Journal of Agricultural and Resource Economics 35(1):18–33

Optimal Length of Moving Average to Forecast Futures Basis

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

The question addressed in this study is which length of historical moving average provides the best forecast of futures basis. Differences in observed forecast accuracy among the different moving averages are usually less than a cent per bushel, and most are not statistically significant. Further, the search for an optimal length of moving average may be futile since the optimal length depends on how much structural change has occurred. Our recommendation is to use moving averages when there has been no structural change and to use last year's basis or an alternative approach if the forecaster perceives that a structural change has occurred.

Key Words: basis forecast, grain, Law of One Price, moving averages, structural change

Introduction

Creating 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" [Chicago Board of Trade (CBOT), 1990, p. 23].

The most popular method of forecasting basis is historical moving averages, as illustrated by their availability on extension websites (e.g., Dhuyvetter, 2009; Farmdoc, 2009a). The attractiveness of moving averages is their ease of application. Access to local prices is generally inexpensive and readily available, allowing basis values to be localized for specific markets. Dhuyvetter and Kastens (1998) concluded that longer averages ranging from three to seven years are optimal. 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. Yet, these methods have failed recently as basis values have deviated greatly from previous levels, resulting in poor forecasts based on historical basis (Irwin et al., 2009). Given this recent failure, a need exists to reassess past recommendations about the best length of moving average to use in forecasting basis. Our study addresses 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. We argue that the optimal length of moving average will not be constant across time, form, and space. Longer moving averages

---

Robert B. Hatchett is a risk analyst at Gavilon, LLC in Omaha, Nebraska, and a former M.S. 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, Oklahoma State University; and Kim B. Anderson is Charles A. Breedlove Professor of Agribusiness and an extension economist in the Department of Agricultural Economics, Oklahoma State University. Partial funding for this research was provided by the Oklahoma Agricultural Experiment Station through Hatch Project No. OKL02170.
Review coordinated by Gary W. Brester.

are optimal when little or no structural change occurs, but shorter moving averages are optimal in response to structural changes. Basis forecasters may need to make subjective choices since it can be difficult to identify structural changes by any other means.

Theoretical Model

One of the primary reasons futures markets were created was to let market participants 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:

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

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

$$\varepsilon_t = Basis_t - Ba\hat{s}is_t, \tag{2}$$

where $Basis_t$ is the actual basis at time t, $Ba\hat{s}is_t$ 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, 2009a; Dhuyvetter, 2009). Moving average models use the simple average of the previous N years:

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

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

$$\min_N E(\varepsilon_t^2) = \min_N E\left(Basis_t - \frac{1}{N}\sum_{i=1}^{N} Basis_{t-i}\right)^2. \tag{4}$$

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

$$\min_N \sum_{t=1}^{T}\left(Basis_t - \frac{1}{N}\sum_{i=1}^{N} Basis_{t-i}\right)^2. \tag{5}$$

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

$$Basis_t \sim N\left(\mu_t, \sigma^2\right), \tag{6}$$

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) is the minimum variance forecast. Basis forecast error variance in this case is given by:

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

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. As equation (7) shows, the reduction in forecast error from using more years in the moving average dissipates quickly.

Structural changes are represented as the mean in (6) changing with time. 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 shock with a random walk process is a permanent increase in transportation costs. With a random walk, as (8) shows, the optimal forecast is with $N = 1$. In (8), all changes are permanent and represent structural changes relative to a model with a constant mean.

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

A new ethanol plant represents a structural change as it is a major source of new demand in corn markets and causes local basis levels to strengthen (McNew and Griffith, 2005). This structural change affects prices permanently, making the previous year's basis the optimal predictor for the year following the jump. After markets adjust to the shock, a new mean is reached, and longer moving averages again become optimal.

Mean-reverting models also can be used to model changes from historical basis levels (Jiang and Hayenga, 1997; Sanders and Manfredo, 2006). The basic mean-reverting model is the autoregressive moving average (ARMA). With large sample sizes, 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. ARMA models also suffer from being more difficult to estimate and to explain to producers.

ARMA models were used to forecast basis for the Illinois soybean complex by Sanders and Manfredo (2006) and Jiang and Hayenga (1997). These studies found little improvement in forecast accuracy over moving average models. We do not consider ARMA models here because there is little evidence to suggest they are more accurate than simple moving averages, because they are more difficult to use due to the need to identify a different model for each location, and compared to moving averages they are not as widely used by agricultural economics extension.

The optimal length of moving average to forecast 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, basis is the difference between two prices, and reflects differences in time, form, and space. Since time differences are held constant, structural change can reflect changes in space or form.

Previous Literature

As the difference in two prices, basis should reflect the differences in time, form, and space of the underlying commodity. Explanatory models of basis have used a wide array of variables to explain basis movements. 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. Cost of storage and transportation measures are accepted components of basis from literature that explain the transformation of prices over time and space. Differences in form, such as quality differences in local cash markets and the quality of grain reflected by futures prices, have been less often considered in past research.

Supply and demand variables at local markets can explain basis over space. Supply variables used in past research include crop production levels, a dummy variable for the presence of loan deficiency payments (LDPs), the ratio of Eastern Canadian corn production to consumption, and Western feed grain availability (Dykema, Klein, and Taylor, 2002; Martin, Groenewegen, and Pidgeon, 1980; Jiang and Hayenga, 1997). Soybean crushing levels, animal units consuming grain (corn), corn usage estimates, and export volumes have all been used as local demand variables (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). The variety of variables used suggests the factors explaining basis movements are case specific, and developing a general model of changes in basis across space is difficult. The explanatory models often use data that are not available at the time forecasts need to be made. Changes in local supply and demand could be temporary or permanent, but most of the variables in the explanatory models reflect variables that are likely mean reverting, so the explanatory models are supportive of using moving averages to forecast basis.

Several studies tested the forecasting accuracy of using moving averages of various lengths to form basis expectations. Hauser, Garcia, and Tumblin (1990) compared several naïve models of soybean basis expectations for 10 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) compared a five-year moving average, the previous year's basis, and more advanced times-series methods. Taylor, Dhuyvetter, and Kastens (2004, 2006) 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 as to which forecast performs the best. From the table, we can see that practical forecasts are about as accurate as forecasts that are more complex. The optimal length of moving average varies across studies and across commodities. These inconsistent findings suggest the need for further research to better explain why findings vary.

Data

The commodities considered are corn, soybeans, soft wheat, and hard wheat. To create basis data, nearby futures prices are subtracted from their corresponding cash prices. Two basis values are used for each year. One is selected to represent basis for a preharvest hedge and the other for a storage hedge. Taylor, Dhuyvetter, and Kastens (2006) similarly used harvest and

Table 1. Results from Previous Basis Forecasting Studies

Study	Optimal Forecasts	Conclusions
"Forecasting Crop Basis: Practical Alternatives" (Dhuyvetter and Kastens, 1998)	■ 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.
"Basis Expectations and Soybean Hedging Effectiveness" (Hauser, Garcia, and Tumblin, 1990)	■ 1- or 3-year historical basis during preharvest ■ Futures price spreads after 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, 1997)	■ 3-year moving average plus current market information best for corn ■ Seasonal autoregressive integrated moving average (ARIMA) best for soybeans	Although the 3-year moving average performs relatively well, it is outperformed 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 does not diminish. Even within closely related markets, there is no rule of thumb for producing the most accurate forecasts.

harvest plus 24 weeks to make the data set and presentation of results more manageable and because the gain in statistical precision to considering more points within a year is expected to be small.

For corn, harvest basis is the cash price in October minus the price of the December futures contract in October. October is selected because the bulk of corn harvest in Illinois is in October, and December futures are selected because this is the contract producers would use to hedge October sales. The storage basis for corn is the cash price in April minus the price of the May futures contract in April. April is selected since it is generally not worthwhile to store corn beyond April (Kim and Brorsen, 2008), and the May futures is selected since it is the nearby contract to April (May cash and July futures would be a reasonable alternative choice). For soybeans, the harvest basis is the cash price in October minus the price of the November futures contract in October, while storage basis is the cash price in April minus the price of the May futures contract in April. Basis values for soft and hard wheat are the June cash price minus the price of the July futures contract in June (harvest basis) and the cash price in November minus the price of the December futures contract in November (storage basis).

Cash and futures prices consist of second Wednesday or Thursday prices for corn, soybeans, and wheat, and when unavailable, monthly average prices are used. One disadvantage of using average monthly prices rather than the price from a single day is that they will underestimate the basis risk a hedger will actually experience—i.e., a hedger will typically sell on a single day rather than spreading sales equally over each day of a month. Another disadvantage is that it is not possible to adjust for limit days when using monthly averages. The advantage of monthly averages is that they may lead to slightly more powerful hypothesis tests.

Daily #2 corn and #1 soybean cash prices are from the Illinois Agricultural Marketing Service and reflect the mid-range of elevator bids for each region on the second Thursday of each month from 1975–2008 (FarmDoc, 2009b). 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 1974–2008. 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, 2008). Simple average monthly wheat prices are from the USDA/Agricultural Marketing Service's *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. For #2 hard red winter wheat, the data included 25 locations in Oklahoma, 16 locations in Kansas, and one in Texas. For the other commodities, the data included three locations for #2 soft red winter, one location for #1 hard red winter, seven locations for corn, and seven locations for soybeans (see Hatchett, 2009, pp. 59–60, for a summary of the locations and the time period for each location).

Futures prices reflect daily closing prices at the Chicago Board of Trade (CBOT) and Kansas City Board of Trade (KCBT) for each commodity (R & C Data, 2008), 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 CBOT 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 none of the days studied fell on a day when the futures price hit the daily limit. The daily price limits for the CBOT were 30 cents/bushel for soybeans, 10 cents/bushel for corn, and 20 cents/bushel for both soft and hard wheat as of 1982 (CBOT, 1982). These values are assumed to have remained constant in the preceding years. The CBOT and KCBT daily price limits were 10 cents/bushel prior to 1973 and were increased when prices began increasing rapidly in 1973. Price limits remained stable until March 12, 1992, when CBOT corn price limits increased from 10 cents to 12 cents/bushel, while soybean and wheat limits remained at 30 cents and 20 cents/bushel, respectively (Park, 2000). On August 14, 2000, daily price limits increased at the CBOT from 12 cents to 20 cents/bushel for corn, from 30 cents to 50 cents/bushel for soybeans, and from 20 cents to 30 cents/bushel for wheat [Commodity Futures Trading Commission (CFTC), 2000]. The KCBT limit changed when the wheat price limit was raised from 25 cents to 30 cents on October 9, 2000 (Summers, 2009). On March 28, 2008, the KCBT and CBOT both doubled the 30 cents price limit for wheat futures to 60 cents, while the CBOT also expanded limits from 50 cents to 70 cents for soybeans and 20 cents to 30 cents for corn (CME Group, 2008). None of the limit days occurred on one of the days of interest to this study.

One concern of note is that recent structural changes could greatly affect this study's findings. Hatchett (2009) investigates this possibility and concludes the findings are not fragile with respect to deleting the 2007–2008 data.

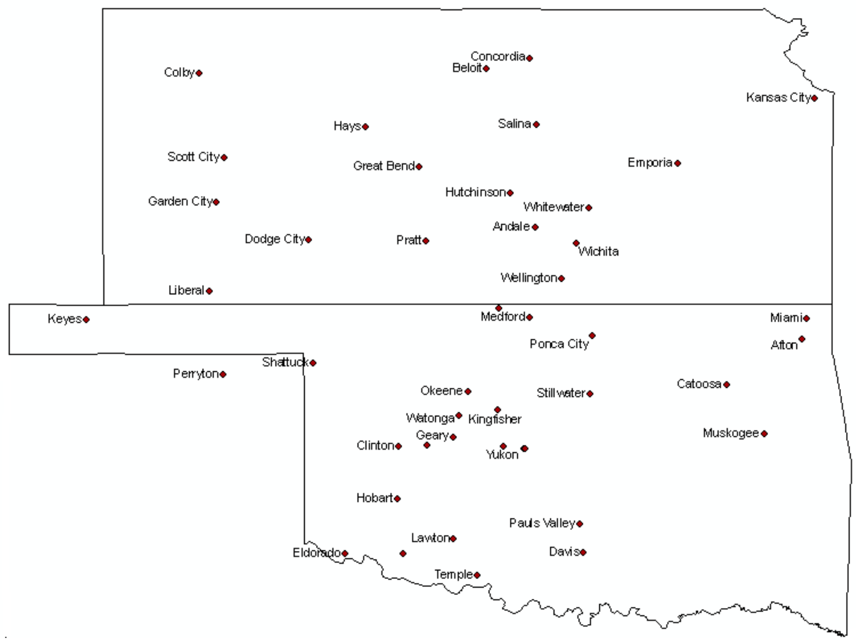

Figure 1. Kansas and Oklahoma hard red winter wheat locations studied

Procedures

Basis values were created by taking the cash market price minus the futures market price. Basis forecasts were created using equation (3), where $N = 1, \ldots, 5$. Following Dhuyvetter and Kastens (1998), we compare forecast accuracy with mean absolute error:

(9)
$$MAE = \frac{1}{T} \sum_{t=1}^{T} \left| Basis_t - Ba\hat{s}is_t \right|,$$

where the absolute value of each forecast error is averaged over the forecast period. The root mean squared error was also considered, but results do 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 computing 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 among 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 among competing forecast models could not be corrected, Dhuyvetter and Kastens conclude that a four-year moving average was more accurate than the three-year benchmark at 0.01 significance. When independence across observations is incorrectly assumed, the standard errors and their ensuing t-tests lead to greatly 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 data set and the individual commodities in this study. The pooled data set contains 15,180 out-of-sample forecast errors. Forecast errors from the first five years of each data series are not used so that an equal number of forecast errors for one-year and five-year moving averages can be generated. Thus, for the data that begin in 1974, the first year forecasted is 1979. To correct for unequal variance using random effects, interactions of indicator variables were considered—such as *period* ∗ *location* and *location* ∗ *year*, where *period* represents the preharvest or storage contract, *location* identifies the market, and *year* identifies the year of the forecast. However, these interaction terms increase the number of parameters and prevent the model from converging. As an alternative, we follow an approach analogous to Irwin, Good, and Martines-Filho (2006) and aggregate across locations and commodities, making the forecast length N the only independent variable, with time included as a random effect. Taylor, Dhuyvetter, and Kastens (2006) also aggregate across locations, but consider only paired differences. Models are estimated for individual commodities as well as the pooled model. The final mixed model is constructed as:

$$AE_{it} = \beta_0 + \sum_{j=1}^{4} \beta_j D_{ij} + v_t + \varepsilon_{it} \, , \tag{10}$$

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

Results

The results are presented first for the pooled model and then separately for each commodity. The commodities considered are #2 hard wheat, #1 hard wheat, soft wheat, corn, and soybeans.

Pooled Model

Table 2 summarizes 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 74.6% (147/194) of the locations studied. The five-year moving average produces the second most optimal forecast at 9.27%, while the two-, three-, and four-year moving averages account, respectively, for 5.67%, 4.64%, and 5.37% of the sample.

Figure 2 graphs the number of optimal forecasts produced by the previous year's basis and five-year moving average for the pooled data. The one-period forecast is usually close to the five-year forecast, but following periods of structural change like the early 1980s (inflation, collapse of land prices, oil price shocks, etc.), 1988 (U.S.-Canada free trade), and 2006 (lack of convergence at contract expiration), many more optimal forecasts occur 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.

Table 2. Number of Locations with a Given Length of Moving Average Having the Lowest Root Mean Squared Forecast Error

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

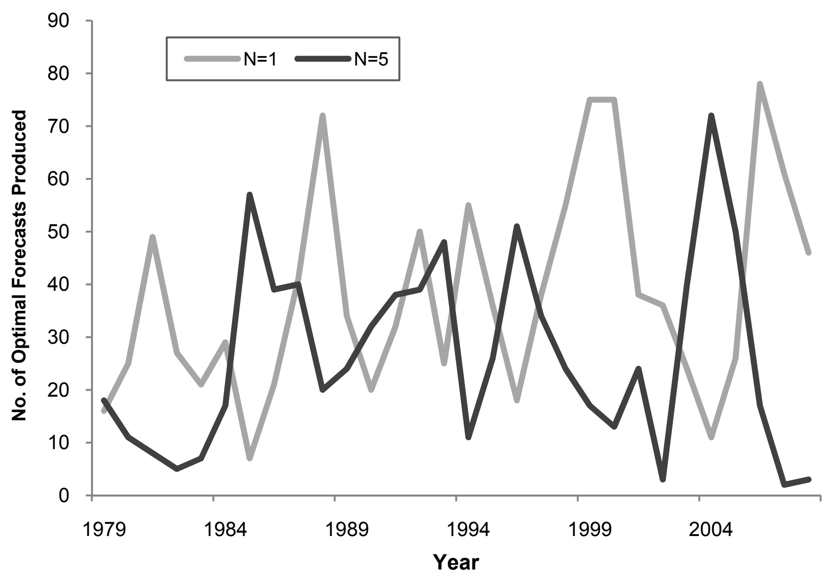

Figure 2. Number of minimum MAE forecasts produced by the previous year's basis vs. the five-year moving average, 1979–2008

Table 3 shows the absolute forecast errors from the pooled model of all the data. The F-value of 1.31 fails to reject any difference among the competing forecast methods. The intercept gives the absolute error of the five-year moving average and is 12.34 cents/bushel. Forecast accuracy increases as less historical information is used, with the previous year's basis providing the lowest pooled MAE at 11.77 cents/bushel. These results are generally within the range of the MAEs reported in previous studies. Dhuyvetter and Kastens (1998) found pooled MAEs of moving average forecasts between 10 cents and 13 cents/bushel for wheat, corn, and soybeans. The individual t-tests show the one-year forecast has significantly lower absolute error than the five-year moving average, but none of the other differences are statistically significant.

Table 3. Absolute Error (cents/bushel) of Basis Forecasts as a Function of Number of Years in the Moving Average, Pooled Data

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.

Table 4. Absolute Error (cents/bushel) of Hard Wheat Basis Forecasts as a Function of Number of Years in the Moving Average

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.

#2 Hard Wheat

Preharvest and storage hard wheat basis forecasting model accuracies are shown in table 4. For the preharvest forecasts, the two-year moving average has a significantly higher forecast error than the five-year benchmark. The only preharvest model to produce a lower MAE than the benchmark is the four-year moving average, which improves by only 0.06 cents/bushel. Over the sample period, any of the five preharvest models considered would result in a forecast error of approximately 13 cents/bushel, 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 five-year benchmark are rejected for all but the four-year moving average. The previous year's basis lowers the benchmark MAE from 13.03 cents/bushel to 11.09 cents/bushel. The improvement in accuracy as the historical period shortens supports using shorter moving averages to forecast hard wheat storage basis.

Table 5. Absolute Error (cents/bushel) of Kansas City Ordinary Protein, #1 Hard Wheat Basis Forecasts as a Function of Number of Years in the Moving Average

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 4 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 for 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.

#1 Hard Wheat

The Kansas City price data allow this study to compare the differences in forecasting both the regular protein #1 hard red wheat, and 13% protein #1 hard red wheat. Table 5 shows the model results for Kansas City ordinary protein #1 hard wheat basis models, while table 6 reports the results for 13% protein #1 hard wheat. The difference between the two prices must be a function of form since time and space are the same. The benchmark intercept for the 13% protein model is 3.22 cents/bushel higher than the ordinary protein forecast model of preharvest basis. The MAE of the best preharvest forecast for 13% protein (N = 3 for the preharvest model in table 6) is still 1.22 cents/bushel more than with the worst ordinary protein forecast model (N = 1 for the preharvest model in table 5).

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. Basis forecast errors are higher for 13% protein wheat because the futures contract reflects the price of ordinary protein wheat.

Soft Wheat

Table 7 displays the results for the soft wheat basis forecasting models. Using the previous year's basis to predict soft wheat preharvest basis gave an average forecast error of 25.95 cents/bushel, while the most accurate method, the two-year moving average, had an MAE of 23.42.

Table 6. Absolute Error (cents/bushel) of Kansas City 13% Protein, #1 Hard Wheat Basis Forecasts as a Function of Number of Years in the Moving Average

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.

Table 7. Absolute Error (cents/bushel) of Soft Wheat Basis Forecasts as a Function of Number of Years in the Moving Average

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.

The storage basis is forecasted considerably more accurately than the preharvest basis. But as with most other commodities, there are no significant differences between the forecasting accuracy of the various lengths of moving averages.

Corn

Table 8 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 five-year

Table 8. Absolute Error (cents/bushel) of Corn Basis Forecasts as a Function of Number of Years in the Moving Average

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

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

Table 9. Absolute Error (cents/bushel) of Soybean Basis Forecasts Based on Moving Averages

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

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

benchmark. 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 vary by moving average length for corn storage basis. Significant differences from the five-year benchmark exist in every model except the four-year moving average at a 0.05 level. This result suggests that shorter moving averages can outperform the five-year moving average at forecasting the corn storage basis. The best model, using the previous year's basis, lowers MAE from the five-year moving average of 7.49 cents/bushel to 6.32 cents/bushel.

Soybeans

Table 9 shows the absolute error of the Illinois soybean basis forecasting models. The preharvest five-year benchmark MAE is 11.23 cents/bushel and can be improved by all of the shorter moving-average models. The most improvement comes from the two-year moving average, which lowers the MAE to 10.62 cents/bushel. Although the benchmark can be improved upon, the improvement is not statistically significant based on the *t*-test. The narrow range (< 0.61 cents/bushel) of MAEs shows little difference across preharvest basis models over the period studied.

The choice of moving average length affects the accuracy of Illinois soybean storage basis forecasts. While all of the shorter moving average models outperform the benchmark, the previous year's basis and the two-year moving average result in 1.98 cents and 1.16 cents/bushel lower forecasts, respectively. As shown by the smaller intercept for soybean storage basis in table 9, the storage basis forecasts have lower MAEs than do preharvest forecasts.

Summary and Conclusions

The most popular method of forecasting basis is historical moving averages. This research reassesses past recommendations about the best length of moving average to use in forecasting basis. Our study employs 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. The forecasting methods considered were last year's basis and moving averages of the previous two to five years. Hypothesis tests about differences in mean absolute error were conducted using aggregated data following Irwin, Good, and Martines-Filho (2006). The regressions used mean absolute error as the dependent variable, with moving average length as the independent variable, and time as a random effect.

This research finds the size of most MAEs to be consistent with previous studies (Dhuyvetter and Kastens, 1998; Taylor, Dhuyvetter, and Kastens, 2004, 2006). These values are generally between 7 cents and 17 cents/bushel, except errors are larger for soft wheat. Forecast errors are larger for harvest basis than for storage basis.

The optimal forecast length is generally shorter than previous recommendations. Using a four-year moving average produces 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. The optimal amount of historical data included in corn and soybean forecasts has 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 cause the shorter moving averages to produce the most accurate basis forecasts. Note that for preharvest hedging, an alternative to using last year's basis when a structural change has occurred is the forward basis. Forward basis contains 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. For postharvest hedging, the harvest basis provides an alternative source of information (Working, 1953; Taylor, Dhuyvetter, and Kastens, 2004, 2006). Further, users may want to adjust forecasts based on local supply and demand information. These adjustments may be done subjectively or perhaps more formally using regression analysis. Structural changes will need to be identified subjectively. While making subjective choices may not be satisfying to those seeking a mechanical rule, it is likely what people are already doing.

The differences between the forecast accuracy of various moving average lengths are rarely statistically significant. When they are, the differences usually are not more than a cent or two per bushel. Consequently, the selected length of moving average to use for basis forecasting may not matter. Nonetheless, many studies and marketers use moving averages to obtain basis expectations (e.g., Kim, Brorsen, and Anderson, 2010). Such studies should be equally valid whether a five-year or a three-year moving average is used. 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 a structural change has occurred, the previous year's basis or an alternative approach should be used.

[*Received October 2009; final revision received January 2010.*]

References

Bates, D. S. "Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options." *Rev. Finan. Stud.* 9(1996):69–107.

Brorsen, B. W., J. Coombs, and K. B. Anderson. "The Cost of Forward Contracting Wheat." *Agribus.: An Internat. J.* 11(1995):349–354.

Chernov, M., A. R. Gallant, E. Ghysels, and G. Tauchen. "Alternative Models for Stock Price Dynamics." *J. Econometrics* 116(2003):225–257.

Chicago Board of Trade (CBOT). *Statistical Annual: Futures Data, Part 1*. Chicago, IL: CBOT, 1982.

———. *Understanding Basis—The Economics of Where and When*. Chicago, IL: CBOT, 1990.

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

Commodity Futures Trading Commission. "CFTC Approves CBT Proposal to Increase Maximum Daily Price Fluctuation Limits for Its Agricultural and Precious Metal Futures and Option Contracts." Release No. 4435-00, 2000. Online. Available at http://www.cftc.gov/opa/press00/opa4435-00.htm.

Dhuyvetter, K. C. Kansas cash wheat prices. Unpub. data, Dept. of Agr. Econ., Kansas State University, 2008.

———. "Interactive Crop Basis Tool." Online. Available at http://www.agmanager.info/marketing/basis/tools/default.asp. [Accessed November 2009.]

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

Dykema, A., N. L. Klein, and G. Taylor. "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, CA, 28–31 July 2002.

Farmdoc. "Illinois Regional Basis Decision Tool." University of Illinois at Urbana-Champaign, 2009a. Online. Available at http://www.farmdoc.illinois.edu/marketing/index.asp.

———. "Illinois Regional Cash Price Database." University of Illinois at Urbana-Champaign, 2009b. Online. Available at http://www.farmdoc.uiuc.edu/marketing/cash/index.asp.

Hatchett, R. B. "Optimal Length of Moving Average to Use When Forecasting Basis." Unpub. M.S. thesis, Dept. of Agr. Econ., Oklahoma State University, 2009.

Hauser, R. J., P. Garcia, and A. D. Tumblin. "Basis Expectations and Soybean Hedging Effectiveness." *N. Cent. J. Agr. Econ.* 12(1990):125–136.

Irwin, S. H., P. Garcia, D. L. Good, and E. L. Kunda. "Poor Convergence Performance of CBT Corn, Soybean, and Wheat Futures Contracts: Causes and Solutions." Marketing and Outlook Res. Rep. No. 2009-02, Dept. of Agr. and Consumer Econ., University of Illinois at Urbana-Champaign, March 2009. Online. Available at http://www.farmdoc.uiuc.edu/marketing/reports.

Irwin, S. H., D. L. Good, and J. Martines-Filho. "The Performance of Agricultural Market Advisory Services in Corn and Soybeans." *Amer. J. Agr. Econ.* 88(2006):162–181.

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

Kim, H. S., and B. W. Brorsen. "Can Real Option Value Explain Why Producers Appear to Store Too Long?" In Proceedings of the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management, St. Louis, MO, 2008. Online. Available at http://ageconsearch.umn.edu/bitstream/37602/2/confp06-08.pdf.

Kim, H. S., B. W. Brorsen, and K. B. Anderson. "Profit Margin Hedging." *Amer. J. Agr. Econ.* 92(2010): forthcoming.

Liu, S. M., B. W. Brorsen, C. M. Oellermann, and P. L. Farris. "Forecasting the Nearby Basis of Live Cattle." *J. Futures Mkts.* 14(1994):259–273.

Martin, L., J. L. Groenewegen, and E. Pidgeon. "Factors Affecting Corn Basis in Southwestern Ontario." *Amer. J. Agr. Econ.* 62(1980):107–112.

McNew, K., and D. Griffith. "Measuring the Impact of Ethanol Plants on Local Grain Prices." *Rev. Agr. Econ.* 27(2005):164–180.

Naik, G., and R. M. Leuthold. "A Note on the Factors Affecting Corn Basis Relationships." *S. J. Agr. Econ.* 23(1991):147–153.

Oklahoma Department of Agriculture, Food, and Forestry. *Oklahoma Market Report*. Various issues, 1974–2008.

Park, C. L. "Examining Futures Price Changes and Volatility on the Trading Day After a Limit-Lock Day." *J. Futures Mkts.* 20(2000):445–466.

Peavey Grain. "Peavey Catoosa Basis Bids." Catoosa, OK, 1988–2008.

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

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

Sanders, D. R., and M. R. Manfredo. "Forecasting Basis Levels in the Soybean Complex: A Comparison of Time-Series Methods." *J. Agr. and Appl. Econ.* 38(2006):513–523.

Shi, W., S. H. Irwin, D. L. Good, and L. A. Hagedorn. "The Cost of Forward Contracting." Paper presented at AAEA annual meeting, Denver, CO, 1–4 August 2004.

Summers, S. Vice President of Marketing, Kansas City Board of Trade. Unpub. e-mail, 10 March 2009.

Taylor, M., K. C. Dhuyvetter, and T. L. Kastens. "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 2004.

———. "Forecasting Crop Basis Using Historical Averages Supplemented with Current Market Information." *J. Agr. and Resour. Econ.* 31(2006):549–567.

Townsend, J. P., and B. W. Brorsen. "The Cost of Forward Contracting Hard Red Winter Wheat." *J. Agr. and Appl. Econ.* 32(2000):89–94.

U.S. Department of Agriculture. *Grain and Feed Market News.* USDA/Agricultural Marketing Service, Washington, DC. Various issues, 1970–2008.

Working, H. "Hedging Reconsidered." *J. Farm Econ.* 35(1953):544–561.

Yang, S. R., and B. W. Brorsen. "Nonlinear Dynamics of Daily Cash Prices." *Amer. J. Agr. Econ.* 74(1993): 706–715.