NCCC-134
APPLIED COMMODITY PRICE ANALYSIS, FORECASTING AND MARKET RISK MANAGEMENT

Optimal Length of Moving Average to Forecast Futures Basis

by

Robert B. Hatchett, B. Wade Brorsen,
and Kim B. Anderson

Optimal Length of Moving Average to Forecast Futures Basis

Robert B. Hatchett,

B. Wade Brorsen,

and

*Kim B. Anderson**

Paper presented at the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management
St. Louis, Missouri, April 20-21, 2009

Robert Hatchett (Rb.hatchett@gmail.com) is a risk analyst at Gavilon, LLC in Omaha, NE and a former MS graduate research assistant at Oklahoma State University; B. Wade Brorsen is a regents professor and Jean & Patsy Neustadt Chair in the Department of Agricultural Economics at Oklahoma State University, and Kim B. Anderson is Charles A. Breedlove Professor of Agribusiness and extension economist in the Department of Agricultural Economics at Oklahoma State University. Partial funding was provided by the Oklahoma Agricultural Experiment Station through Hatch project OKL02170.

Optimal Length of Moving Average to Use When Forecasting Basis

Futures prices when combined with a basis forecast provide a reliable way to forecast cash prices. The most popular method of forecasting basis is historical moving averages. Given the recent failure of longer moving averages proposed by previous studies, this research reassesses past recommendations about the best length of moving average to use in forecasting basis. This research compares practical preharvest and storage period basis forecasts for hard wheat, soft wheat, corn and soybeans to identify the optimal amount of historical information to include in moving average forecasts. Only with preharvest hard wheat forecasts are the best moving averages longer than 3 years. The differences in forecast accuracy among the different moving averages are small and in most cases the differences are not statistically significant. The recommendation is to use longer moving averages during time periods (or at locations) when there have been no structural changes and use last year's basis after it appears that a structural change has occurred.

Keywords: Basis forecast, grain, Law of One Price, moving averages, structural change

Introduction

Creating reliable preharvest price expectations and making postharvest storage decisions depend heavily on accurate basis forecasts. Without accurate forecasts of basis "it is impossible to make fully informed decisions about…whether to accept or reject a given price; (and) whether and when to store your crop" (CBT, 1990, p.23).

The most popular method of forecasting the basis is historical moving averages. The attractiveness of these models is their ease of application. Access to local prices is cheap and readily available, allowing basis values to be localized for specific markets. Studies have applied forecasts of various lengths in order to determine the optimal length of years to include. These models generally conclude that longer averages ranging from 3 to 7 years are optimal (Dhuyvetter and Kastens, 1998; Sanders and Manfredo, 2006). The idea is that these longer moving averages can smooth out temporary deviations in markets.

In stable market conditions, the longer historical average forecasts proposed by previous studies should form the most accurate basis expectations. These methods have failed recently as basis values have deviated greatly from previous levels, resulting in poor forecasts based on historical basis. Given this recent failure, there is a need to reassess past recommendations about the best length of moving average to use in forecasting the basis. This study meets this need by determining which length of moving average has been most accurate in forecasting basis in terms of mean absolute error. Four commodities are considered: soft wheat, hard wheat, corn, and soybeans.

Theoretical Model

One of the primary reasons futures markets were created was to provide market participants the opportunity to exchange cash price risk for more manageable basis risk. Basis risk is preferred to price risk because price levels are more variable than basis levels. This price variability can be shown mathematically as

(1) $$\sigma^2_{price} > \sigma^2_{basis},$$

where σ^2_{price} is the variance of the cash market price and σ^2_{basis} is the variance of the basis. Basis forecasting seeks to reduce σ^2_{basis} by reducing forecast error (ε_t):

(2) $$\varepsilon_t = Basis_t - Ba\hat{s}is_t$$

where $Basis_t$ is the actual basis at time t, and $Ba\hat{s}is_t$ is basis forecast, and $\varepsilon_t \sim N(0, \sigma^2_{basis})$ assuming unbiased forecasts.

The most popular practical approach to forecasting basis is historical moving averages (FarmDoc 2009; Dhuyvetter 2009). Moving average models use the simple average of the previous N years:

(3) $$Ba\hat{s}is_t(N) = \frac{1}{N}\sum_{i=1}^{N} Basis_{t-i}.$$

By substituting (3) into (2) we can define how the optimal moving average length is selected to minimize basis forecast error

(4) $$\min_N E(\hat{e}_t^2) = \min_N E(Basis_t - \frac{1}{N}\sum_{i=1}^{N} Basis_{t-i}).$$

.

Rather than take the partial derivative of (4) with respect to N, this equation must be solved through enumeration due to the choice variable N being discrete. Once these individual forecasts are aggregated, the optimal forecast minimizes the error for the entire sample, T by

(5) $$\min_N \sum_{t=1}^{T}(Basis_t - \frac{1}{N}\sum_{i=1}^{N} Basis_{t-i})$$

The variance minimizing moving average length depends on the underlying stochastic process. Under normality and homoskedasticity the stochastic process for basis is

(6) $$Basis_t \sim N(\mu_t, \sigma^2)$$

where μ_t is the time varying mean and σ^2 is variance. The optimal moving average forecast length depends on μ_t.

Without structural change, the mean basis is $\mu_t = \mu$, and the longest moving average (largest N) would result in the minimum variance forecast. Basis forecast error variance in this case is

(7) $$\sigma_{forecast}^2 = \frac{\sigma^2}{N} + \sigma^2 \,.$$

These two sources of error originate in equations (5) and (6), in the variance of the moving average forecast, and in the current basis variance. So long as $\mu_t = \mu$, then as $N \to \infty, \sigma^2/N \to 0$, and the primary source of basis forecast error is σ^2. Therefore markets that are not prone to structural changes would find longer moving average forecasts optimal.

Structural changes within grain markets can change the dynamics of price relationships, and the resulting basis values. An extreme example of a stochastic process that could explain changes in markets is a random walk:

(8) $$\mu_t = Basis_{t-1} \,.$$

An example of a random walk process would be a permanent increase in transportation costs, which would widen the basis. With a random walk, as (8) shows, the optimal forecast is with *N*=1.

A more general stochastic process that includes both the constant mean and random walk models as special cases is a variation in a normal jump process. Diffusion-jump processes that combine a normal and a Poisson jump process are popular processes for modeling stochastic volatility in equity, stock and options markets (Yang and Brorsen 1993; Anderson et al. 2002; Chernov et al. 2003; Bates 1996). A discrete-time version such as in Pebe Diaz et al. (2002) is more appropriate here rather than the usual continuous-time model. With this model, the mean is constant and then occasionally changes as

(9) $$\mu_t = \mu_{t-1} + J_t P_t$$

where $J_t \sim N(\theta, \delta^2)$ and P_t is the jump process that is often assumed to follow a Poisson distribution. The difficulty in measuring this process is that the jump parameters and probability of the jumps occurring varies over time. Equation (9) could result in a random walk if $P_t = 1$ and $\delta^2 = 0$ in (6), and it gives a constant mean if $P_t = 0$. Ethanol plants are a major source of new demand in corn markets and cause the basis levels near the plant to strengthen. The structural change reflected by the jump affects prices initially, making the previous year's basis the optimal predictor for the year following the jump. The size of the shock in basis drastically changes the current basis levels so that all data before the change no longer reflect the current market. As the supply feeds the plant and markets adjust, bids will gradually decrease and the effects from the initial jump will result in a new mean and longer moving averages will then become optimal.

Mean-reverting models can also be used to model changes from historical basis levels (Jiang and Hayenga 1998; Sanders and Manfredo 2006). The basic mean-reverting model is the autoregressive moving average, or ARMA(p, q),

(10) $$Basis_t = \alpha + \varepsilon_t + \sum_{i=1}^{p} \emptyset_i Basis_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

where α is an intercept, $\emptyset_1 \ldots \emptyset_p$ are the autoregressive parameters, p is the number of autoregressive terms, q is the number of moving average terms, $\theta_1 \cdots \theta_q$ are the moving average parameters, and $\varepsilon_t \sim N(0, \sigma^2)$. If ($\emptyset_i = 1, p = 1$) and ($q = 0$) then it is a random walk, and if ($\emptyset_i = 1/p$) and ($q = 0$) then it is a simple moving average.

If the ARMA model in (10) is stationary, then the basis will converge toward its long-run mean of $\alpha / \sum_{i=1}^{p} \phi_i$. If the ARMA model is nonstationary (has a unit root) then the long-run mean will change over time. While Wang and Tomek (2007) argue that cash prices do not have unit roots, it is hard to argue that the mean of the basis is constant over time.

If plenty of observations are available, estimating an ARMA model should outperform the simple moving average of basis. But time series are often too short or structural changes are too frequent to estimate an ARMA model. Even if ARMA models could provide slightly more accurate forecasts, ARMA models may still not be preferred because of the difficulty in estimating and explaining them to producers.

ARMA (p, q) models, and another generalization, a seasonal autoregressive integrated moving average or SARIMA(p, d, q), have been used to forecast the basis (Sanders and Manfredo, 2006; Jiang and Hayenga, 1998). These studies found little improvement in forecast accuracy when compared to the moving average models. In order to identify the correct level of $\emptyset_i$, the appropriate covariance function of the process must be identified by the partial autocorrelation and autocorrelation plots. This econometric technique is too complicated for many producers to understand, and is not modeled in this study for that reason. Instead, this research focuses on simple moving average forecasts, which are ARMA $(p, 0)$ processes where $\emptyset_i = 1/p$ and $\theta_i = 0$.

The optimal length of moving average to forecast the basis is expected to depend on the size and frequency of structural changes. When conditions are static, longer moving averages are optimal. However, after a structural change occurs, the optimal length of a moving average is one. According to the Law of One Price, the basis is the difference between two prices and so it can reflect differences in time, form, and space. Since time differences are held constant, the structural change can reflect changes in space or form.

Previous Literature

Numerous variables have been used in regression models to explain the basis. Most of these variables correspond to differences in time, form, and space, but the theoretical basis for some of these variables is not as clear. Differences in form are explained through components of the futures price not reflected in the cash market price. Cost of storage and transportation measures are accepted components of the basis from literature that explain the transformation of prices over time and space, but the theoretical support for supply and demand variables used to explain the basis over space is not as clear.

Supply and demand variables at local markets can explain the basis over space. Supply variables for markets include crop production levels, a dummy variable for the presence of loan deficiency payments (LDP), the ratio of Eastern Canadian corn production to consumption, and

Western feed grain availability (Dykema, Klein, and Taylor, 2002; Martin, Groenewegen, and Pigeon, 1980; Jiang and Hayenga, 1997). Soybean crushing levels, animal units consuming grain (corn), corn usage estimates, and export volumes were all used as demand variables to identify the differences in markets (Jiang and Hayenga, 1997, Dykema, Klein, and Taylor, 2002). Models of livestock basis have also considered a wide variety of explanatory variables (Naik and Leuthold 1991; Liu et al. 1994). These supply and demand variables represent proxy variables used to identify the factors that constitute the basis at a particular location. A wide variety of variables are used to explain the basis. These variables should correspond to aspects of the Law of One Price, and explaining the basis through time, form, and space. A major drawback of the explanatory models is that usually use data that would not be available at the time forecasts need to be made.

Several studies have applied moving averages of various lengths to identify the most accurate method of forming basis expectations. Hauser et al. (1990) compared several naïve models in forming their soybean basis expectations for ten Illinois elevators. Dhuyvetter and Kastens (1998) forecast nearby basis for wheat, corn, soybeans and grain sorghum for multiple Kansas locations using historical moving averages and current market information. Sanders and Manfredo (2006) tested a 5-year moving average, the previous year's basis, and the expected nearby basis is the ending basis models are compared against more advanced times series methods. Taylor, Dhuyvetter, and Kastens (2004) revisited Dhuyvetter and Kastens (1998), and included models to determine the optimal amount (weight) of current market information, the current basis deviation from the moving average, needed to improve forecast accuracy.

Table 1 lists the results from these forecasting studies. These results do not provide a clear pattern in what forecast performs the best. From the table we can see that practical forecasts perform comparably to more complex forecasts. The optimal amount of historical data included in the forecasts does not follow any rule of thumb. And the inclusion of current information is shown to increase forecast accuracy over short horizons, but its effectiveness diminishes greatly with time. No clear patterns in the amount nor kind of current information to consistently improve basis forecasts exists. These inconsistent findings suggest the need for further research.

Data

The commodities considered are corn, soybeans, soft wheat, and hard wheat. To create the basis data, futures prices are subtracted from their corresponding cash price.

Two basis values are used for each year. One is selected to represent the basis for a preharvest hedge and the other for a storage hedge. For corn, the December contract in October represents the harvest basis, while the May contract in April represents storage hedges. For soybeans, the November contract in October represents the harvest basis, while the May contract in April represents storage hedges. The basis values used for soft and hard wheat are the July contract in June and the December contract in November.

Table 1. Results from Previous Basis Forecasting Studies

Study	Optimal Forecasts	Conclusions
"Forecasting Crop Basis: Practical Alternatives" -Dhuyvetter and Kastens (1997)	• 4-year moving average for wheat. • 7-year moving average for corn. • 7-year moving average for soybeans. • 5-year moving average for milo.	Futures price spreads and current nearby basis increased accuracy, but futures price spreads were best. The benefit from incorporating current market information diminished beyond 4-12 weeks.
"Incorporating Current Information into Historical-Average-Based Forecasts to Improve Crop Price Basis Forecasts" – Taylor, Dhuyvetter, and Kastens (2004)	• 3-year moving average for wheat. • 2-year moving average for corn. • 3-year moving average for soybeans. • 2-year moving average for milo.	Futures price spreads and current basis deviations from historical levels helpful in post-harvest and harvest (only 4 weeks prior to harvest). As the post harvest horizon approached, the optimal amount of current market information increased.
"An Analysis of Anticipatory Short Hedging Using Predicted Harvest Basis" - Kenyon and Kingsley (1980)	• Regression equation using initial local cash and Chicago futures market prices, the Chicago cash price at planting, and the residual of open interest.	The regression estimates predicted 73-81% of the change in corn basis, and 95%-97% of the change in soybean basis as harvest approached using initial basis and the difference between actual and predicted open interest.
"Basis Expectations and Soybean Hedging Effective" – Hauser, Garcia, and Tumblin (1990)	• 1 or 3-year historical basis during preharvest. • Futures price spreads after the harvest.	Forecasts that include the implied return to storage outperform historical averages in 2 of the 3 contract periods. Historical average models perform comparably to models incorporating current market information.
"Corn and Soybean Basis Behavior and Forecasting: Fundamental and Alternative Approaches" - Jiang and Hayenga (1998)	• 3-year moving average plus current market information best for corn. • Seasonal ARIMA best for soybeans.	Although the 3-year moving average performs relatively well, it is out performed by models that include current market information and seasonal ARIMA models.
"Forecasting Basis Levels in the Soybean Complex: A Comparison of Time Series Methods" - Sanders and Manfredo (2006)	• ARMA model best for soybeans. • VAR model best for soybean meal. • Previous year's basis best for soybean oil.	Over time, the accuracy of the 1 and 5-year moving averages do not diminish. Even within closely related markets there is no rule-of-thumb for producing the most accurate forecasts.

Cash and futures prices consist of second Wednesday or Thursday prices for corn, soybeans and wheat, and when unavailable, monthly-average prices are used. Daily #2 corn and #1 soybean cash prices are from the Illinois Agricultural Marketing Service, and reflect the midrange of elevator bids for each region on the second Thursday of each month from 1975-2008 (FarmDoc, 2009). When the second Thursday fell on a holiday, the third Thursday was used. Second Wednesday daily Oklahoma reported prices paid to producers for #2 hard red winter wheat were taken from the Oklahoma Department of Agriculture, Food and Forestry's weekly "Oklahoma Market Report" from 1974 through 2008. This report also provides the Galveston Gulf Port prices. When a holiday prevented the release of the report, the third Wednesday was used. Second Wednesday prices from an additional Oklahoma location, the Port of Catoosa, are for 1988-2008 (Peavey Grain, 1988-2008). Second Wednesday Kansas cash prices cover 1982-2007 (Dhuyvetter, 1982-2007). Simple average monthly wheat prices were taken from the USDA AMS "Grain and Feed Market News" for #2 soft red winter wheat at Chicago, IL, Toledo, OH, and St. Louis, MO, along with #1 hard red winter wheat at Kansas City, MO over 1970-2008. Figure 1 shows the Kansas and Oklahoma hard red winter wheat locations studied.

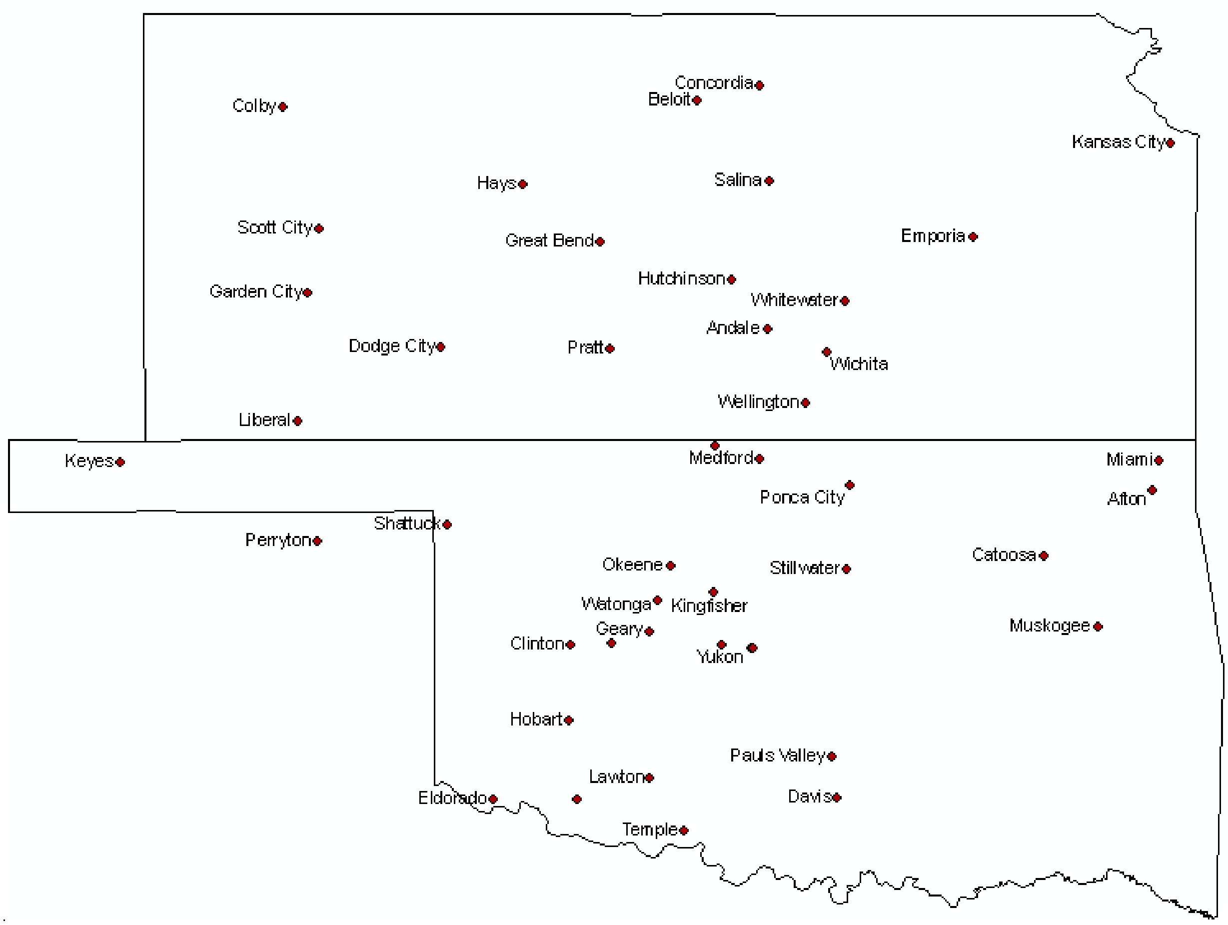


Figure 1. Kansas and Oklahoma Hard red winter wheat locations studied

Futures prices reflect daily closing prices at the CBT and KCBT for each commodity (R & C Data), and match the same days as the cash prices. When only monthly cash prices were available, average monthly futures prices were used. Corn, soybeans, and soft wheat futures prices are reflected by CBT contracts, while KCBT wheat contracts reflect hard wheat. These futures prices, along with their corresponding cash prices, provide the nearby basis values used.

The data series were checked to ensure that none of the days studied happened to fall on days when the futures price hit the daily limit. The earliest reported historical daily price limits for the CBT were found to be 30 cents per bushel for soybeans, 10 cents per bushel for corn, and 20 cents per bushel for both soft and hard wheat as of 1982 (CBT, 1982). The earliest change to KCBT daily price limits occurred when the limit increased from 10 cents per bushel in 1973, and it is assumed that these levels rose to the CBT limit of 20 cents per bushel. These values were assumed to have remained constant in the preceding years. Price limits remained stable until March 12, 1992 when CBT corn price limits increased from 10 to 12 cents per bushel, while soybean and wheat limits remained at 30 and 20 cents per bushel, respectively (Park, 2000). On August 14, 2000 daily price limits increased at the CBT from 12 to 20 cents per bushel for corn, from 30 to 50 cents per bushel for soybeans, and from 20 to 30 cents per bushel for wheat (CFTC). The KCBT limit changed when the wheat price limit was raised from 25 to 30 cents on October 9, 2000 (Summers). On March 28, 2008 the KCBT and CBT both doubled the 30 cent price limit for wheat futures to 60 cents, while the CBT also expanded trading limits from 50 to 70 cents for soybeans and 20 to 30 cents for corn (CMEGroup). None of the limit days occurred on one of the days of interest to this study.

Note that one concern is that this study's findings that favor shorter moving averages could be due to the recent structural changes. Hatchett (2009) investigates this possibility and concludes that the findings are not fragile with respect to deleting the 2007-2008 data.

Procedures

Basis values were created by taking the cash market price less the futures market price. Basis forecasts were created using equation (3), where N=1,…,5. The resulting forecast errors from each model were then evaluated.

As in Dhuyvetter and Kastens (1998), we compare forecast accuracy with mean absolute error:

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|Basis_t - Ba\hat{s}is_t| \tag{11}$$

where the absolute value of each forecast error is averaged over the forecast period. This measure of forecast accuracy will be used in this study to identify the optimal historical period to include in basis forecasts. The root mean squared error was also considered, but results did not differ substantially (Hatchett 2009) and so they are not included here.

The complex nature of the variance-covariance matrix of the error term when modeling time-series cross sectional data makes misspecification a concern when modeling basis forecast errors. Econometric problems prevalent with this type of data include spatial autocorrelation, cross correlations, and heteroskedasticity. Failing to correct for these correlations and unequal error variance can lead to biased and inconsistent standard errors and hypothesis tests. Dhuyvetter and Kastens (1998) tested for heteroskedasticity, and identified groupwise heteroskedasticity amongst forecast methods and time horizon variables for corn, soybeans, and wheat forecasting models. To correct for this heteroskedasticity, interaction terms of methods and forecast time horizon squared were included in each of their separate models. Although the dependence of the errors amongst competing forecast models could not be corrected, Dhuyvetter and Kastens (1998) conclude that a 4-year moving average was more accurate than the 3-year benchmark at 0.01 significance. When independence across observations is incorrectly assumed, the standard errors and their ensuing t-tests lead to overstated significance (Irwin, Good, and Martines-Filho, 2006).

A variation of the Dhuyvetter and Kastens (1998) approach to correct for heteroskedasticity was attempted with both the aggregate dataset and the individual commodities in this study. The pooled data set contains 15,180 observations. To correct for unequal variance using random effects, combinations of variables such as *period*location* and *location*year*, where *period* represented the preharvest or storage contract, *location* identified the market, and *year* identified the year of the forecast, were considered. However, these interaction terms resulted in too many parameters, which prevented the model from converging. As an alternative, we follow Irwin, Good, and Martines-Filho (2006) and pool the data, leaving the forecast length *N* as the only independent variable, and with time included as a random effect . This model was also run for the individual commodities by period to identify any patterns that would be lost in the pooled model. The final mixed model is:

$$AE_{it} = \beta_0 + \sum_{j=1}^{4} \beta_j D_{ij} + v_t + \varepsilon_{it} \tag{12}$$

where AE_{it} is the absolute error of the ith forecast, at time t, β_0 is an intercept term created for the 5-year moving average to serve as a benchmark for model comparison, and β_j , $j = 1, \dots, 4$, are the coefficients for moving averages of j length, where D_{ij}=1 when $i = j$, v_t is the random-effects vector for years at time t and ε_{it} is the stochastic error term for the observation i at time t. The random-effects vector and stochastic error term are uncorrelated, and are distributed $v_t \sim N(0, \sigma_v^2)$ and $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$.

Pooled Model Results

Table 1 shows the optimal forecast length by year for the pooled data. From this table we can see that the previous year's basis provides the optimal forecast for 37.51% (1144/3,050) of the values. The 5-year moving average produces the second most optimal forecasts at 25.77%, while the 2, 3, and 4-year moving averages account for 14.59, 11.64, and 10.49% of the sample, respectively.

Table 1. Number of Locations with a Given Length of Moving Average Having the Lowest Root Mean Squared Forecast Error, 1975-2008

Commodity	Period	*N*=1	*N*=2	*N*=3	*N*=4	*N*=5
Hard wheat	Preharvest	25	2	5	7	6
	Storage	34	2	4	1	4
Soft wheat	Preharvest	3	0	0	0	0
	Storage	0	2	0	0	1
Corn	Preharvest	0	0	0	0	7
	Storage	7	0	0	0	0
Soybeans	Preharvest	2	5	0	0	0
	Storage	7	0	0	0	0

Figure 2 graphs the number of optimal forecasts produced by the previous year's basis and 5-year moving average for the pooled data. The one-period forecast is usually close to the 5-year forecast, but following periods of structural change like the early 1980's (inflation, collapse of land prices, oil price shocks, etc.), 1988 (US-Canada free trade) and 2006 (lack of convergence at contract expiration) there are many more optimal forecasts using the one-period forecast. This preference for shorter moving averages shows the inferiority of basing expectations on longer moving averages after times of structural change.

Figure 2. Number of minimum MAE forecasts produced by the previous year's basis vs. the 5-year moving average, 1979-2008

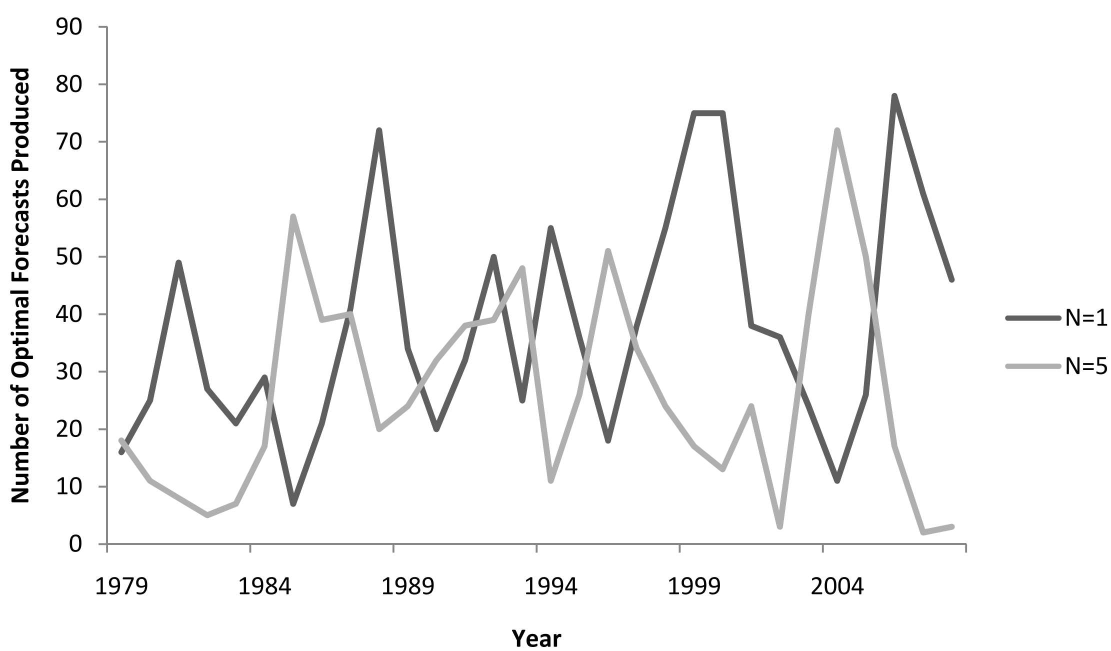


Table 2 shows the results from the pooled model of absolute forecast errors for the entire study. The F-value of 0.92 fails to reject any difference amongst the competing forecast methods. The intercept gives the absolute error of the 5-year moving average, and is 12.89 cents/bu. Forecast accuracy increases as the amount of historical information used decreases, with the previous year's basis providing the lowest pooled MAE at 12.34 cents/bu. These results are generally within the range of the MAE's found by previous studies. Dhuyvetter and Kastens (1998) find the pooled MAE's of moving average forecasts to be between 10-13 cents/bu. for wheat, corn, and soybeans. The individual t-tests show that the one year forecast has a significantly lower absolute error than the 5-year moving average, but none of the other differences are statistically significant.

Table 2. Absolute Error (cents/bu.) of Basis Forecasts as a Function of Number of Years in the Moving Average, Pooled Data, 1975-2008

Effect	Estimate	t-value	p-value
Intercept	12.34	12.06	0.000
N=1	-0.57	-2.06	0.040
N=2	-0.22	-0.79	0.427
N=3	-0.16	-0.58	0.562
N=4	-0.05	-0.18	0.858
N=5	-	-	-
F-statistic[a]	1.31	-	0.263

[a] The null hypothesis is that all values of *N* have the same forecast accuracy.

#2 Hard Wheat Model Results

Preharvest and storage hard wheat basis forecasting model results are in Table 3. For the preharvest forecasts, the 2-year moving average has significantly higher forecast error than the 5-year benchmark. The only preharvest model to produce a lower MAE than the benchmark is the 4-year moving average, which improves by only 0.04 cents/bu. These results indicate that, over the sample, any of the 5 preharvest models considered would result in a forecast error of approximately 13 cents/bu so the differences between methods are small.

The storage model results for hard wheat reject the joint test of no differences in forecast accuracy with an F-statistic of 10.85. Individual t-tests of no difference from the 5-year benchmark are rejected for all but the 4-year moving average. The previous year's basis lowers the benchmark MAE from 13.03 cents/bu. to 11.09 cents/bu. The improvement in accuracy as the historical period shortens supports using shorter moving averages to forecast the hard wheat storage basis.

Table 3. Absolute Error (cents/bu.) of Hard Wheat Basis Forecasts as a Function of Number of Years in the Moving Average, 1978-2008

Period	Effect	Estimate	t-value	p-value
Preharvest	Intercept	12.77	8.71	0.000
	N=1	0.35	1.06	0.291
	N=2	0.68	2.06	0.040
	N=3	0.41	1.24	0.216
	N=4	-0.06	-0.19	0.853
	N=5	-	-	-
	F-statistic[a]	1.73	-	0.141
Storage	Intercept	13.03	9.06	0.000
	N=1	-1.94	-5.90	0.000
	N=2	-1.09	-3.32	0.001
	N=3	-0.77	-2.33	0.020
	N=4	-0.23	-0.70	0.481
	N=5	-	-	-
	F-statistic[a]	10.85	-	0.000

[a] The null hypothesis is that all values of *N* have the same forecast accuracy

Table 3 shows a pattern consistent throughout the forecast results. By studying the preharvest and storage basis separately, we can see that MAEs are greater for preharvest than storage models. One possible explanation of this difference comes from Dhuyvetter and Kastens (1998), who found that forecast errors peak during critical production periods. Local conditions are much more variable around harvest, and spatial differences between cash and futures markets may not reflect the same supply and demand.

Hard Wheat Changes over Time

Figure 2 is a map of the 1975-1980 average hard-wheat harvest basis values for Oklahoma. Basis values tend to be weakest in the northern part of the state, and grow stronger when moving south.

Figure 3 shows the 2008 harvest basis values. The trend from the first map is now reversed, with basis strengthening from the southern to the northern part of the state. A major shift in the primary market for Oklahoma wheat occurred over the period studied. Oklahoma wheat was shipped via rail to the Gulf Port at Houston, but now a portion of it near Catoosa, travels by barge to New Orleans. This change in the transportation of Oklahoma wheat over the time period studied may partially explain why Oklahoma wheat basis changed over the 30 plus years studied.

Figure 2. 1975-1980 average Oklahoma wheat harvest basis

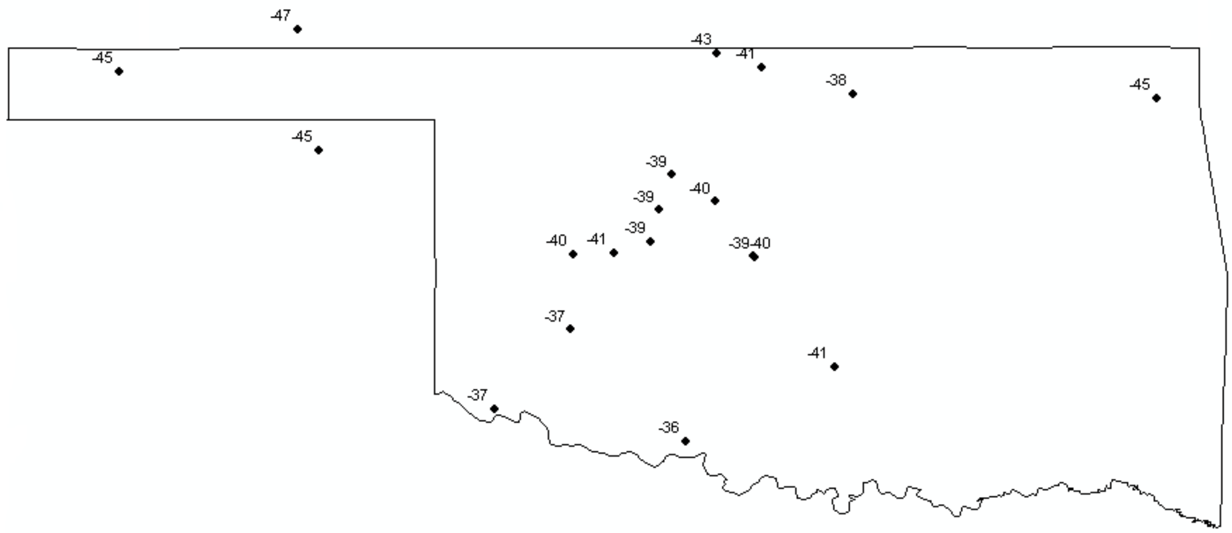


Figure 3. 2008 Oklahoma wheat harvest basis

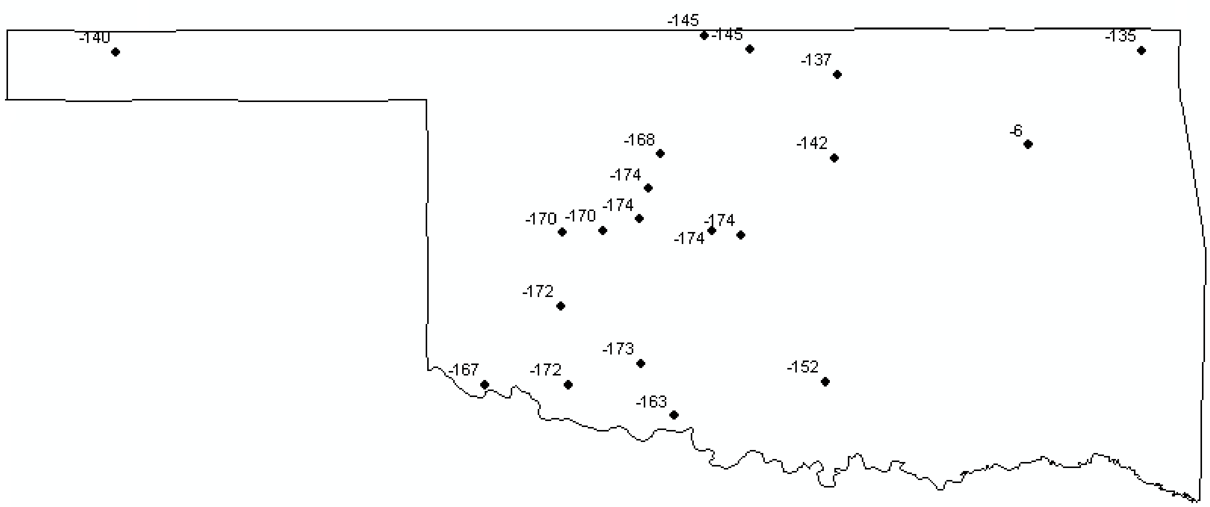
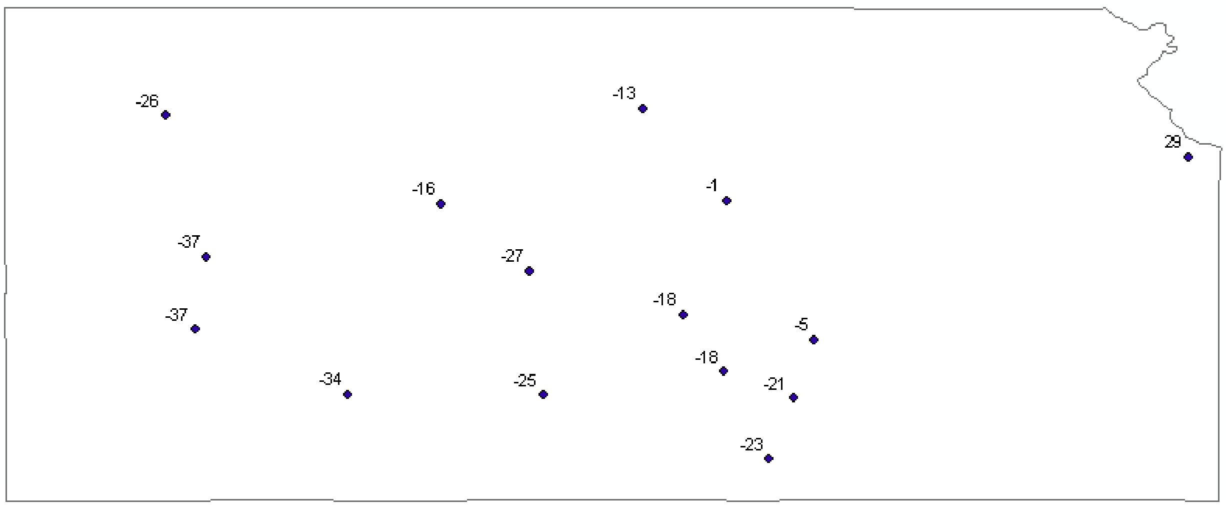


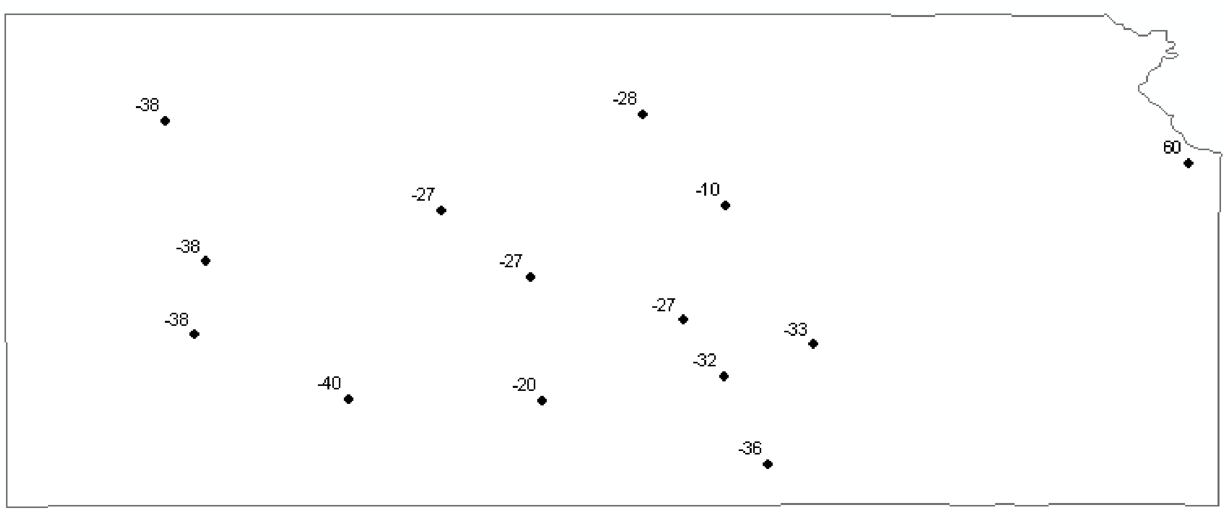
Figure 4 shows the 5-year average basis for Kansas locations over 1982-1986. The trend in this map is that the basis weakens when moving south and west away from Kansas City.

Figure 4. 1982-1986 Average Kansas wheat harvest basis



The 2007 Kansas harvest basis is shown in Figure 5. Similar to the relationships in Figure 3, the harvest basis tends to weaken moving away from Kansas City to the southwest. O'Brien (2009) discusses the importance of rail transportation to Kansas wheat producers, and rail rates can explain this weakness when moving away from Kansas City.

Figure 5. 2007 Kansas wheat harvest basis



The greatest difference between the two time periods is that most locations seem to be more closely aligned with the markets surrounding them. In Figure 4 there were isolated markets that experienced much stronger basis than their closest neighbors. These locations typically represent mills, which may have had greater discounts or premiums in the earlier period. Figure

5 shows that in 2007, nearly all of the locations are within a few cents of their surrounding locations. O'Brien (2009) proposes that the consolidation of coops throughout Kansas may explain the increased alignment of basis.

#1 Hard Wheat Model Results

The Kansas City price data allows this study to compare the differences in forecasting both the regular protein #1 hard red wheat, and 13% protein #1 hard red wheat. Table 4 shows the model results for the Kansas City ordinary protein #1 hard wheat basis models.

Table 4. Absolute Error (cents/bu.) of Kansas City Ordinary Protein, #1 Hard Wheat Basis Forecasts as a Function of Number of Years in the Moving Average, 1976-2008

Period	Effect	Estimate	t-value	p-value
Preharvest	Intercept	15.72	8.33	0.000
	N=1	0.47	0.29	0.770
	N=2	0.21	0.13	0.897
	N=3	-0.12	-0.07	0.942
	N=4	-0.96	-0.60	0.552
	N=5	-	-	-
	F-statistic[a]	0.23	-	0.923
Storage	Intercept	12.99	5.24	0.000
	N=1	2.17	1.35	0.180
	N=2	1.89	1.17	0.244
	N=3	1.83	1.14	0.259
	N=4	0.76	0.47	0.638
	N=5	-	-	-
	F-statistic[a]	0.65	-	0.629

[a] The null hypothesis is that all values of *N* have the same forecast accuracy.

Table 5 reports the results of the model for the 13 percent #1 hard wheat. This data allows space to be held constant, and shows the difference in forecast accuracy between two types of a commodity delivered on the same futures contract. The benchmark intercept for the 13 percent protein model is 3.22 cents/bu. higher than the ordinary protein forecast model. This best preharvest forecast is still 1.22 cents/bu. more than the worst ordinary protein forecast model.

Comparing the forecast results of ordinary and 13% protein #1 hard wheat shows the effect of differences in grain form on forecast accuracy. Forecast errors are lower in both periods for ordinary protein. Higher forecast errors for 13% protein are likely the result of changes in the variable premiums for protein content at KCBT. Rather than using a fixed premium similar to what exists between #1 and #2 grade wheat, the market posts a protein premium scale that allows for market adjustments to premiums (KCBT). Differences in supply and demand for ordinary and 13% protein wheat qualities explain the difference in form of these hard wheat markets.

Table 5. Absolute Error (cents/bu.) of Kansas City 13% Protein, #1 Hard Wheat Basis Forecasts as a Function of Number of Years in the Moving Average, 1976-2008

Period	Effect	Estimate	t-value	p-value
Preharvest	Intercept	18.94	4.11	0.000
	N=1	0.26	0.18	0.856
	N=2	-1.37	-0.95	0.344
	N=3	-1.53	-1.07	0.289
	N=4	-0.32	-0.22	0.824
	N=5	-	-	-
	F-statistic[a]	0.64	-	0.636
Storage	Intercept	19.34	3.75	0.001
	N=1	5.52	1.88	0.064
	N=2	3.55	1.20	0.231
	N=3	1.06	0.36	0.721
	N=4	0.77	0.26	0.793
	N=5	-	-	-
	F-statistic[a]	1.21	-	0.310

[a] The null hypothesis is that all values of N have the same forecast accuracy.

Soft Wheat Model Results

Table 6 displays the results for the soft wheat basis forecasting models. Using the previous year's basis to predict soft wheat preharvest basis would lead to an average forecast error of 25.95 cents/bu., while the most accurate method, the 2-year moving average, only lowers the MAE to 23.42. Only the 2-year moving average proves to be a better forecast of the storage basis than the 5-year benchmark for soft wheat. Although it decreases the MAE to nearly 13 cents/bu., the 2-year moving average is not significantly different from the benchmark. The storage basis is forecasted considerably more accurately than the preharvest basis.

Corn Model Results

Table 7 shows the results for the corn models across all regions of Illinois. Results from the preharvest model indicate that using the previous year's basis outperforms the 5-year benchmark over all Illinois locations. The F-statistic and individual t-tests both fail to indicate any significant differences in forecast choice. The F-statistic of 4.10 for the storage models rejects the null hypothesis, and concludes that model forecast accuracy does differ over the sample for corn storage basis. Significant differences from the 5-year benchmark exist in every model except the 4-year moving average at a 0.05 level. This result indicates that shorter moving averages can outperform the 5-year moving average at forecasting the corn storage basis. The best model, using the previous year's basis, lowers the MAE from the 5-year moving average of 7.49 cents/bu. to 6.32 cents/bu.

Table 6. Absolute Error (cents/bu.) of Soft Wheat Basis Forecasts as a Function of Number of Years in the Moving Average, 1975-2008

Period	Effect	Estimate	t-value	p-value
Preharvest	Intercept	23.45	4.64	0.000
	N=1	2.50	0.78	0.434
	N=2	-0.03	-0.01	0.991
	N=3	0.20	0.06	0.951
	N=4	0.48	0.15	0.882
	N=5	-	-	-
	F-statistic[a]	0.22	-	0.926
Storage	Intercept	13.91	8.52	0.000
	N=1	0.59	0.45	0.654
	N=2	-0.93	-0.71	0.479
	N=3	0.16	0.13	0.900
	N=4	0.51	0.39	0.696
	N=5	-	-	-
	F-statistic[a]	0.43	-	0.788

[a] The null hypothesis is that all values of *N* have the same forecast accuracy.

Table 7. Absolute Error (cents/bu.) of Corn Basis Forecasts as a Function of Number of Years in the Moving Average, 1980-2008

Period	Effect	Estimate	t-value	p-value
Preharvest	Intercept	11.74	9.60	0.000
	N=1	-0.12	-0.21	0.836
	N=2	0.63	1.07	0.286
	N=3	0.55	0.94	0.349
	N=4	0.52	0.87	0.385
	N=5	-	-	-
	F-statistic[a]	0.70	-	0.594
Storage	Intercept	7.49	7.59	0.000
	N=1	-1.17	-3.63	0.000
	N=2	-0.68	-2.12	0.034
	N=3	-0.66	-2.05	0.041
	N=4	-0.18	-0.56	0.574
	N=5	-	-	-
	F-statistic[a]	4.10	-	0.003

NOTE: Storage model forecasts begin in 1981 due to the time series available.

[a] The null hypothesis is that all values of *N* have the same forecast accuracy.

Soybean Model Results

Table 8 shows the absolute error of the basis forecasting models for Illinois soybean basis. The preharvest 5-year benchmark MAE is 11.23 cents/bu., and can be improved by all of the shorter moving-average models. The most improvement comes from the 2-year moving average, which lowers the MAE to 10.62 cents/bu. Although the benchmark can be improved upon, the improvement is not enough to be statistically significant based on the t-test. The narrow range (< 0.61 cents/bu.) of MAEs shows that little difference exists across preharvest basis models over the period studied.

The choice of forecasting model affects the accuracy of the Illinois soybean storage basis forecasts. The F-statistic of 8.58 indicates that the choice of models can result in different forecasting accuracies. While all of the shorter moving average models outperform the benchmark, both the previous year's basis and the 2-year moving average result in 1.98 and 1.16 cents/bu. lower forecasts, respectively. Compared to the soybean preharvest model, the storage basis forecasts result in decreased MAEs of over 1.6 cents/bu.

Table 8. Absolute Error (cents/bu.) of Soybean Basis Forecasts as a Function of Number of Years in the Moving Average, 1980-2008

Period	Effect	Estimate	t-value	p-value
Preharvest	Intercept	11.23	8.58	0.000
	N=1	-0.47	-0.78	0.438
	N=2	-0.61	-1.00	0.318
	N=3	-0.50	-0.82	0.410
	N=4	-0.20	-0.32	0.748
	N=5	-	-	-
	F-statistic[a]	0.43	-	0.852
Storage	Intercept	9.61	8.25	0.000
	N=1	-1.98	-4.99	0.000
	N=2	-1.16	-2.92	0.004
	N=3	-0.66	-1.66	0.100
	N=4	-0.08	-0.19	0.846
	N=5	-	-	-
	F-statistic[a]	8.58	-	0.000

Note: Storage model forecasts begin in 1981, due to the time-series available.
[a] The null hypothesis is that all values of N have the same forecast accuracy.

Conclusions

The most popular method of forecasting the basis is historical moving averages. Given the recent failure of longer moving averages proposed by previous studies, this research reassesses past recommendations about the best length of moving average to use in forecasting basis. Our study uses a longer time series with more locations and crops than these previous studies to determine the optimal length of historical data to forecast basis. The hypothesis testing procedure using the pooled data is valid in the presence of cross correlations.

Basis values for hard wheat, soft wheat, corn and soybeans were used to create basis forecasts. The methods considered included the previous year's basis and moving averages of the previous 2-5 years. Mean absolute error was modeled for the pooled data following Irwin, Good, and Martines-Filho (2006). The mean absolute error was the dependent variable, the forecast length was the independent variable, and time was the random effect.

This research found the size of most MAEs to be consistent with previous studies (Dhuyvetter and Kastens, 1998; Taylor, Dhuyvetter and Kastens, 2002). These values were generally between 10 and 17 cents/bu.

The optimal forecast length found for each commodity is generally shorter than previous recommendations. Using a 4-year moving average produced the minimum MAE preharvest wheat forecast, consistent with Dhuyvetter and Kastens (1998), but the optimal storage forecast model has lower forecast error using shorter historical information. This study finds that the optimal amount of historical data included in corn and soybean forecasts have shortened to one or two years for both preharvest and storage periods. Most differences in forecast accuracy among the different models are not statistically significant and most of the significant differences are with the storage basis forecasts..

Structural changes over the time period studied have led to recommending shorter historical moving averages to forecast the basis. Markets within this study undergo varying amounts of structural change for different reasons. Kansas wheat markets, for example, maintained consistent basis relationships over space, which may be due to their proximity to the KCBT hard wheat market delivery points. Toledo, OH and St. Louis, MO experienced more structural change when they became futures contract delivery points. The structural changes apparent in the basis data in this study cause the shorter moving averages to produce the most accurate basis forecasts in terms of mean absolute error.

Note that for preharvest hedging an alternative to last year's basis is the forward basis. Forward basis has been shown to contain a risk premium (Townsend and Brorsen 2000; Brorsen, Coombs, and Anderson 1995; Shi et al. 2004). This risk premium may be higher in times of structural change so forward basis has its own disadvantages.

The differences between the forecast accuracy of various moving average lengths was rarely statistically significant. Many studies use moving averages to obtain basis expectations (e.g. Kim, Brorsen, and Anderson 2007). The conclusion here is that it should not matter whether such studies use a five year or a three year moving average.

Although our individual models produced varied results, the general rule of thumb supported by this research is: When a location or time period does not undergo structural change longer moving averages produce optimal forecasts, but when it appears that a structural change has occurred, the previous year's basis should be used.

References

Bates, D.S. 1996. "Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options." *Review of Financial Studies* 9:69-107.

Brorsen, B.W, J. Coombs, and K. Anderson, 1995. "The Cost of Forward Contracting Wheat." *Agribusiness*, 11:349-354.

Chicago Board of Trade (CBT). 1982. "Statistical Annual: Futures Data, Part 1." Chicago Board of Trade.

Commodity Futures Trading Commission. 2000. "CFTC Approves CBT Proposal to Increase Maximum Daily Price Fluctuation Limits for its Agricultural and Precious Metal Futures and Option Contracts." Release: #4435-00. http://www.cftc.gov/opa/press00/opa4435-00.htm.

CME Group. February 8, 2008. "CME Group, KCBT and MGEX Announce Price Limit Expansion for Wheat Futures." http://files.shareholder.com/downloads/CME/609827704x0x168404/d77eca7a-cc28-4eac-80bc-fd54ba54a7bd/CME_News_2008_2_8_General.pdf

Chernov, M., A.R. Gallant, E. Ghysels, and G. Tauchen. 2003. "Alternative Models for Stock Price Dynamics." *Journal of Econometrics* 116:225-257.

Chicago Board of Trade (CBT). *Understanding Basis-The Economics of Where and When*. Chicago, 1990.

Dhuyvetter, K. 2009. Interactive Crop Basis Tool. [http://www.agmanager.info/marketing/basis/tools/default.asp]

Dhuyvetter, K. 2008. Kansas cash wheat prices. Unpublished data. Kansas State University.

Dhuyvetter, K. C., and T. L. Kastens. 1998. "Forecasting Crop Basis: Practical Alternatives." Paper presented at the NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management. Chicago, IL, April.

Dykema, A., N.L. Klein, and G. Taylor. 2002. "The Widening Corn Basis in South Dakota: Factors Affecting and the Impact of the Loan Deficiency Payment." Paper presented at WAEA annual meeting, Long Beach, California, 28-31 July.

Farm Doc. 2009. "Illinois Regional Cash Price Database." University of Illinois. http://www.farmdoc.uiuc.edu/marketing/cash/index.asp

Hatchett, R.B. 2009. "Optimal Length of Moving Average to Use When Forecasting Basis." MS thesis, Oklahoma State University.

Hauser, R.J., P. Garcia, and A.D. Tumblin. 1990. "Basis Expectations and Soybean Hedging Effectiveness." *North Central Journal of Agricultural Economics* 12:125-136.

Irwin, S.H., D.L. Good, and J. Martines-Filho. 2006. “The Performance of Agricultural Market Advisory Services in Corn and Soybeans.” *American Journal of Agricultural Economics* 88:162-181.

Irwin, S.H., P. Garcia, D.L. Good, and E.L. Kunda. 2009. “Poor Convergence Performance of CBOT Corn, Soybean and Wheat Futures Contracts: Causes and Solutions." Marketing and Outlook Research Report 2009-02, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, March [http://www.farmdoc.uiuc.edu/marketing/reports]

Jiang, B., and M. Hayenga. 1997. “Corn and Soybean Basis Behavior and Forecasting: Fundamental and Alternative Approaches.” Paper presented at the NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management. Chicago, IL, April.

Kansas City Board of Trade (KCBT). 1974-1976. “Historical Data.” Accessed at [http://www.kcbt.com/historical_data.asp] in June, 2008.

Kenyon, D.E., and S.E. Kingsley. 1973. “An Analysis of Anticipatory Short Hedging Using Predicted Harvest Basis.” *Southern Journal of Agricultural Economics* 5:199-203.

Kim, H.S., B.W. Brorsen, and K.B. Anderson. 2007. “Profit Margin Hedging.” Proceedings of the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management. Chicago, IL. [http://www.farmdoc.uiuc.edu/nccc134].

Klumpp, J.M., B.W. Brorsen, and K.B. Anderson. 2007. “Determining Returns to Storage: Does Aggregation Matter?” *Journal of Agricultural and Applied Economics* 39:571-579.

Liu, S.M., B.W. Brorsen, C.M. Oellermann, and P.L. Farris. 1994. “Forecasting the Nearby Basis of Live Cattle.” *Journal of Futures Markets* 14:259-273.

Martin, L., J.L. Groenewegen, and E. Pidgeon. 1980. “Factors Affecting Corn Basis in Southwestern Ontario.” *American Journal of Agricultural Economics* 62:107-112.

Naik, G., and R.M. Leuthold. 1991. “A Note on the Factors Affecting Corn Basis Relationships.” *Southern Journal of Agricultural Economics* 23:147-153.

O’Brien, D. 2009. “The Effects of the Micro-Market Structure for Kansas Grain Elevators on Spatial Grain Price Differentials.” Paper presented at the NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management. St. Louis, April 20.

Oklahoma Department of Agriculture, Food and Forestry. 1974-2008. “Oklahoma Market Report.”

Park, C.L. 2000. “Examining Futures Price Changes and Volatility on the Trading Day after a Limit-Lock Day.” *Journal of Futures Markets* 20:445-466.

Pebe Diaz, E.F., B.W. Brorsen, K.B. Anderson, F.G.C. Richter, and P. Kenkel. 2002. "The Effect of Rounding on the Probability of Regrading in the U.S. Peanut Industry." *Journal of Agribusiness* 20:103-116.

Peavey Grains. 1988-2008. "Peavey Catoosa Basis Bids." Catoosa, OK.

R & C Data. 1970-2008. "Financial Price Data." Marietta, GA.

Sanders, D.R., and M.R. Manfredo. 2006. "Forecasting Basis Levels in the Soybean Complex: A Comparison of Time Series Methods." *Journal of Agricultural and Applied Economics* 38:513-523.

Summers, S. VP of Marketing, KCBT. March 10, 2009. Unpublished e-mail.

Shi, W., S.H. Irwin, D.L. Good, and L.A. Hagedorn. 2004. "The Cost of Forward Contracting." American Agricultural Economics Association annual meeting, Denver, CO, August 1-4.

Taylor, M., K.C. Dhuyvetter, and T.L. Kastens. 2004. "Incorporating Current Information into Historical-Average-Based Forecasts to Improve Crop Price Basis Forecasts." Paper presented at NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management, St. Louis, MO, 19-20 April.

Townsend, J.P., and B.W. Brorsen, 2000. "The Cost of Forward Contracting Hard Red Winter Wheat." *Journal of Agricultural and Applied Economics,* 32:89-94.

U. S. Department of Agriculture. 1970-2008. "Grain and Feed Market News." Agricultural Market Service.

Wang, D., and W.G. Tomek. 2007. "Commodity Prices and Unit Root Tests." *American Journal of Agricultural Economics* 89:873-889.

Yang, S.R., and B.W. Brorsen. 1993. "Nonlinear Dynamics of Daily Cash Prices." *American Journal of Agricultural Economics* 74:706-715.