

Title: Empirical Confidence Intervals for USDA Commodity Price Forecasts

Running Title: Empirical Confidence Intervals

Authors:

Olga Isengildina-Massa (Corresponding Author)

Assistant Professor, Department of Applied Economics & Statistics Clemson University

295 Barre Hall

Box 340313

Clemson, SC 29634-0313

Phone: 864-656-2440

Fax: 864-656-5776

E-mail: olga123@clemson.edu

Scott Irwin

Laurence J. Norton Chair of Agricultural Marketing, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign

344 Mumford Hall

1301 W. Gregory Dr.

Urbana, IL 61801

Darrel L. Good

Professor, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign

344 Mumford Hall

1301 W. Gregory Dr.

Urbana, IL 61801

Luca Massa

Assistant Professor, Mechanical and Aerospace Engineering,

University of Texas at Arlington,

500 W. First St., WH 214 C

Arlington, 76019-0018, TX.

Empirical Confidence Intervals for USDA Commodity Price Forecasts

Abstract

Conventional procedures for calculating confidence limits of forecasts generated by statistical models provide little guidance for forecasts based on a combination or a consensus process rather than formal models, as is the case with USDA forecasts. This study applied and compared several procedures for calculating empirical confidence intervals for USDA forecasts of corn, soybean, and wheat prices over the 1980/81 through 2006/07 marketing years. Alternative procedures were compared based on out-of-sample performance over 1995/96 through 2006/07. The results of this study demonstrate that kernel density, quantile distribution and best fitting parametric distribution (logistic) methods provided confidence intervals calibrated at the 80% level prior to harvest and 90% level after harvest. The kernel density-based method appears most accurate both before and after harvest with the final value falling inside the forecast interval 77% of the time before harvest and 92% after harvest, followed by quantile regression (73% and 91% before and after harvest, respectively), logistic distribution (73% and 90% before and after harvest, respectively), and histogram (66% and 84% before and after harvest, respectively). Overall, this study demonstrates that empirical approaches may be used to construct more accurate confidence intervals for USDA corn, soybean, and wheat price forecasts.

Keywords: Agriculture, Commodity, Evaluating forecasts, Fixed-event forecasts, Government forecasting, Judgemental forecasting, Prediction intervals, Price forecasting

Empirical Confidence Intervals for USDA Commodity Price Forecasts

1. Introduction

Commodity price volatility, as recent years have amply demonstrated, is a fact of life. Managing this uncertainty is an important issue for commodity producers, processors, food manufacturers and numerous other firms with dealings in commodity markets (e.g., Tufano, 1996). For agricultural commodities, analysis and forecasts provided by the U.S. Department of Agriculture (USDA) typically serve as market benchmarks in formulating risk management strategies. USDA price forecasts are generated using a balance sheet approach, with published intervals reflecting uncertainty associated with prices in the future (Vogel and Bange, 1999). For example, the October 2008 USDA forecast of the 2008/09 marketing year average farm price was \$4.20-\$5.20/bushel for corn, \$9.60-\$11.10/bushel for soybeans and \$6.60-\$7.40/bushel for wheat. Note that the confidence level associated with published intervals is not revealed.

Isengildina, Irwin, and Good (2004) showed that USDA interval forecasts of corn and soybean prices during the 1980/81 through 2001/02 marketing years had relatively low hit rates, ranging from 36% to 82% for corn and from 59% to 89% for soybeans depending on the forecast month. In addition, actual prices were more likely to be above the forecast intervals, suggesting that observed symmetric USDA forecast intervals did not reflect the true asymmetry in the distribution of underlying prices. The authors argued that specific confidence levels should accompany forecast intervals in order to minimize confusion and misunderstanding in forecast interpretation.

While numerous procedures have been proposed to calculate confidence limits of forecasts generated by statistical models, these procedures provide little guidance for forecasts based on a combination or a consensus process rather than formal models, as is the case with USDA forecasts. In reviewing the prediction interval literature, Chatfield (1993) observes that when theoretical formulae are not available or there are doubts about model assumptions the use of empirically-based methods should be considered as a general purpose alternative. Empirical methods are based on the notion that forecast confidence limits may be estimated by evaluating historical forecast errors. An empirical method was first applied to construction of forecast confidence limits by Williams and Goodman (1971). The method has been used successfully in a variety of fields since the initial work of Williams and Goodman (e.g., Murphy and Winkler, 1977; Stoto, 1983; Keilman, 1990; Zarnowitz, 1992; Shlyakhter et al, 1994; Jorgensen and Sjoberg, 2003).

The purpose of this study is to investigate alternative approaches to estimating empirical forecast error distributions and associated confidence limits for USDA forecasts of marketing year average corn, soybean, and wheat prices. USDA interval price forecasts over the 1980/81 to 2006/07 marketing years are included in the analysis. Empirical error distributions are estimated using non-parametric approaches, parametric distributions, and quantile regressions. The procedures are compared based on out-of-sample performance, where the first 15 observations (1980/81-1994/95) are used to generate confidence limits for the 16th year (1995/96); the first 16 observations are used to generate confidence limits for the 17th year (1996/97) and so on. The accuracy of the out-of-sample empirical confidence intervals is compared to each other

and to published intervals in terms of hit rates. Statistical significance of the differences in hit rates from a target confidence level is assessed using an unconditional coverage test developed by Christoffersen (1998).

This study extends previous literature by providing a side-by-side comparison of the out-of-sample performance of several non-parametric and parametric methods for estimating empirical forecast confidence intervals. While empirical methods are gaining popularity there is no consensus to date on the best way to estimate forecast error distributions. The results of this study will provide new evidence regarding the most appropriate methods of calculating empirical confidence intervals. Furthermore, the results of this research will provide valuable information that can be used to more accurately estimate confidence limits for USDA price interval forecasts.

2. Empirical Methods

The empirical approach to estimation of forecast confidence intervals is based on the notion that confidence limits for future forecasts can be based on the distribution of historical forecast errors. The empirical approach was first applied to construction of confidence limits for economic forecasts by Williams and Goodman (1971). Their approach consists of splitting the data in two parts and fitting the method or model to the first part in order to find forecast errors. The model is then refitted each year adding an additional observation in the first part and increasing the part of the sample used to estimate forecast errors. The key assumption of this method is that future forecast errors belong to approximately the same distribution as past forecast errors.¹ Williams and Goodman (1971) argue that this assumption is less restrictive than the standard

assumption that a forecasting model describes the series adequately in the future. Therefore, by accumulating forecast errors through time one can obtain an empirical distribution of forecast errors and determine confidence limits for future forecasts by using the percentage points of the empirical distribution generated from past errors. The benefit of this method is that it can be applied in a straightforward manner to any type of error distribution, including fat-tailed and/or asymmetric distributions. This method is also particularly useful when the forecasting model is not fully identified and based on consensus or a combination of several sources, which makes the standard model-based methods for confidence limit calculation impossible to apply.

Empirical methods have also been recommended for cases when theoretical formulae yield prediction intervals that are too narrow. For example, Williams and Goodman (1971) found that between 20 and 26% of 18-month-ahead forecasts fell outside the 90% prediction intervals. Makridakis et al. (1987) found that 13 of the 15 univariate forecasting methods had more than 30% of the 4-year-ahead forecasts outside the 95% confidence interval. Chatfield (1993) argues that one of the reasons for narrow theoretical prediction intervals may be that innovations are not normally distributed. This shortcoming may be easily corrected with an empirical method which does not rely on the assumption of normality.

Empirical methods of constructing confidence intervals have been used successfully in a variety of fields (e.g., Murphy and Winkler, 1977; Stoto, 1983; Keilman, 1990; Zarnowitz, 1992; Shlyakhter et al, 1994; Jorgensen and Sjoberg, 2003). One of the main limitations of empirical methods is the data requirement. Therefore, these methods have been most widely-used in areas where data limitations are less

common, such as weather, population, and software development forecasting.

Recently, the importance of empirical control of a model's probability assessments has been recognized in engineering applications (e.g., Rebba and Mahadevan, 2006).

While empirical methods are gaining popularity there is no consensus on how the forecast error distributions should be estimated. Gardner (1988) proposed the use of Chebychev's inequality to derive prediction intervals as it provides an upper bound for any distribution with known mean and standard deviation. Based on the 1,001 series of the M-competition, Gardner found that Chebychev's inequality produced more accurate confidence intervals than when normally distributed errors are assumed. However, Bowerman and Koehler (1989) objected to the use of Chebychev's inequality because it gives unnecessarily wide prediction intervals for a correctly specified and unchanging model. Makridakis and Winkler (1989) provide evidence that actual forecast errors based on this approach tend to be larger, on average, than in-sample fit errors. With data on daily electricity peak loads, Allen and Morzuch (1995) demonstrated that probability forecasts calculated from Chebychev's inequality were the worst-calibrated. Their study also included four parametric distributions, normal, gamma, Cauchy and Weibull. They found that the Weibull distribution produced better calibrated post-sample forecasts for their data set. Williams and Goodman (1971) found that for their dataset on the number of telephone lines in use in designated cities a gamma distribution provided the best fit. Several other studies (e.g., Stoto, 1983; Cohen, 1986) found that forecast error distributions of population growth forecasts approximately followed a normal distribution. Shlyakhter et al. (1994) proposed the use of exponential distributions for data with large deviations from normality, such as population and

energy data. Jorgensen and Sjoberg (2003) used a non-parametric histogram approach to find desired points of the software development effort distribution.

Taylor and Bunn (1999a, 1999b) suggested an approach to empirical confidence interval estimation that addresses the small sample problem by pooling data across time and forecasting horizons and estimating forecast error distributions via quantile regression. The authors develop forecast error quantile models that are functions of lead time, k , as suggested by theoretically-derived variance expressions. The use of quantile regression avoids the normality and optimality assumptions underlying theoretical forecast variance expressions. Another benefit of this approach is that it relaxes the assumption that error distributions for each forecast month are independent. This assumption is likely to be violated in the case of USDA price forecasts, since forecast errors tend to decline from the beginning to the end of the forecasting cycle as more information becomes available.

Empirical methods described in this section use historical forecast errors for construction of confidence intervals based on the assumption that future forecast errors will follow approximately the same distribution as past errors. These methods differ in the way historical error distributions are estimated. Our study will provide insight regarding the relative accuracy of the alternative methods by providing a side-by-side comparison of their out-of-sample performance.

3. Data

USDA corn, soybean, and wheat interval price forecasts are released monthly, usually between the 9th and 12th of the month within WASDE (World Agricultural Supply and Demand Estimates) reports. USDA price forecasts are published in the

form of an interval and generated using a balance sheet approach, with published intervals reflecting uncertainty associated with prices in the future (see Vogel and Bange (1999) for a detailed description of the forecast generation process). The price in these forecasts refers to the average marketing year farm price, which is an average of the price received by farmers weighted by the monthly amounts marketed. The first price forecast for a marketing year is released in May preceding the U.S. marketing year (September through August for corn and soybeans and June through May for wheat). Estimates are typically finalized by August (for wheat), and November (for corn and soybeans) of the following marketing year (figure 1). Note that USDA price forecasts are considered fixed-event forecasts because the series of forecasts is related to the same terminal event (q_{FM}^i), where FM is the release month (November for corn and soybeans and September for wheat) of the final estimate of average farm price for the i^{th} marketing year. The forecast of the terminal event for month fm is denoted as q_{fm}^i , where $fm=1, \dots, FM$, and $i=1980/81, \dots, 2006/07$. Thus, each subsequent forecast is essentially an update of the previous forecast as it describes the same terminal event. The WASDE forecasting cycle generates 18 updates ($FM=19$) of marketing year average price for both corn and soybeans and 15 updates ($FM=16$) for wheat. While the forecasts are published in the form of an interval, the probability with which the realized price is expected to fall within the forecast interval is not specified.

[INSERT FIGURE 1 HERE]

Descriptive statistics for USDA interval price forecasts for corn, soybeans, and wheat over 1980/81 through 2006/07 are presented in Table 1. Prior to harvest (the first six forecasts from May to October for corn and soybeans and the first three forecasts

from May to July for wheat), soybean price interval forecasts were the widest, averaging about 19% of the forecast price compared to 17% for corn and 14% for wheat.² Soybean interval forecasts prior to harvest were also the most accurate with an average hit rate (proportion of times forecast intervals contain the final or “true” value) of 67% compared to 46% for both corn and wheat. After harvest (the last ten forecasts from November to August for corn and soybeans and from August to May for wheat),³ corn interval forecasts were the widest and the most accurate, averaging about 7% of the forecasted price (relative to 6% and 5% for wheat and soybeans, respectively) with an average hit rate of 71% (relative to 65% for both soybeans and wheat). Poor forecast accuracy after harvest is associated with the fact that forecast intervals were usually collapsed to point estimates in May after harvest for wheat and soybeans and in August after harvest for corn. No trends in the magnitude of interval forecasts over time were detected. Thus, intervals in the same months did not become smaller (or larger) from the beginning to the end of the sample period.

[INSERT TABLE 1 HERE]

Another issue is whether forecast intervals accurately reflect the shape of the underlying price distribution. Statistics on the proportion of misses above and below the forecast interval reported in Table 1 provide insight on this issue. If the forecast intervals accurately reflected the shape of the underlying price distribution, one would expect equal probability of misses above and below the forecast interval, which is about what we observe for corn price forecasts after harvest. Table 1 demonstrates that for soybean price forecast intervals the proportion of misses above the interval was 2.2 times greater than the proportion of misses below the interval prior to harvest and 2.9

times greater after harvest. The other two commodities did not exhibit such persistent tendencies.

An important basic assumption of empirical approaches to estimation of confidence limits is that the distribution of forecast errors remains stable over time. Forecast errors are calculated in this study as percentage errors, which allows us to compare results across commodities and may be preferred to unit errors when mean price levels change (as they did for most agricultural commodities after 2006). The validity of this assumption for corn, soybean and wheat price forecast errors is tested by dividing the sample in two parts, from 1980/81 through 1994/95 and from 1995/96 through 2006/07 and examining whether the first two moments of forecast error distributions differed between two sub-periods. An independent sample t-test was used to test the difference in the mean forecast errors across the two sub-samples and Levene's F-statistic was used to test the difference in the variance of forecast errors.

Results of this analysis are presented in Table 2 and demonstrate that the mean forecast error for soybean price forecasts prior to harvest was significantly higher in the more recent sub-period. These results also suggest that the variance of corn price forecast errors after harvest was significantly smaller and the variance of wheat price forecast errors prior to harvest was significantly larger in the 1995/96 through 2006/07 sub-period. Forecast errors were not statistically different across the two sub-periods in all other cases, indicating that error distributions remained relatively stable over time. These results suggest that empirical approach to calculating confidence intervals would be most efficient for corn price forecasts prior to harvest and soybean and wheat price forecasts after harvest.

[INSERT TABLE 2 HERE]

This section described several major concerns regarding WASDE interval forecasts of corn, soybeans, and wheat prices: 1) these intervals are characterized by relatively low hit rates; 2) in soybeans, symmetric forecast intervals do not accurately reflect the shape of the underlying price distribution; and 3) confidence levels associated with these forecast intervals are not specified. Since error distributions of these forecasts remained relatively stable over time, confidence limits calculated using empirical methods described in this study may provide a powerful tool for improving WASDE price forecast intervals.

4. Forecast Error Distribution Estimation

Empirical methods of calculating confidence intervals are based on splitting the data in two parts and fitting the method or model (e.g., Williams and Goodman, 1971, and Gardner, 1988) to the first part in order to find the forecast error distribution, which is then used to provide confidence limits for the forecast intervals in the second part of the sample. This study uses actual USDA forecast errors from the first part of the sample to estimate forecast error distributions for the second part of the sample.

In order to determine a split point between the first and the second part of the sample, data availability issues should be taken into account. In this study, 27 annual observations are available for each monthly forecast for each commodity. Dividing the sample in two parts, from 1980/81 through 1994/95 and from 1995/96 through 2006/07 provides us with a minimum of 15 years of data to estimate forecast error distributions and 12 years out-of-sample to evaluate the performance of the proposed approaches. One drawback of an empirical method is the requirement of enough historical data in

order to estimate forecast error distributions. Taylor and Bunn (1999b) argue that, "...the number of forecast errors needed depends on the accuracy required, but it is probably fair to say that empirical methods need at least fifty observations to produce reasonable results (p. 328)." It has been shown that variance estimates using Gardner's approach are unreliable for small sample sizes and are based on model-fitting errors rather than true post-sample forecast errors (e.g., Bowerman and Koehler, 1989). At the same time Williams and Goodman (1971) demonstrate that empirical confidence intervals using 24 observations were satisfactory at the 80% level, but 90% and 95% levels improved when more observations were included. These arguments suggest that some confidence intervals calculated within this study may not be very precise due to small sample sizes, but the use of the larger sample for calculation of the most recent forecast intervals should provide reasonable estimates.

Another issue is selection of the confidence level for interval forecasts. As was mentioned, confidence levels associated with published USDA interval price forecasts are not specified. Isengildina, Irwin, and Good (2004) conducted a survey of USDA officials involved in compiling USDA corn and soybean price interval forecasts inquiring about the confidence levels associated with published forecasts. Analyst responses were variable across respondents (by as much as 30% in the beginning of the season) and over the forecasting cycle (from 65% in May prior to harvest to 95% in April after harvest). The average confidence level prior to harvest was 81% for corn and 78% for soybeans; the average confidence level after harvest was 91% for corn and 87% for soybeans. Based on this information, and assuming that wheat analysts

provide interval forecasts for similar confidence levels, this study calculates confidence intervals for 80% level prior to harvest and 90% level after harvest.

Finally, the most important question is how the historical error distributions should be estimated. This study explores several non-parametric and parametric procedures for estimating forecast error distributions. The analysis consists of estimating error distributions using the first part of the sample, which starts with 15 observations (1980/81 through 1994/95) and adds one observation for each additional year.⁴ Upper and lower interval bounds are calculated based on the percentiles of the estimated error distributions. These interval bounds are added to the midpoint of the first observation in the second (prediction) part of the sample (e.g., 1995/96) to construct the empirical confidence interval. Thus, the empirical confidence intervals for the 2006/07 estimates are based on 26 observations (1980/81 through 2005/06). The procedures for estimating forecast error distributions are introduced and discussed in the remainder of this section.

4.1. Histograms

The simplest procedure for estimation of non-parametric forecast error distributions is via histograms. The histograms are constructed by sorting forecast errors from negative to positive. Figure 2 gives an example of such histograms for November soybean price forecasts based on the first 15 and the first 26 observations. The confidence intervals based on the histograms can be approximated by dropping extreme observations to achieve the desired confidence level.⁵ For the sample ranging from 15 to 26 observations, the confidence intervals constructed by dropping the largest positive and the largest negative error correspond to confidence levels from 87% (13 out of 15

observations) to 92% (24 out of 26 observations). Thus, the empirical 90% level confidence intervals for the post-harvest months can be constructed by adding the second largest positive and second largest negative error to the midpoint of the interval forecast. Similarly, confidence intervals constructed by dropping the two largest positive and two largest negative errors correspond to confidence levels from 73% (11 out of 15 observations) to 85% (24 out of 26 observations). These confidence intervals are used to approximate an 80% confidence level for the pre-harvest months. While this non-parametric procedure is straightforward, several drawbacks should be pointed out: 1) the discrete nature of the histograms limits the flexibility of selecting a specific confidence level; 2) this procedure concentrates on the tails of the distribution (its extreme values in this case), without taking into account the complete shape of the probability distribution and 3) the histogram may be sensitive to the choice of origin and bin width⁶.

[INSERT FIGURE 2 HERE]

4.2. Kernel Density Estimation

An alternative non-parametric approach to estimation of empirical error distributions is the use of kernel density estimators. The kernel density estimator replaces the “boxes” in a histogram by “bumps” that are smooth (Silverman, 1986). Smoothing is accomplished by putting less weight on observations that are further from the point being evaluated. Specifically, the kernel density estimate of series X at a point x is estimated by,

$$(1) \quad f(x) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{x - X_i}{h}\right)$$

where N is the number of observations, h is the bandwidth (or smoothing parameter) and $K(\cdot)$ is a kernel function that integrates to one. Silverman (1986, p. 43) shows that the optimal kernel function is given by,

$$(2) \quad K_e(t) = \begin{cases} \frac{3}{4\sqrt{5}} \left(1 - \frac{1}{5}t^2\right) & -\sqrt{5} \leq t \leq \sqrt{5}, \\ = 0 & \text{otherwise,} \end{cases}$$

often referred to as the Epanechnikov kernel. Optimal bandwidth is given by (Silverman, 1986, pp. 47-48),

$$(3) \quad h_{opt} = 0.9AN^{-1/5}$$

where A is an adaptive estimate of spread given by,

$$(4) \quad A = \min(\text{standard deviation, interquantile range}/1.34).$$

Figure 3 provides an example of kernel density estimates for November soybean price forecasts based on the first 15 and the first 26 observations. Once the kernel density is estimated, the appropriate upper and lower bounds for the 80% and 90% confidence levels are found by integrating the density function using the Lobatto adaptive quadrature method (Ueberhuber, 1997). The empirical forecast intervals are constructed by adding these upper and lower bounds to the midpoint of the published interval forecasts.

[INSERT FIGURE 3 HERE]

While this empirical approach is not as straight forward as the approach based on a histogram, it offers the advantages of continuous non-parametric estimation of error distributions. Limitations of this approach include the changing shape of a kernel density estimate as observations are added or dropped. Additionally, the choice of

bandwidth (h) may impact the results of smoothing. Silverman's h_{opt} (equation 3) has been used in numerous studies as a rule-of-thumb, but may not fit all types of data equally well. The issue of bandwidth selection is still a point of debate in statistics literature. Kernel density estimates were computed using the MATLAB library of statistical functions.

4.3. Parametric Distribution Estimation

Some of the limitations of the non-parametric approaches can be addressed by fitting parametric probability distributions to forecast errors from the first part of the sample. This procedure includes finding a parametric distribution that best fits the empirical data. Five parametric distributions were used as likely alternatives: normal, logistic, extreme value, uniform, and Rayleigh distributions.⁷ Goodness-of-fit of the parametric distributions to forecast errors was evaluated using the Anderson-Darling statistic,

$$(5) \quad A_N^2 = N \int_{-\infty}^{+\infty} \left[F_N(x) - \hat{F}(x) \right]^2 \psi(x) \hat{f}(x) dx$$

where $\Psi^2 = \frac{1}{\hat{F}(x)[1 - \hat{F}(x)]}$, $\hat{f}(x)$ is the hypothesized density function, $\hat{F}(x)$ is the

hypothesized cumulative distribution function, $F_N(x) = \frac{N_x}{N}$, and N_x is the number of

observations less than x . The Anderson-Darling statistic is preferred to the Chi-squared statistic because it does not require binning.⁸ Furthermore, the Anderson-Darling statistic is selected over the Kolmogorov-Smirnov statistic because it concentrates on the differences between the tails of the fitted distribution and input data rather than the middle of the distribution.

Figure 4 presents the best fitting probability distributions for November soybean price forecasts based on the first 15 and the first 26 observations. The logistic distribution was selected as best-fitting in both cases. The main drawback of this approach is that a different best-fitting distribution may be chosen from one year to the next as more observations are added. Allowing distributions to vary over time can be viewed as a violation of the underlying assumption of the empirical approach that future forecast errors belong to approximately the same distribution as past forecast errors. To avoid this shortcoming the probability distribution that is most often ranked as the best fitting distribution is selected. For example, Williams and Goodman (1971) report that forecast errors from their data set approximately followed a gamma distribution. It was found that the forecast errors from the data set used in the present study most often followed a logistic distribution. The logistic distribution was selected as best-fitting using an Anderson-Darling statistic in 72% of cases in corn and soybeans, and in 86% of cases in wheat.⁹ Therefore, this distribution was applied in all cases for calculation of the confidence interval. The upper and lower bounds for the 80% and 90% confidence levels were calculated by integrating the logistic density function using the Lobatto adaptive quadrature method. The empirical forecast intervals were constructed by adding these upper and lower bounds to the midpoint of the published forecasts. A benefit of this approach is that the shape and properties of the distribution are well-known and do not change from one year to the next as observations are added/dropped. A disadvantage of this approach is that logistic distribution is not always the best fitting parametric distribution. Finally, since the logistic distribution is not as flexible as kernel density estimator, it may not fit the data as well as the kernel density function.

Logistic distribution confidence intervals were computed using MATLAB computing library.

[INSERT FIGURE 4 HERE]

4.4. *Quantile Regression Estimation*

One of the shortcomings of the previous non-parametric and parametric procedures for generating empirical confidence intervals is that they are not based on large samples. Each method is based on the available annual observations for each forecast horizon, with samples ranging from 15 to 26 observations. Small sample problems may be addressed by pooling the data across months and marketing years and estimating forecast error distributions using quantile regressions. The quantile regression approach to the construction of prediction intervals was first applied by Taylor and Bunn (1999a, 1999b). The authors argue that this method is a hybrid of the empirical and theoretical approaches in that it uses the empirical fit errors and produces forecast error quantile models which are functions of lead time, k , as suggested by theoretically derived variance expressions. The use of quantile regression avoids the normality and optimality assumptions which underlie theoretical forecast variance expressions. Another benefit of this approach is that it relaxes the assumption that error distributions for each forecasting month are independent, since forecast errors tend to decline from the beginning to the end of the forecasting cycle as more information becomes available.

Assuming lead time k corresponds to the forecast error series for k -step ahead forecasts, Taylor and Bunn (1999a, 1999b) specify the following quantile regression for a given level of τ , a quantile for which the regression is estimated:

$$(6) \quad y_t(\tau) = \beta_0 + \beta_1 k + \beta_2 k^2 + \varepsilon_t .$$

In the present study, this approach is adapted to fixed-event forecasts and k is substituted for its reverse, fm , the forecast month from the beginning to the end of the forecast cycle (1 through 19 for corn and soybeans and 1 through 16 for wheat),

$$(7) \quad y_t(\tau) = \beta_0 + \beta_1 fm + \beta_2 fm^2 + \varepsilon_t .$$

Again following Taylor and Bunn, standard errors were estimated by bootstrap resampling in order to correct for heteroscedasticity. Bootstrap resampling used the XY -pair method with 100 replications and samples the same size as the original data. This specification was used in the empirical analysis to generate upper ($\tau=0.90$) and lower ($\tau=0.10$) bounds of 80% confidence intervals and upper ($\tau=0.95$) and lower ($\tau=0.05$) bounds of 90% confidence intervals pre- and post- harvest, respectively, for each commodity. An example of the forecast error distribution generated using quantile regression based on 15 and 26 years of data is shown in Figure 5. The quantile regression approach offers the benefit of pooling data across months and years, and thus substantially increasing the statistical power of estimation. Quantile regression relaxes the assumption that forecast errors for separate months are uncorrelated. All quantile regressions were estimated using *Eviews* econometric software.

[INSERT FIGURE 5 HERE]

5. Results

Forecast error distributions estimated using histogram, kernel density, logistic distribution, and quantile regression procedures described in the previous section were used to calculate empirical confidence intervals for USDA corn, soybean, and wheat price forecasts over the 1995/96 through 2006/07 marketing years. These empirical

confidence intervals are computed using the first 15 observations (1980/81-1994/95) to generate confidence limits for the 16th year (1995/96); the first 16 observations to generate confidence limits for the 17th year (1996/97) and so on. Table 3 presents descriptive statistics of empirical confidence intervals as well as those of published USDA forecast intervals of corn, soybean, and wheat prices over 1995/96 through 2006/07 marketing years. Because of the generally low hit rates of the published forecasts shown in Table 1, it was expected that the empirical confidence intervals calibrated at the 80% confidence level prior to harvest and 90% confidence level after harvest would be wider than published intervals. Table 3 demonstrates that prior to harvest, empirical confidence intervals were 1.6 times (histogram) to 2 times (kernel density) wider on average than published forecast intervals of corn prices.¹⁰ After harvest, the difference was not as dramatic with histogram-based intervals about 25% wider, and kernel density-based intervals about 63% wider, on average, than published intervals. Logistic distribution and quantile regression-based procedures generated intervals that were wider than histogram-based intervals but narrower than kernel density-based intervals. A similar pattern is observed for both soybean and wheat price forecast intervals.

[INSERT TABLE 3 HERE]

Another issue with the USDA price forecasts is whether the symmetric forecast intervals accurately reflected the distributions of the underlying commodity prices. Because empirical confidence intervals were calculated using actual forecast errors, they take into account the asymmetries of error distributions. If the forecast intervals accurately reflected the shape of the underlying price distribution, one would expect equal probability of misses above and below the forecast interval. Table 4 demonstrates

that for corn, soybeans and wheat the price forecast intervals prior to harvest the proportion of misses above the interval was much greater than the proportion of misses below the interval. This is at least partially attributable to the fact that the mean of the forecast error distribution of soybean prices prior to harvest shifted to the right in the prediction sub-period as shown in Table 2.¹¹ Another violation of constant error distribution assumption highlighted in Table 2 was the increase in variance of the forecast errors of wheat prices prior to harvest in the prediction sub-sample. This resulted in low accuracy of empirical confidence intervals for wheat prior to harvest that was significantly different from the target level of 80% across all methods. These findings emphasize that empirical methods may be sensitive to changes in variance of the forecast error distribution.

The next step of the analysis is to assess and compare the performance of the empirical confidence intervals to the target confidence level. Accuracy of forecast intervals is traditionally examined in terms of hit rates and forecast coverage. Hit rates, which as noted earlier describe the proportion of times forecast intervals contain the final or “true” value (y_t), may be defined as an indicator variable, I_t^k ,

$$(8) \quad I_t^k = \begin{cases} 1, & \text{if } y_t \in [l_{t/k}(\alpha), u_{t/k}(\alpha)] \\ 0, & \text{if } y_t \notin [l_{t/k}(\alpha), u_{t/k}(\alpha)] \end{cases}$$

where $[l_{t/k}(\alpha), u_{t/k}(\alpha)]$ are the lower and upper limits of the interval forecast for y_t made at time k with confidence level α . The closer the hit rate to the stated confidence level, the more accurate is the forecast. Forecast coverage is based on the expectation of the indicator variable, I_t^k and examines whether the proportion of times the forecast interval includes the true value is equal to the target (stated) confidence level. Thus,

forecast coverage may be examined by testing the hypothesis $H_0: E(I_t^k) = \alpha$ against $H_1: E(I_t^k) \neq \alpha$. If H_0 is not rejected and the interval hit rate is equal to the stated confidence level, forecasts are said to be calibrated. Since the indicator variable I_t^k has a binomial distribution (Christoffersen, 1998), the likelihood function under the null hypothesis is,

$$(9) \quad L(\alpha) = (1 - \alpha)^{n_0} \alpha^{n_1}$$

where L is a likelihood function. Under the alternative hypothesis, the likelihood function is,

$$(10) \quad L(p) = (1 - p)^{n_0} p^{n_1}$$

where n_1 and n_0 are the number of times an interval was “hit” (1) or “missed” (0) in the indicator sequence I_t^k . Then, forecast coverage may be tested via the likelihood ratio test,

$$(11) \quad LR_c = -2 \ln \left(\frac{L(\alpha)}{L(\hat{p})} \right) \xrightarrow{asy} \chi^2(1)$$

where $\hat{p} = n_1 / (n_0 + n_1)$ is the maximum likelihood estimator of p . This test is described by Christoffersen (1998) as an unconditional coverage test.¹²

Accuracy test results presented in Table 4 reveal relative performance across different methods of error distribution estimation. When monthly hit rates are averaged pre- and post-harvest across all three commodities, the kernel density-based method appears most accurate both before and after harvest with an average hit rate of 77% before harvest and 92% after harvest, followed by quantile regression (73% and 91% before and after harvest, respectively) and logistic distribution (73% and 90% before and after harvest, respectively), and histogram (66% and 84% before and after harvest,

respectively). Coverage tests revealed that hit rates of published intervals over 1995/96 through 2006/07 marketing years were significantly different from 80% prior to harvest and 90% after harvest in all cases except after harvest soybean price forecasts. Among empirical confidence intervals, histogram-based forecasts performed the worst with hit rates statistically different from target levels in 4 out of 6 cases. The other three approaches performed relatively well with only minor deviations from the target level in all cases except wheat prior to harvest. It appears that logistic distribution did not capture the asymmetry in soybean price distribution after harvest as well as the other two methods. The accuracy of the kernel density and logistic distribution methods was significantly higher than the target level in wheat after harvest.

[INSERT TABLE 4 HERE]

6. Summary and Conclusions

This study applied and compared several procedures for calculating empirical confidence intervals of USDA forecasts of corn, soybean, and wheat prices over the 1980/81 through 2006/07 marketing years. The approach was first introduced by Williams and Goodman (1971), and is based on the notion that by accumulating forecast errors through time one can obtain an empirical distribution of forecast errors. If the key assumption of this method is satisfied and future forecast errors belong to approximately the same distribution as past forecast errors, confidence limits for future forecasts can be determined by using the percentage points of the empirical distribution generated from past errors. Based on the results of independent sample *t*-tests and Levene's F-tests, forecast error distributions of monthly USDA corn, soybean and wheat forecasts appeared stable most of the time (with the exception of higher mean in

soybean price forecast errors prior to harvest and greater variance of wheat forecast errors prior to harvest).

Four procedures were used to estimate forecast error distributions, including histograms, kernel density estimation, logistic distribution, and quantile regression. The procedures were compared based on out-of-sample performance, where the first 15 observations (1980/81-1994/95) were used to generate confidence limits for the 16th year (1995/96); the first 16 observations were used to generate confidence limits for the 17th year (1996/97) and so on. The results demonstrate that even using this fairly limited number of observations (15-26), empirical approaches provide reasonable estimates for confidence interval calculation at the 80% and 90% confidence levels. The larger variance of the forecast errors of wheat prices prior to harvest in the prediction sub-sample resulted in empirical confidence intervals that were significantly different from the target level of 80% across all methods. A significant shift in the mean forecast errors of soybean prices prior to harvest resulted in misses occurring more often on one side of the distribution rather than the other, but didn't deteriorate the overall accuracy of empirical confidence intervals. Thus, empirical methods may be sensitive to changes in variance of the forecast error distribution.

Based on the results of accuracy tests for empirical confidence intervals over 1995/96 through 2006/07, the kernel density-based method appears most accurate both before and after harvest with an average hit rate of 77% before harvest and 92% after harvest, followed by quantile regression (73% and 91% before and after harvest, respectively) and logistic distribution (73% and 90% before and after harvest, respectively), and histogram (66% and 84% before and after harvest, respectively). The

kernel density based approach generated the best fitting distribution for the empirical approach, as the shape of the kernel distribution is dictated entirely by data. This approach, however, may be unstable in small samples as additional observations are added or dropped. Quantile regression combines the benefits of the kernel density approach with the ability to pool the data across years and forecasting cycles thus greatly increasing the number of observations used for distribution estimation. The best fitting parametric distribution (logistic in this case) also provided reasonable results and is attractive due to the known properties of the distribution. The histogram approach, on the other hand, demonstrated poor performance and should be avoided in construction of empirical confidence intervals. While the results of the accuracy tests do not provide a clear indication about the preferred method, it appears that the benefits offered by quantile regression approach (flexibility of the distribution estimation and ability to pool data) lend it to be somewhat superior to the other methods. Further research on the details of quantile regression use for empirical confidence interval estimation is currently under way.

The findings of this study can be used to estimate confidence intervals for corn, soybean, and wheat price forecasts. Specifically, confidence limits for future forecasts can be calculated based on percentage points of the distributions of past forecast errors. The use of an empirical approach to confidence interval calculation provides accurate intervals that reflect the properties of the underlying price distributions. While this study reports results for the 80% and 90% confidence level, other confidence levels may be selected by USDA analysts and calculated in a similar manner.

Overall, this study demonstrates how various empirical approaches may be used to construct confidence intervals for USDA corn, soybean, and wheat price forecasts. The results of this study demonstrate that kernel density, quantile distribution and best fitting parametric distribution (logistic) methods provided calibrated confidence intervals. These findings may be extended to calculation of confidence intervals for other forecasts. Finally, this research concentrated on accuracy of the empirical confidence intervals without regard to interval width. Investigation of USDA forecast users' preferences for accuracy versus informativeness of forecast intervals is an interesting topic for future research.

Acknowledgements.

The funding support of the U.S. Department of Agriculture under cooperative Agreement 43-3AEK-5-80076 is gratefully acknowledged. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the view of the U.S. Department of Agriculture.

References

- Allen, P.G., and B.J. Morzuch. (1995). Comparing Probability Forecasts Derived from Theoretical Distributions. *International Journal of Forecasting*, 11, 147-157.
- Bowerman, B.L., and Koehler, A.B. (1989). The Appropriateness of Gardner's Simple Approach and Chebychev Prediction Intervals. Unpublished paper presented at the 9th International Symposium on Forecasting in Vancouver, British Columbia, June 18-20, 1989.
- Chatfield, C. (1993). Calculating Interval Forecasts. *Journal of Business and Economics Statistics*, 11, 121-135.
- Christoffersen, P.F. (1998). Evaluating Interval Forecasts. *International Economic Review*, 39, 841-62.
- Cohen, J.E. (1986). Population Forecasts and Confidence Intervals for Sweden: A Comparison of Model-Based and Empirical Approaches. *Demography*, 23, 105-126.
- Gardner, E. S., Jr. (1988). A Simple Method of Computing Prediction Intervals for Time-Series Forecasts. *Management Science*, 34, 541-546.
- Isengildina, O., S.H. Irwin, and D.L. Good. (2004). Evaluation of USDA Interval Forecasts of Corn and Soybean Prices. *American Journal of Agricultural Economics*, 84, 990-1004.
- Jorgensen, M. and D.I.K. Sjoberg. (2003). An Effort Prediction Interval Approach Based on the Empirical Distribution of Previous Estimation Accuracy. *Information and Software Technology*, 45, 123-136.

- Keilman, N.W. (1990). *Uncertainty in National Population Forecasting: Issues, Backgrounds, Analyses, Recommendations*. Amsterdam, Swets and Zeitlinger, Publications of the Netherlands Interdisciplinary Demographic Institute (NIDI) and the Population and Family Study Centre (CBGS), Volume 20.
- Makridakis, S., M. Hibon, E. Lusk, and M. Belhadjali. (1987). Confidence Intervals: An Empirical Investigation of the Series in the M-Competition. *International Journal of Forecasting*, 3, 489-508.
- Makridakis, S., and R.L. Winkler. (1989). Sampling Distributions of Post-Sample Forecasting Errors. *Applied Statistics*, 38, 331-342.
- Murphy, A.H., and R.L. Winkler. (1977). Reliability of Subjective Probability Forecasts of Precipitation and Temperature. *Journal of the Royal Statistical Society, Series C*, 26, 41-47.
- Rebba, R. and S. Mahadevan. (2006). "Model Predictive Capability Assessment Under Uncertainty." *American Institute of Aeronautics and Astronautics Journal*, 44 , 2376-2384.
- Shlyakhter, A.I., D.M. Kammen, C.L. Broido, and R. Wilson. (1994). Quantifying the Credibility of Energy Projections from Trends in Past Data. *Energy Policy*, 22, 119-130.
- Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. Bristol: J.W. Arrowsmith Ltd.
- Stoto, M.A. (1983). The Accuracy of Population Projections. *Journal of American Statistical Association*, 78, 13-20.

- Taylor, J.W., and D.W. Bunn. (1999a). A Quantile Regression Approach to Generating Prediction Intervals. *Management Science*, 45, 225-237.
- Taylor, J.W., and D.W. Bunn. (1999b). Investigating Improvements in the Accuracy of Prediction Intervals for Combinations of Forecasts: A Simulation Study. *International Journal of Forecasting*, 15, 325-339.
- Tufano, P. (1996). "Who Manages Risk? An Empirical Examination of Risk Management Practices in the Gold Mining Industry." *Journal of Finance*, 51, 1097-1137.
- Ueberhuber, C. W. (1997). *Numerical Computation 2: Methods, Software, and Analysis*. Berlin: Springer-Verlag, p. 105.
- Vogel, F.A., and G.A. Bange. (1999). *Understanding USDA Crop Forecasts*. Miscellaneous Publication No. 1554, US Department of Agriculture, National Agricultural Statistics Service and Office of the Chief Economist, World Agricultural Outlook Board.
- Williams, W.H., and Goodman, M.L. (1971). A Simple Method for the Construction of Empirical Confidence Limits for Economic Forecasts. *Journal of the American Statistical Association*, 66, 752-754.
- World Agricultural Supply and Demand Estimates. Selected issues from 1980 through 2007. Available at: <http://www.usda.gov/oce/commodity/wasde/index.htm>
- Zarnowitz, V. (1992). *Business Cycles: Theory, History, Indicators, and Forecasting*. The University of Chicago Press.

Table 1. Descriptive and Accuracy Statistics for Published USDA Interval Forecasts of Corn, Soybean, and Wheat Prices, 1980/81-2006/07 Marketing Years.

Statistic	Corn		Soybeans		Wheat	
	Prior to harvest	After harvest	Prior to harvest	After harvest	Prior to harvest	After harvest
Number of observations	162	270	162	270	81	270
			-----%-----			
Average Interval	17	7	19	5	14	6
Hit Rate	46	71	67	65	46	65
Misses Below	23	15	10	9	31	16
Misses Above	30	14	22	26	23	19

Note: Average interval is calculated by dividing interval range by its midpoint. Hit rate is the proportion of times the interval contained the final estimate. Misses above and below describe cases when the final estimate fell above or below the forecast interval. The sum of hit rate, misses below and misses above may not add up to 100 due to rounding

Table 2. Comparison of the First Two Moments of Percentage Error Distributions of USDA Interval Forecasts of Corn, Soybean, and Wheat Prices, 1980/81-1994/95, and 1995/96-2006/07 Marketing Years.

Statistic	Corn		Soybeans		Wheat	
	Prior to harvest	After harvest	Prior to harvest	After harvest	Prior to harvest	After harvest
Number of observations	162	270	162	270	81	270
1980/81-1994/95 Mean (%)	-0.40	-0.27	-1.53	0.20	0.16	0.36
1995/96-2005/06 Mean (%)	0.52	-0.02	5.20	0.85	0.48	-0.09
t-test	-0.45	-0.52	-3.82 ^{***}	-1.51	-0.13	1.15
1980/81-1994/95 Variance	1.63	0.21	1.35	0.14	0.84	0.12
1995/96-2005/06 Variance	1.73	0.11	1.10	0.13	1.74	0.11
Levene's F-Statistic	0.36	2.89 [*]	0.60	0.50	10.05 ^{***}	0.43

Note: Forecast error is calculated as the percentage difference between the final estimate and the midpoint of the forecast interval times 100. One asterisk indicates significance at 10% level, two asterisks indicate significance at 5% level, three asterisks indicate significance at 1% level.

Table 3. Descriptive Statistics for Empirical Confidence Intervals for USDA Corn, Soybean, and Wheat Price Forecasts, 1995/96-2006/07 Marketing Years.

Method/Statistic		Corn		Soybeans		Wheat		All	
		Prior to harvest	After harvest	Prior to harvest	After harvest	Prior to harvest	After harvest	Prior to harvest	After harvest
Number of observations		72	120	72	120	36	120	180	360
Published Intervals	Average Interval (%)	18	8	18	6	17	6	18	7
Histogram Intervals	Average Interval (%)	29	10	25	8	26	9	27	9
	Average Below (%)	-15	-5	-14	-4	-11	-4	-14	-4
	Average Above (%)	14	5	11	4	14	5	13	5
Kernel Density Intervals	Average Interval (%)	36	13	31	10	30	11	33	11
	Average Below (%)	-19	-6	-17	-5	-14	-5	-17	-5
	Average Above (%)	17	7	15	6	16	6	16	6
Logistic Distribution Intervals	Average Interval (%)	32	11	27	9	26	10	29	10
	Average Below (%)	-16	-6	-14	-4	-14	-5	-15	-5
	Average Above (%)	15	5	13	4	12	5	14	5
Quantile Regression Intervals	Average Interval (%)	33	12	30	10	28	10	31	11
	Average Below (%)	-18	-6	-16	-4	-13	-4	-16	-5
	Average Above (%)	15	6	14	6	16	6	15	6

Note: Average interval is calculated by dividing interval range by its midpoint. Average below refers to the average size of the interval below the midpoint, average above refers to the average size of the interval above midpoint.

Table 4. Accuracy Statistics for Empirical Confidence Intervals for USDA Corn, Soybean, and Wheat Price Forecasts, 1995/96-2006/07 Marketing Years.

Method/Statistic	Corn		Soybeans		Wheat						
	Prior to harvest	After harvest	Prior to harvest	After harvest	Prior to harvest	After harvest					
Number of observations	72	120	72	120	36	120					
Published Intervals	Average Interval (%)	18	8	18	6	17	6				
	Hit Rate (%)	51	79	65	92	44	71				
	Misses Below (%)	18	8	6	8	25	17				
	Misses Above (%)	31	13	29	0	31	12				
	Unconditional Coverage Test	29.42	***	13.12	***	8.47	***	0.39	22.06	***	33.04
Histogram Intervals	Average Interval (%)	29	10	25	8	26	9				
	Hit Rate (%)	71	88	69	75	50	89				
	Misses Below (%)	8	6	0	12	25	8				
	Misses Above (%)	21	6	31	13	25	3				
	Unconditional Coverage Test	3.43	*	0.57	4.50	**	22.16	***	16.07	***	0.05
Kernel Density Intervals	Average Interval (%)	36	13	31	10	30	11				
	Hit Rate (%)	81	93	79	86	64	96				
	Misses Below (%)	6	3	0	7	11	2				
	Misses Above (%)	14	4	21	8	25	2				
	Unconditional Coverage Test	0.01	1.50	0.03	2.30	5.02	**	6.59	**		
Logistic Distribution Intervals	Average Interval (%)	32	11	27	9	26	10				
	Hit Rate (%)	78	93	76	80	58	96				
	Misses Below (%)	7	2	0	5	14	2				
	Misses Above (%)	15	5	24	15	28	2				
	Unconditional Coverage Test	0.22	1.50	0.56	10.07	***	8.75	***	6.59	**	
Quantile Regression Intervals	Average Interval (%)	33	12	30	10	28	10				
	Hit Rate (%)	78	93	75	85	58	94				
	Misses Below (%)	6	5	0	8	17	5				
	Misses Above (%)	17	2	25	7	25	2				
	Unconditional Coverage Test	0.22	1.50	1.06	3.11	*	8.75	***	2.37		

Note: Empirical price forecast intervals were calculated using percentage errors from the 1980/81 marketing year forward. Target confidence level is 80% prior to harvest and 90% after harvest. One asterisk indicates significance at 10% level, two asterisks indicate significance at 5% level, three asterisks indicate significance at 1% level.

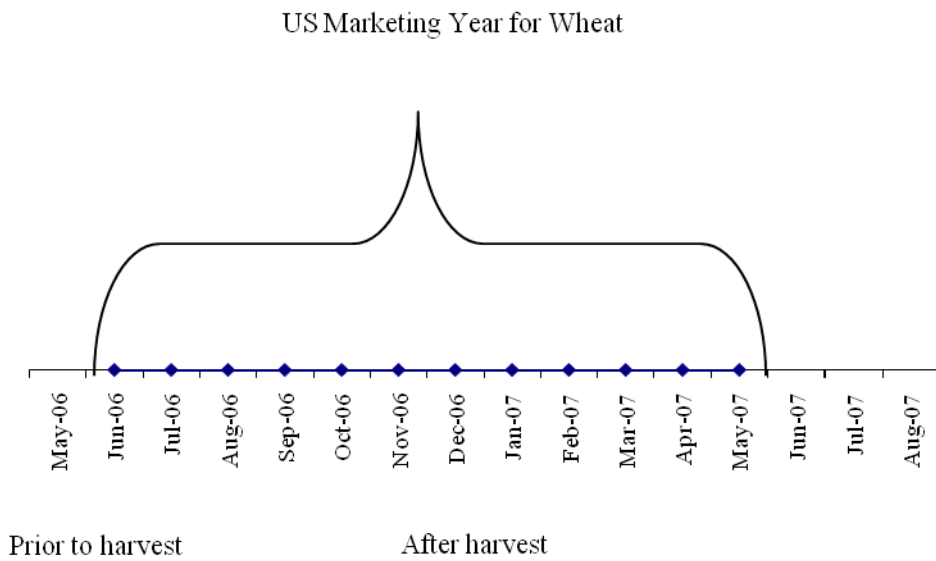
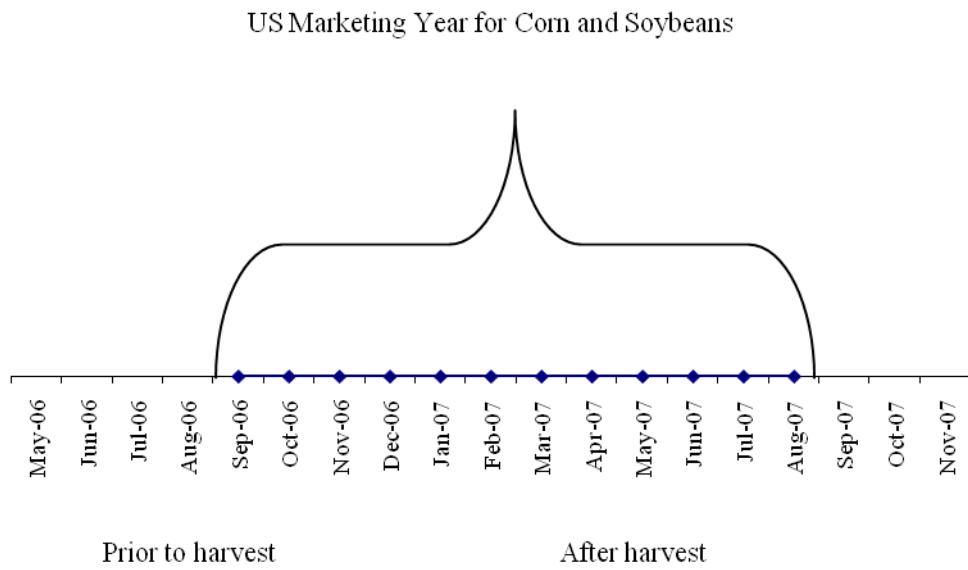


Figure 1. The 2006/07 WASDE Forecasting Cycle for Corn, Soybeans, and Wheat Relative to the U.S. Marketing Year.

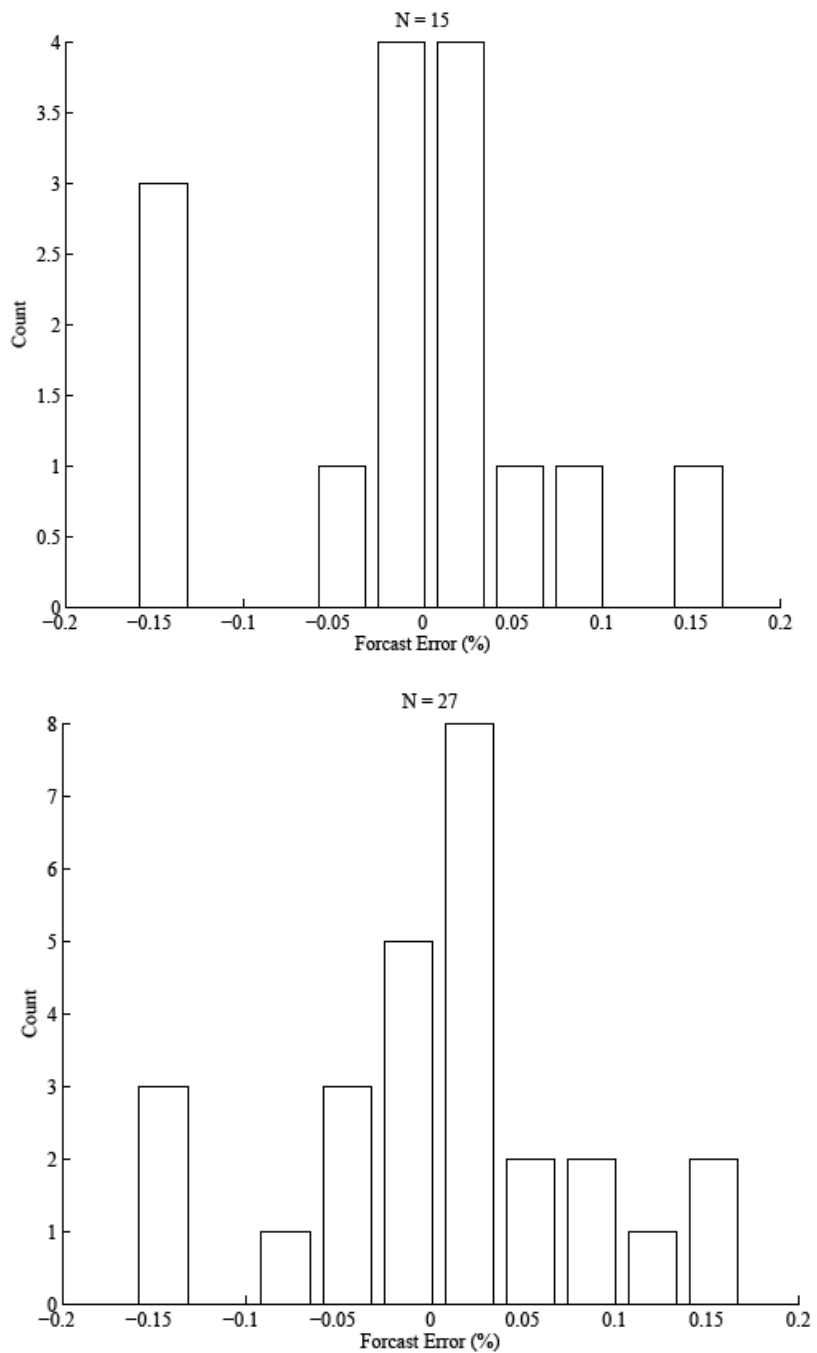


Figure 2. Histogram Estimates for USDA Soybean Price Forecasts Made in November Based on 1980/81-1994/95 (upper panel) and on 1980/81-2006/07 (lower panel) sub-samples.

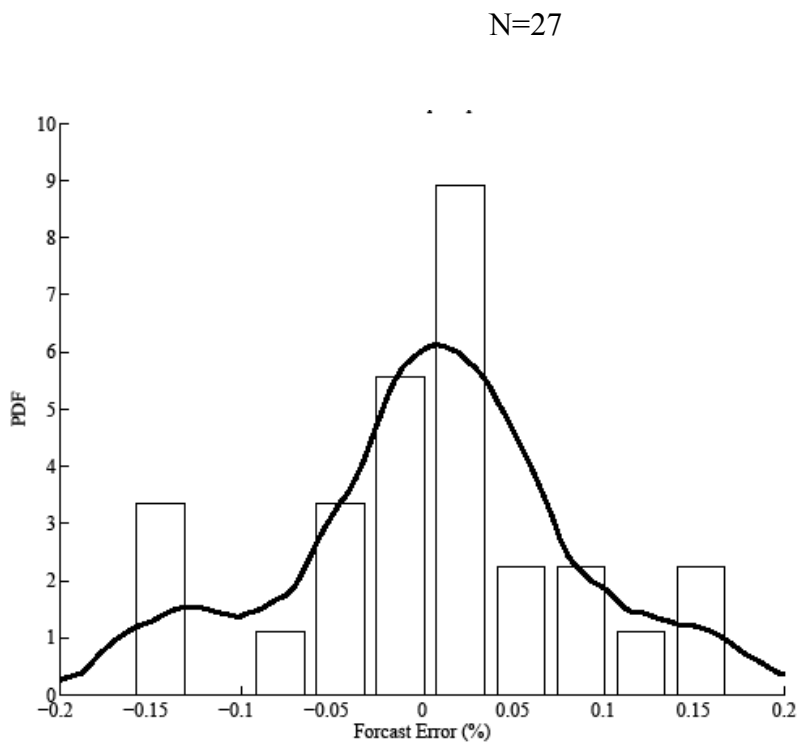
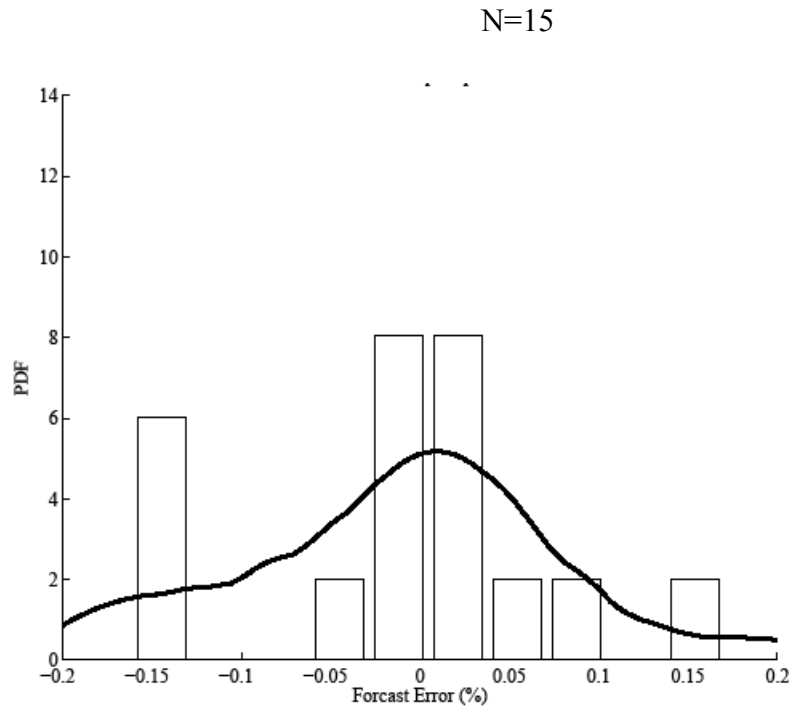


Figure 3. Kernel Density Estimates for USDA Soybean Price Forecasts Made in November Based on 1980/81-1994/95 (upper panel) and on 1980/81-2006/07 (lower panel) sub-samples.

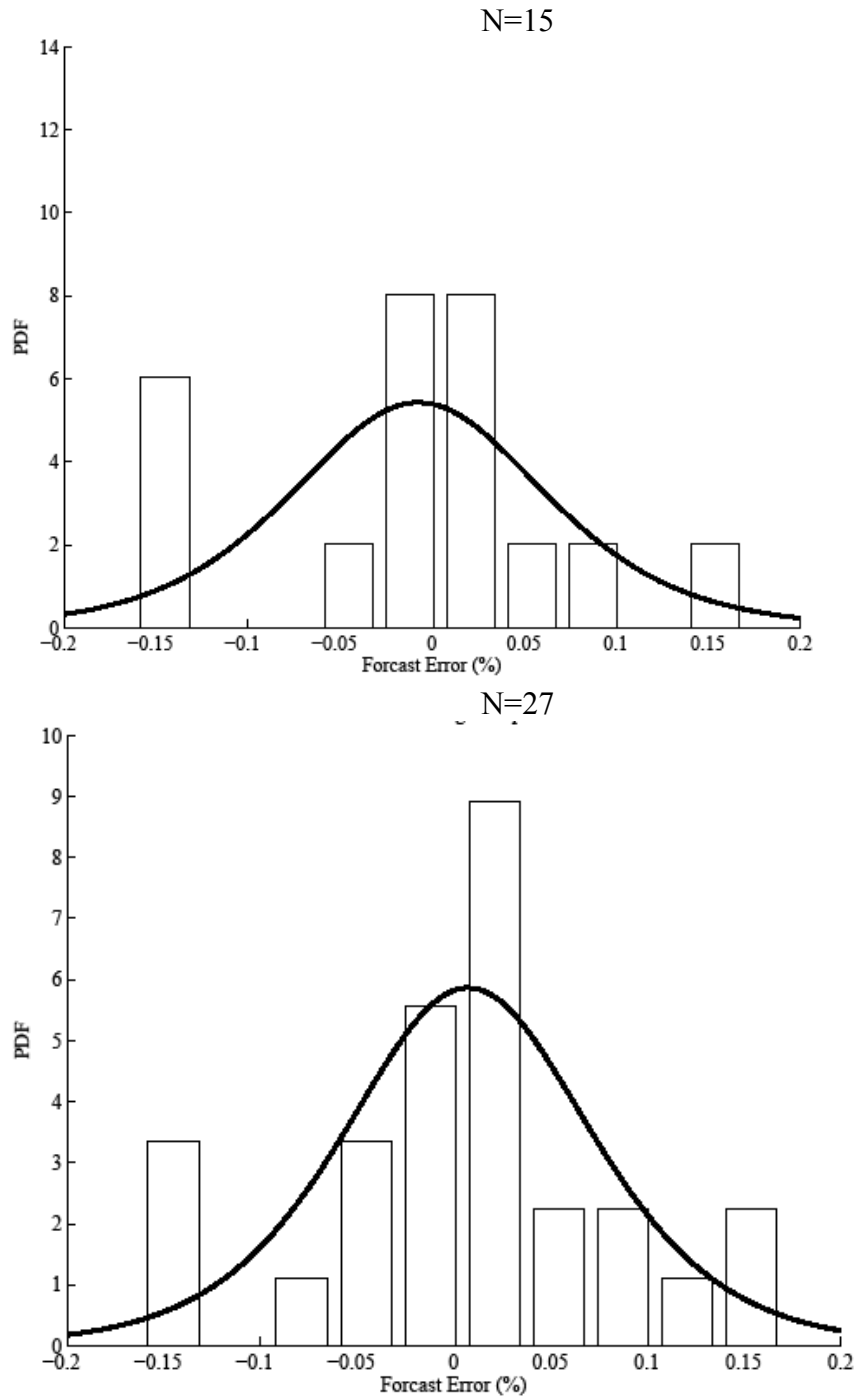


Figure 4. Logistic Distribution Estimates for USDA Soybean Price Forecasts Made in November Based on 1980/81-1994/95 (upper panel) and on 1980/81-2006/07 (lower panel) sub-samples.

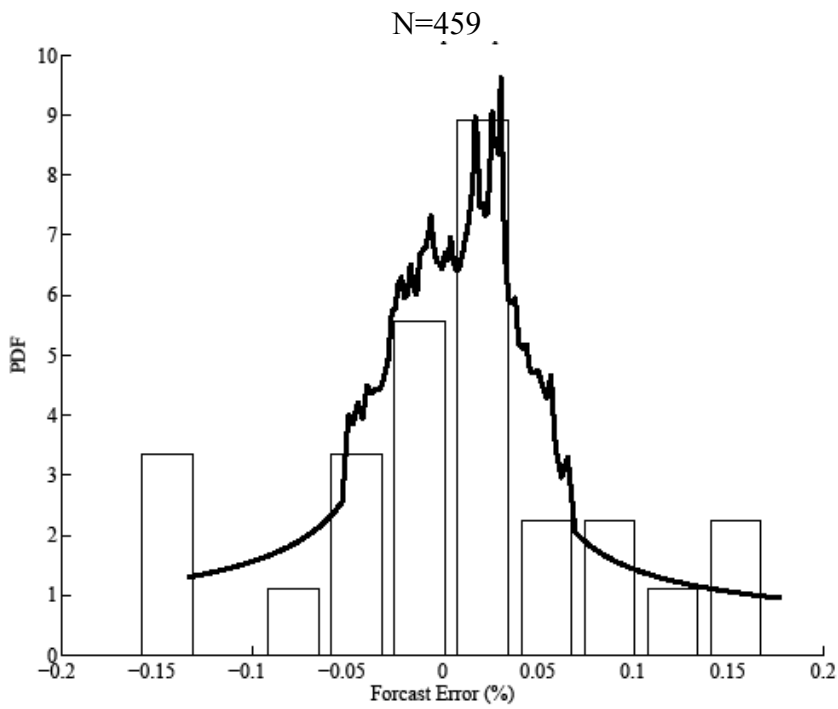
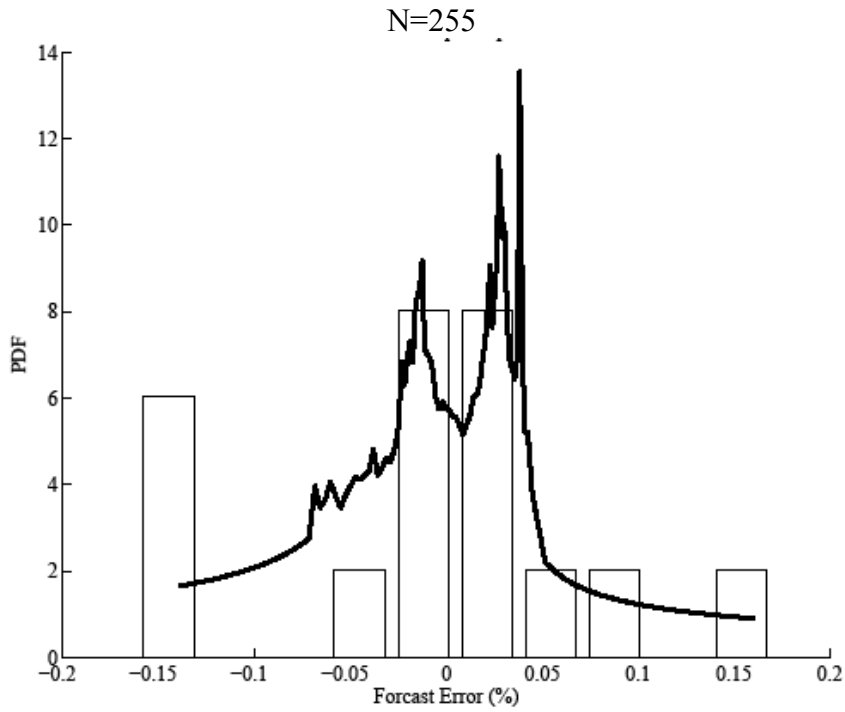


Figure 5. Quantile Regression Estimates for USDA Soybean Price Forecasts Made in November Based on 1980/81-1994/95 (upper panel) and on 1980/81-2006/07 (lower panel) sub-samples.

Footnotes

¹ It is worth noting that most theoretical variance expressions are based on the same assumption.

² Isengildina, Irwin, and Good (2004) provide survey evidence that USDA price intervals are symmetric. That is, a midpoint is forecast and then an equal interval is added to each side of the midpoint. Therefore, the average forecast price is computed in this study by taking an average of the midpoint of forecast prices for each month.

³ The last several months (17 and 18 for corn and soybeans and 14 and 15 for wheat) were not included in the analysis because errors were often zero, so the distributions were impossible to estimate.

⁴ It is important to note that the forecasting cycle for each of the commodities included in this analysis exceeds one calendar year. This implies that the final estimates (and thus forecast errors) for the previous year are not known until the 6th month into the next forecasting cycle for corn and soybeans and the 4th month for wheat. Therefore, the estimation sub-sample is one observation smaller in these early months of the forecasting cycle, i.e. while most intervals for 2006/07 are constructed using 26 prior observations, only 25 observations are available for the first five months of the corn and soybean forecasting cycle and the first three months of the wheat forecasting cycle.

⁵ This procedure was used by Isengildina, Irwin, and Good (2004) to construct 95% confidence level forecast intervals and implied confidence level forecast intervals as described in footnote 9 of their paper.

⁶ The bin width of 0.05% was selected for this study, which was the finest bin based on the data, the coarser bin could have resulted in wider intervals.

⁷ Other distributions, such as gamma and Weibull were not included in this analysis because they are not defined for negative values.

⁸ To calculate the chi-squared statistic, one must break up the x-axis domain into “bins” or classes of data. However, there are no clear guidelines for selecting the number and location of the bins and different conclusions can be reached from the same data depending on how bins are specified.

⁹ Computations were also made where the best-fitting distribution was allowed to change from year-to-year for corn. These alternative computations yielded quite similar results to those limited to the logistic distribution and are available from the authors upon request.

¹⁰ It is important to keep in mind that these interval widths are conditional on the selected confidence levels (80-90%). If the selected confidence level was equal to the actual hit rates, the widths of the empirical confidence intervals would be approximately equal to that of published intervals.

¹¹ There is some evidence that mean forecast error of corn prices prior to harvest also shifted to the right, but the shift was not statistically significant.

¹² Christoffersen (1998) also proposed additional tests that examine interval forecast independence and forecast coverage conditional on independence. These tests, however, should only be applied across marketing years. For example, the May forecast for 2005/06 should be independent from the May forecast for 2006/07. Independence, however, is not possible between April and May forecasts for 2006/07, as these are forecasts of the same event. Therefore, our approach of combining pre- and post- harvest forecasts for accuracy analysis would yield meaningless results due to violation of independence across forecasts of the same events. Thus, due to a small number of observations, these tests cannot be applied reliably to the prediction part of the sample (1995/96-2006/07).