing and Investment Performance. II. Statistical Procedures for Evaluating Forecasting Skills." *Journal of Business*. 1981(54):513-533.
Kastens, T. L., Schroeder, T.C., and R. Plain. "Evaluation of Extension and USDA Price and Production Forecasts. *Journal of Agricultural and Resource Economics.* 23(1):244-261.
Leuthold, R.M. "On the Use of Theil's Inequality Coefficients." *American Journal of Agricultural Economics*. 57(1975):344-346.
McIntosh, C.S. and J.H. Dorfman. "Qualitative Forecast Evaluation: A Comparison of Two Performance Measures." *American Journal of Agricultural Economics*. 74(1992):209-214.
Mills, T.C. and G.T. Pepper. "Assessing the Forecasters: An Analysis of the Forecasting Records of the Treasury, the London Business School and the National Institute." *International Journal of Forecasting*. 15(1999):247-257.
Mintert, J. "KSU Livestock Market Update: Special Electronic Edition." Kansas State Research and Extension, Manhattan, Kansas, October 8, 2001 (http://www.agecon.ksu.edu/livestock).
National Agricultural Statistics Service. "About NASS." http://www.usda.gov/nass/nassinfo/importnt.htm. June 14, 2001.
Nordhaus, W.D. "Forecasting Efficiency: Concepts and Applications." *The Review of Economics and Statistics*. 69(1987):667-674.
Pesaran, M.H. and A.G. Timmermann. "A Generalization of the Non-Parametric Henriksson-Merton Test of Market Timing." *Economic Letters*. 44(1994): 1-7.
Pettee, E.W. "Long-Term Commodity Price Forecasting—1850 to 1930. I" *The Journal of Business*. 9(1936):95-113.
Pons, J. "The Accuracy of IMF and OECD Forecasts for G7 Countries." *Journal of Forecasting*. 19(2000):53-63.
Pons, J. "The Rationality of Price Forecasts: A Directional Analysis." *Applied Financial Economics.* 11(2001):287-290.
Sanders, D.R. and M.R. Manfredo. "USDA Production Forecasts for Pork, Beef, and Broilers: An Evaluation," *Journal of Agricultural and Resource Economics*. 27(2002): 114-127.
Schaefer, M.P. and R.J. Meyers. "Forecasting Accuracy, Rational Expectations, and Market Efficiency in the US Beef Cattle Industry." Selected Paper, American Agricultural

Economics Association Meetings, Nashville, Tennessee, 1999.
Smyth, D.J. "Effect of Public Price Forecasts on Market Price Variation: A Stochastic Cobweb Example." *American Journal of Agricultural Economics*. 55(1973):83-88.
Steckler, H.O. and M.H. Schnader. "Evaluating Predictions of Change: An Application to Inflation Forecasts." *Applied Financial Economics*. 1(1991):135-137.
Stein, J.L. "Speculative Price: Economic Welfare and the Idiot of Chance." *Review of Economics and Statistics*. 63(1981):223-232.
Tyson Foods, Inc. "Tyson Provides Guidance for First Quarter and FY2003." Springdale, Arkansas, September 4, 2002. (http://www.tysonfoodsinc.com/IR/newsInfo/viewNews.asp?article=1042).
United States Department of Agriculture. *World Agricultural Supply and Demand Estimates*. Office of the Chief Economist, Washington, D.C., (various issues, 1980-2002).

Table 1. Summary Statistics, 1982.3-2002.3

	Actual Prices		
	Cattle[a]	Hogs	Broilers
Mean	0.0008	-0.0147	0.0092
Standard Deviation	0.0831	0.2183	0.1230
	USDA Forecasts		
	Cattle	Hogs	Broilers
Mean	0.0046	-0.0161	-0.0151
Standard Deviation	0.0615	0.1856	0.1032
	AR(4) Forecasts		
	Cattle	Hogs	Broilers
Mean	0.0177*	0.0156	0.0323*
Standard Deviation	0.0607	0.1768	0.0999

[a]Note: The numbers in Table 1 are interpreted as percents. For instance, the mean annual change in hog prices over the sample interval was -1.47% with a standard deviation of 21.83%.
*Statistically different from zero at the 5% level (two-tailed t-test).

Table 2. Forecast Accuracy Measures, 1982.3-2002.3

	USDA Forecasts		
	Cattle	Hogs	Broilers
RMSE[a]	0.0504	0.0882	0.0602
MAE	0.0392	0.0710	0.0456
Theil's U	0.6100	0.4059	0.4915
	AR(4) Forecasts		
	Cattle	Hogs	Broilers
RMSE	0.0648	0.1328	0.0758
MAE	0.0512	0.1059	0.0628
Theil'sU	0.7841	0.6110	0.6188
RMSE MDM test[b]	-2.496	-3.864	-2.267
MAE MDM test	-2.620	-4.662	-2.896

[a]Note: RMSE is the root mean squared error and MAE is the mean absolute error.
[b]The t-tests from the modified Diebold-Mariano test for equality of prediction errors. The USDA forecasts' RMSE and MAE are statistically smaller than those of the time series forecasts at the 5% level.

Table 3. Forecast Bias Test, $e_t = \gamma + \mu_t$, 1982.3-2002.3

	USDA Forecasts		
	Cattle	Hogs	Broilers
Estimated γ	0.0038	-0.0014	-0.0242
(t-statistic)	(0.56)†	(-0.14)	(-3.31)†
p-value	0.5783	0.8855	0.0014
	AR(4) Forecasts		
	Cattle	Hogs	Broilers
Estimated γ	0.0169	0.0302	0.0231
(t-statistic)	(2.42)	(2.09)	(2.86)
p-value	0.0179	0.0396	0.0054

†Newey-West covariance estimator.

Table 4. Beta Efficiency Test, $e_t = \alpha_1 + \beta\ FP_t + \mu_t$, 1982.3-2002.3

	USDA Forecasts		
	Cattle	Hogs	Broilers
Estimated β	-0.0744	-0.0027	-0.0650
(t-statistic)	(-0.81)	(-1.45)	(-1.59)†
p-value	0.4213	0.1505	0.1147
	AR(4) Forecasts		
	Cattle	Hogs	Broilers
Estimated β	0.1002	0.0092	0.0066
(t-statistic)	(0.86)	(0.15)†	(0.08)
p-value	0.3906	0.8807	0.9358

†Newey-West covariance estimator.

Table 5. Rho Efficiency Test, $e_t = \alpha_2 + \rho e_{t-1} + \mu_t$, 1982.3-2002.3

	USDA Forecasts		
	Cattle	Hogs	Broilers
Estimated ρ	0.2465	0.1804	0.3161
(t-statistic)	(2.26)	(1.63)	(3.03)
	0.0264	0.1078	0.0034
	AR(4) Forecasts		
	Cattle	Hogs	Broilers
Estimated ρ	0.0228	-0.2108	0.1745
(t-statistic)	(0.20)	(-1.91)	(1.60)
p-value	0.8388	0.0602	0.1146

Table 6. Forecast Encompassing Test, $e_{1t} = \alpha_3 + \lambda(e_{1t} - e_{2t}) + \varepsilon_t$, 1982.3-2002.3

	Cattle	USDA Encompass Time Series Hogs	Broilers
Estimated λ	0.1986	-0.0190	0.0973
(t-statistic)	(1.93)[†]	(-0.15)[a]	(1.12)[†]
p-value	0.0286	0.4425	0.1330
MDM Statistic[b]	1.660	-0.1460	0.9287
p-value	0.0650	0.4615	0.1779

[†]Newey-West covariance estimator.
[a]White's covariance estimator.
[b]The t-statistic from modified Diebold-Mariano test for forecast encompassing.

Table 7. Time Improvement Test, $|e_t| = \theta_1 + \theta_2 Trend_t + \mu_t$, 1982.3-2002.3

	Cattle	USDA Forecasts Hogs	Broilers
Estimated $\theta_2 \times 10^2$	-0.0085	0.0377	-0.0462
(t-statistic)	(-0.56)	(1.52)	(-2.53)[a]
p-value	0.5789	0.1328	0.0133

	Cattle	AR(4) Forecasts Hogs	Broilers
Estimated $\theta_2 \times 10^2$	-0.0284	0.0933	-0.0603
(t-statistic)	(-1.51)	(2.28)[a]	(-3.13)
p-value	0.1344	0.0254	0.0025

[a]White's covariance estimator.

Table 8. Contingency Table to Forecast Market Direction

	Actual		
Forecast	$\Delta A>0$	$\Delta A\leq 0$	Subtotal
$\Delta F>0$	n_{11}	n_{12}	FN_1
$\Delta F\leq 0$	n_{21}	n_{22}	FN_2
Subtotal	N_1	N_2	N

Note: ΔF is the forecasted direction of change, ΔA is the actual direction of change, and n is the number of observations in each cell of the table.

Table 9. Directional Forecasting Ability versus Prior Quarter

		USDA Forecasts	
	Cattle	Hogs	Broilers
% Correct	76.5	71.6	72.8
X^2 statistic	22.82	15.11	20.23
p-value	0.0000	0.0001	0.0000
		AR(4) Forecasts	
	Cattle	Hogs	Broilers
% Correct	65.4	65.4	75.3
X^2 statistic	8.49	7.65	20.18
p-value	0.0036	0.0057	0.0000
z-score[a]	1.37	0.83	0.07

[a]The z-score testing the null hypothesis of equal directional forecasting ability between the USDA and the time series model.

Table 10. Directional Forecasting Ability versus Prior Year

	USDA Forecasts		
	Cattle	Hogs	Broilers
% Correct	85.2	87.7	80.2
X^2 statistic	40.10	46.06	31.17
p-value	0.0000	0.0000	0.0000
	AR(4) Forecasts		
	Cattle	Hogs	Broilers
% Correct	80.2	77.8	75.3
X^2 statistic	30.52	24.95	25.10
p-value	0.0000	0.0000	0.0000
z-score[a]	0.67	1.26	0.50

[a]The z-score testing the null hypothesis of equal directional forecasting ability between the USDA and the time series model.

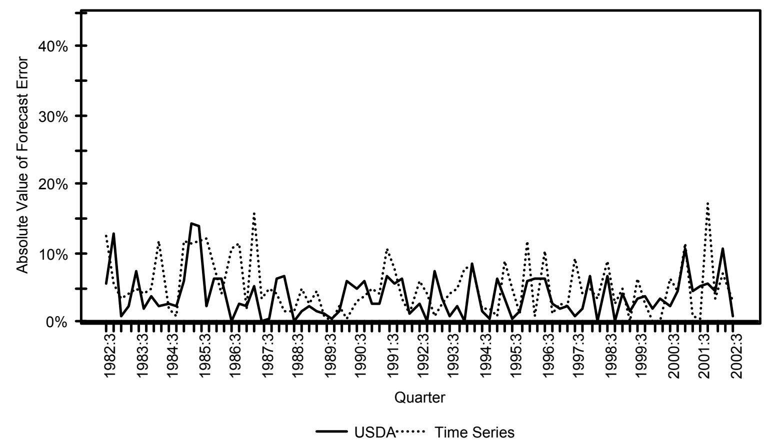
Table 11. Accuracy of Forecast Ranges

	USDA Forecasts		
	Cattle	Hogs	Broilers
% Correct	40.7	34.6	48.1
	Naïve Forecasts		
	Cattle	Hogs	Broilers
% Correct	29.6	34.6	40.7
z-score[a]	1.48	0.00	0.95

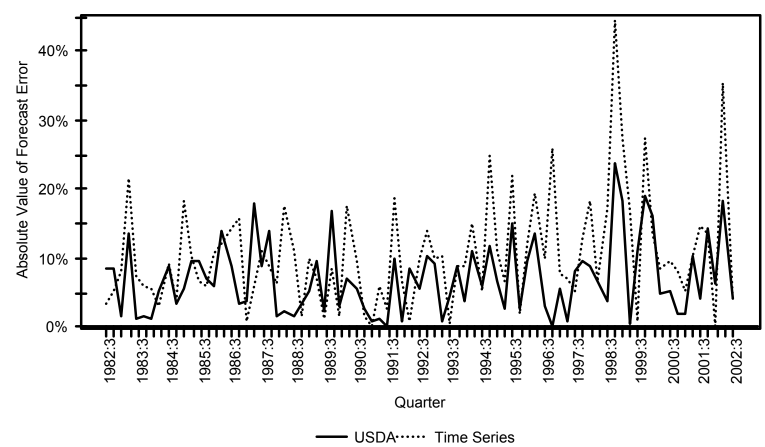
[a]The z-score testing the null hypothesis of equal sample proportions between the USDA and the time series model.

Figure 1. Absolute Value of Forecast Errors, $|e_t|$, 1982.3-2002.3.

Panel A: Cattle



Panel B: Hogs



Panel C: Broilers

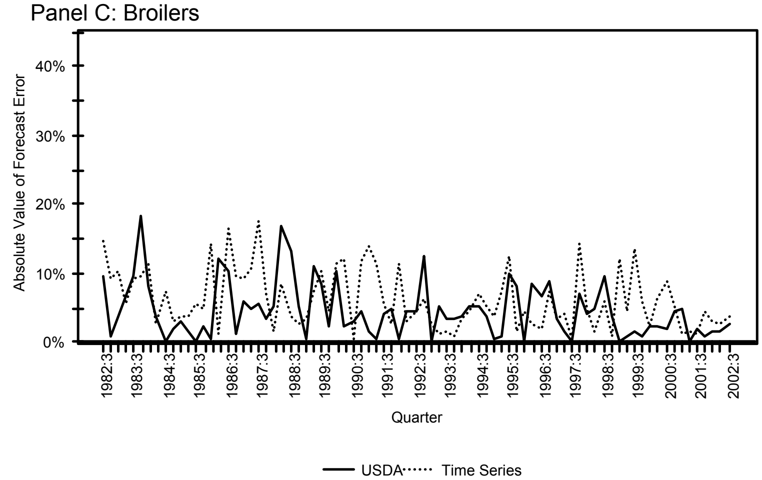
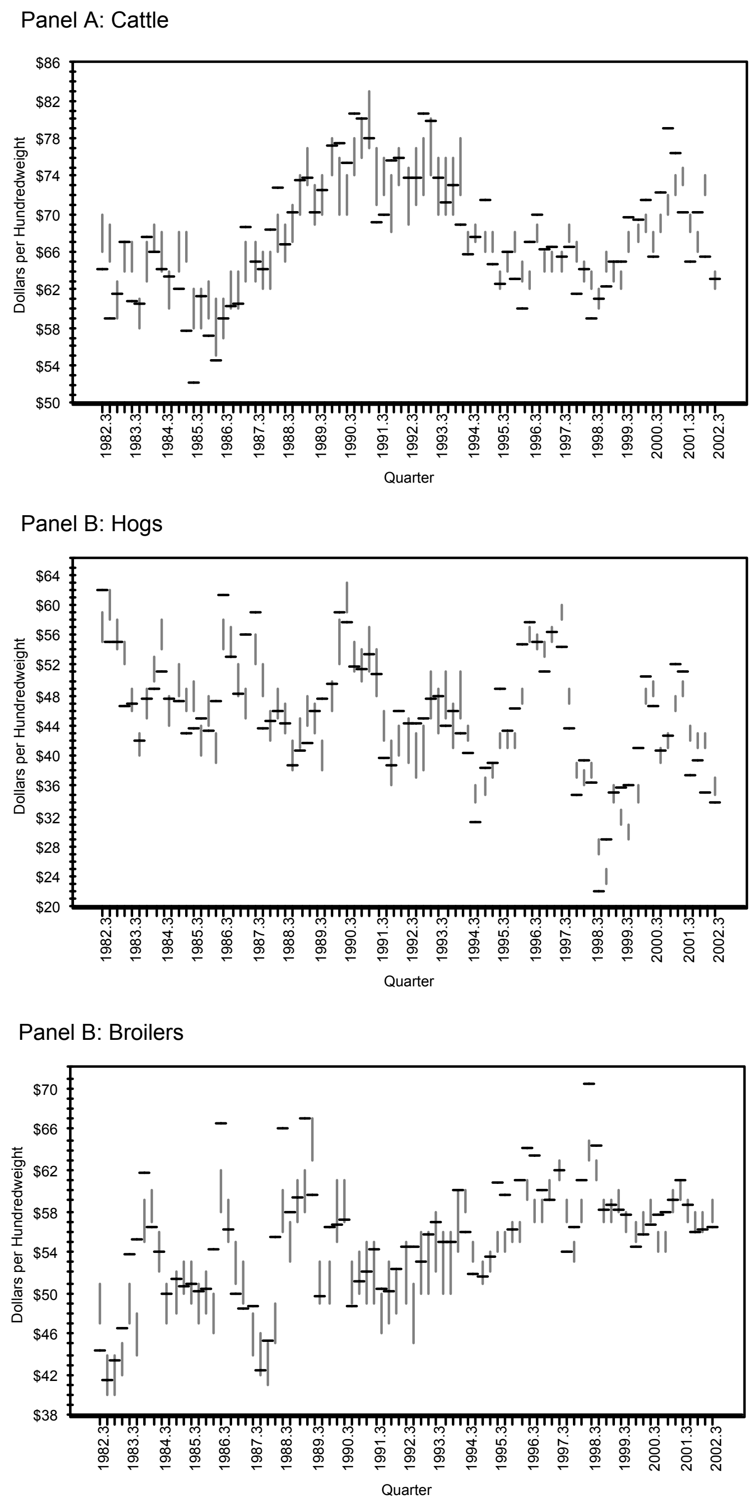


Figure 2. USDA Forecast Ranges and Realized Prices, 1982.3-2002.3.



Note, the gray vertical bars are the USDA's forecasted price range. The black horizontal dash is the realized price.