Corn and Soybean Basis Behavior and Forecasting: Fundamental and Alternative Approaches

by

Bingrong Jiang and Marvin Hayenga

Suggested citation format:

Jiang, B., and M. Hayenga. 1997. "Corn and Soybean Basis Behavior and Forecasting: Fundamental and Alternative Approaches." Proceedings of the NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management. Chicago, IL. [http://www.farmdoc.uiuc.edu/nccc134].

Corn and Soybean Basis Behavior and Forecasting: Fundamental and Alternative Approaches

Bingrong Jiang and Marvin Hayenga*

This basis study covers corn and soybean markets across the U.S. Corn and soybean bases have seasonal patterns, as does the relative importance of factors (storage costs, barge rates, production levels) determining the basis. Corn and soybean basis behavior in port locations is different than in major production areas. Though three-year-average basis forecasts are reasonably accurate in recent years, forecasts based on a three-year-average supplemented with additional fundamental variables or seasonal ARIMA model forecasts slightly improved basis forecasting accuracy in out of sample tests.

Introduction

The basis is defined as the difference between cash and futures prices. In grain merchandising, the basis is usually defined as the difference between the local cash price and the nearby futures price, i.e. the current price of the nearest futures delivery contract. It has been argued by many researchers that the key to successful hedging is understanding the basis (Garcia and Good, 1983; Hieronymous, 1978; Leuthold et. al., 1989; Karlson et. al.., 1993; Tomek, 1996; and so on). This is because most hedging involves two opposite positions: one in the cash market and another in the futures market. It is the difference between cash and futures prices, together with the futures price, that determines the return from hedging and hedging's effectiveness in reducing risk.

Grain merchandisers and processors routinely need to accurately forecast basis to offer forward purchase or sales contracts. Corn and soybean producers need to know the basis to evaluate contracts offered to them, or in making hedging decisions. The Chicago Board of Trade (CBOT,1990) asserts "Without a knowledge of the usual basis and basis patterns for your particular community, it is impossible to make fully informed decisions about, for example, whether to accept or reject a given price; whether and when to store your crop; whether, when, and in what delivery month to hedge; when to close (or 'lift') a hedge; or when and how to turn an unusual basis situation into a possible profit opportunity."(p.23)

---

* Authors are Graduate Research Assistant and Professor of Economics, respectively, at the Department of Economics, Iowa State University, Ames, IA 50011

Though basis is extremely important, there have been only few basis behavior studies published, and even fewer basis forecasting studies (with the exception of forecasts using simple moving averages of historical basis). The objective of this study to investigate corn and soybean basis behavior and to improve the accuracy of basis forecasts.

This paper first reviews the theory of storage and basis studies on grain. The theory of storage is related closely to the temporal and spatial price relationships in the grain market. The basis of storable commodities is actually a temporal price relationship, the difference between the current spot price and futures price. Local basis has another component of spatial price relationship, which is the difference between a local cash price and the cash price at Chicago Board of Trade delivery points. The previous basis studies on grain are reviewed to assess the current state of knowledge, and identify where further contributions could be useful.

Since basis patterns differ from location to location, what is important to hedgers is the basis at their locations. To provide some diversity in locations and to better represent the U.S. corn and soybean market, several local markets are studied. These markets include Chicago, St. Louis, Toledo, Gulf Coast, NE Iowa, Central Illinois, Richmond, and the Pacific Northwest market (corn only). Seven markets for soybeans and eight markets for corn are analyzed in this study.

Several approaches are utilized in this study to explain and subsequently forecast the local grain basis at these markets. A fundamental structural model incorporating storage cost, transportation costs, and regional supply and demand variables is developed to explain basis behavior. Several forecasting techniques are used in forecasting corn and soybean basis. These include including traditional methods such as a simple three-year-average forecasts, structural econometric model, modified three-year average model, artificial neural networks, time series methods such as seasonal ARIMA and state space models, and composite forecasts. The ability of the structural model to explain past basis behavior is examined, and out-of-sample forecast performance of these alternative basis forecasting approaches is evaluated.

Theory of Storage

The efficient market hypothesis states that prices reflect information to the point where the marginal benefits of acting on information do not exceed the marginal costs. Correspondingly, the theory of storage suggests that basis--the difference between contemporaneous spot and futures prices--should equal the cost of storage. Otherwise, there will opportunities for profitable arbitrage between the spot and futures markets. This suggests that basis has a predictable temporal pattern. As the cost of storage decreases as one gets closer to delivery (maturity), the cash price will gain in relation to futures price. On the first day of delivery at the par delivery point, cash and futures prices should be equal theoretically, and the basis should be zero (but the basis at non-par delivery points is usually not zero because of transportation cost differences). Similarly, the price spread of two futures contracts within a given crop year, at least when it is positive will not exceed cost of storage between the two delivery months. These are the clear aspects of the theory of price of storage as stated by Working (1949).

Empirical studies have found that it is not uncommon to have the basis to be less than the full cost of storage, and even negative returns for storage (spot price is higher than futures price) are possible. Several studies have been offered to explain this discrepancy between theory and empirical evidence. The popular explanations for the failure of the theory are convenience yield and risk premium. Supporters of convenience yield include these classical papers: Kaldor (1939, 1940), Working (1948, 1949); and Telser (1958). The risk premium literature can trace back to Keynes (1930). The most recent studies show that theory should not have been rejected, because the analyst's misconception or mis-measurement of the concept led to its rejection (Wright and Williams 1989; Brennan et al. 1997; and Benirschka and Binkley 1995).

Grain Basis Studies

Besides various studies dealing with the theory of storage, there have been few empirical studies related to the basis behavior and basis forecasting of grain. Heifner (1966) set up a prediction equation for both cash price change and basis change, which stated that the change variables over a particular interval are a function of the basis at the beginning of the interval. Kenyon and Kingsley (1973) predicted the harvest time basis at planting time and compare the performance of this harvest time forecasted basis with other historical average basis estimates in the hedging effectiveness. Martin, Groenewegen and Pidgeon (1980) modeled the factors affecting corn basis in Southwestern Ontario over crop years from 1962 to 1976. Kahl (1982) focused on the change corn basis patterns from the sixties to seventies in Chicago. Garcia and Good (1983), based on the theory of carrying charge, analyzed the factors influencing the Illinois corn basis for the period from 1971 to 1981. Powers and Johnson (1983) studied Wisconsin corn basis during the storage season from 1978-1980. Taylor and Tomek (1984) developed a simple model to forecast the November corn basis in Batavia, New York. Brorsen et al. (1985) investigated dynamic price relationships of corn, sorghum, and soybeans in different locations. Both price relationships across space and among commodities in the same location were studied. Kahl and Curtis (1986) analyzed various factors that influence the magnitude and variation of a grain surplus area (Illinois) and a grain deficit area (North Carolina). Hauser, Garcia and Tumblin (1990) evaluated alternative soybean basis expectation (forecast) models and how basis expectations play a role in measuring hedge effectiveness. Thompson, Eales and Hauser (1990) investigated the relationship of spatial grain (corn and soybean) basis changes between cash market locations in the North Central region and explained how cash price is linked to other cash market prices and futures price. Naik and Leuthold (1991) empirically tested the components of the corn basis using cash prices for East Central Illinois elevators and CBOT futures prices. Karlson, Anderson and Dahl (1993) analyzed the role of futures markets in corn marketing decisions. In a recent paper, Tomek (1996) outlined a simple model of price level and basis behavior based on supply and demand of storage.

Structural Basis Behavior Model

Basis at a location involves two price relationships: (1) delivery point cash and futures price relationship, and (2) local and the delivery point cash prices relationship. The first component is a temporal price difference which, according to the theory of storage, should equal to storage return (or price of storage). The second component represents spatial price difference. The law of one price suggests that it will be equal to transportation cost between two locations. That is, the basis can be viewed as the sum of storage return and transportation cost. The price of storage (par delivery point basis) is determined by both the supply of storage and the demand for storage. Tomek's (1996) supply of storage equation expresses the price of storage is a function of opportunity cost, direct storage cost and convenience yield, and the demand for storage equation shows inventory is a function of relative demand for consumption over two periods, production and the price of storage. A reduced form for the price of storage, derived from these two equations, shows that the price for storage is a function of production, opportunity cost, relative demand in two periods, direct storage cost, and convenience yield. Given the transportation cost, basis, the sum of storage and transportation costs, is affected conceptually by storage and transportation costs, production and stocks, and local economic conditions such as local grain consumption, and constraints of storage and transportation capacities.

All previous basis behavior studies (Martin et al. 1980; Kahl 1982; Garcia and Good 1983; Powers and Johnson 1983; Kahl and Curtis 1986) used models based on this general approach, though the actual data used in estimation were quite different. These studies were conducted in early 80s', and they were confined to one or two markets. They studied basis in Ontario, Chicago, Illinois, Wisconsin, and Illinois and North Carolina, respectively. There have been significant changes in grain marketing since then.

According to the law of one price, one major component in spatial price difference is transportation cost, yet this is not well captured in previous studies. Only Garcia and Good (1983) used barge rates on the Mississippi River. Powers and Johnson (1983) used a trend variable in the place of transportation cost, Kahl and Curtis (1986) took U.S. rail rate index as the transportation cost variable, and Martin et al. (1980) did not include a transportation cost variable. Four out of the five fundamental basis models discussed above used a single equation approach, and OLS in estimating the basis equation. Grain markets may be well integrated, and cash prices in different locations may interact with each other. A system of equations approach may capture market behavior better. Kahl and Curtis (1986) found that Seemingly Unrelated Regression outperformed Ordinary Least Squares.

Corn and soybean nearby basis behavior of several grain markets across the U.S. will be studied in this paper, taking into account the regional price interrelationships. This basis behavior study can also provide more updated study on corn and soybean basis, as basis patterns have changed dramatically since most of the previous studies were done in the 1980s.

Empirical model

The basis behavior model is estimated contract by contract for both corn and soybean. The following equation is an example of the empirical model being estimated for Northeast Iowa corn:

(EQ. 1)
$$BS\_nia_j = \alpha_j + \beta_{1j}*RP\_nia_j + \beta_{2j}*BR\_nia_j + \beta_{3j}*CN\_nia_j + \beta_{4j}*EX\_nia_j + \beta_{5j}*AUC\_G_j + \beta_{6j}*TTM_j,$$
j = Mar., May, Jul., Sep., and Dec.

Table 1. Variable Definition

Variable	Definition
BS_nia	Nearby corn basis of Northeast Iowa
RP_ nia	Storage cost of Northeast Iowa (Prime interest rate * Northeast Iowa corn cash price)
BR_ nia	Barge rate for Northeast Iowa, St. Louis barge rate is used
CN_ nia	Ratio of corn production and storage capacity for the states around Iowa
EX_ nia	Export volume through Mississippi River
AUC_G	Animal units consuming grain
TTM	Time to maturity, measured by months before futures contract expires

The study utilizes monthly time series data from January, 1980 to December, 1995.[1] Futures prices for corn and soybeans from January, 1980 to April, 1996 were generously provided by Chicago Board of Trade. The settlement prices are selected as the futures prices. The nearby future prices (NBFPs) are compiled from the futures prices of all the contracts. For any month (either delivery or non-delivery month), the current futures prices of next closest futures contract are taken as this month's nearby futures prices. Monthly corn and soybean cash prices for various market locations are all from the USDA. Other series that are obtained from USDA include barge rates for shipping points on the Mississippi, Illinois and Ohio Rivers, corn and soybean monthly export volumes by ports, and animal units consuming grain. Prime interest rate is the prime rate charged by banks on short-term business loans from the Annual Statistical Digest and Federal Reserve Bulletin. Soybean monthly crush volumes[2] are from U.S. Bureau of Census publication, Fats and Oil.

[1] Since weekly cash price time series are not available for some locations or back to the 1980s, therefore monthly data is used in this study.

[2] It is for soybean basis model

A stacked model, which stacks all contract models together (with appropriate slope and intercept dummy variables for each contract or each month), is used to test for seasonality and to test whether SUR can improve the model fit. The nested F-test statistics reject the null hypothesis that seasonal dummies, either contract or month, are zero in most of the market equations. Similar results are found for the soybean equations. This suggests that both corn and soybean basis have seasonal pattern. Though the stacked model with seasonal dummies can capture the seasonal basis patterns, estimation by contract or month[3] will be more appropriate. This is because the Goldfeld-Quandt tests show that variances of the contract models are not all equal. Contract by contract modeling results in a higher R-Squares and lower RMSEs on average. Therefore, the basis behavior model is estimated contract by contract for both corn and soybean models. There are five corn basis models for each of the eight markets and six soybean basis contract models for each of the seven selected markets.[4]

Estimation Results

Corn contract models

The model explains about 50-80% of the corn basis variation There are seasonal variations in size and significance of estimated coefficients throughout the year. Storage costs are usually negative and significant only in the early storage season (for contracts March and May), except in Northeast IA, Central IL and Gulf Port where storage costs are also marginally significant (at 10% level) for July and September contract models (see Table 2). Barge rates are mostly significant in the months of May, June, September, October and November. That makes sense because the upper Mississippi River is closed in some of the winter months due to ice and in few summer months with low water level or flooding. Corn production relative to storage capacity is an important factor only for the months right after harvest (from December to February), and it does not affect corn basis significantly after that.

Corn exports only have significant coefficients in two port markets: Pacific NW and Richmond. Animal Units Consuming Grain is more important in Pacific NW than other markets. December contract models have the most significant AUC_G coefficients. Time to maturity matters from December to February for all eight markets.

Soybean contract models

The mean R^2 of six soybean contract models ranges from 0.5 to 0.8. There are also noticeable differences in significance and magnitude of the coefficients for the same variables across the soybean contracts. Storage costs affect the soybean basis in every market right after harvest season (November and December), and does not influence the basis in May and June when the demand for storage is much less (see Table 3). Barge rate coefficients are

[3] The number of observation is a problem for estimating the basis model by month, therefore, we only focus on contract model in this study.

[4] The soybean futures contract August and September are combined into a single contract model since each of them covers one month for nearby basis and there in not enough observation for a good estimation.

significantly different from zero mostly in May, July and November contract models. Most of the significant production coefficients are in the March and May contract models, which suggests that the size of the new crop impacts soybean basis from January to April.

The demand factors results are mixed. Soybean crushing generally has coefficients that are inconsistent with expectation. Animal units consuming grain, not significant in general, has both positive and negative significant coefficients. Export volume does not affect soybean basis significantly throughout the year. Time to maturity variables are significant in the January contract model for all markets (except Northeast Iowa) and in other contract models for some markets.

Table 2. Number of significant coefficients with expected sign in each contract model for eight corn markets

	Contracts				
Variables	March	May	July	September	December
	Total number of significant coefficients				
Storage cost	7	5	3	3	1
Transport cost		1	7	2	6
Production	7	1		2	
Export	1		1		
AUC_G	1		1	1	4
TTM	8	2			
	Number of significant coefficient at different significance levels				
Storage cost	6,6,7	4,5,5	0,0,3	0,0,3	0,0,1
Transport cost		0,0,1	3,6,7	1,1,2	6,6,6
Production	3,7,7	0,0,1		0,1,2	
Export	0,1,1		0,1,1		
AUC_G	1,1,1		1,1,1	1,1,1	2,4,4
TTM	8,8,8	0,2,2			

Note: The three numbers represents the number of coefficients significant at 1%, 5% and 10% significance levels, respectively.

Basis Forecasting Models

Several forecast techniques were utilized in prior studies: a basis change model[5] (Heifner 1966; and Kenyon and Kingsley 1973), which considers that basis change over a time interval as a function of initial basis; a structural econometric forecasting model (Taylor and Tomek 1984; and Strobl etc. 1996); a seasonal ARIMA technique (Strobl etc. 1996); and a

[5] This model, also called basis convergence model, is not used in this paper because it is not appropriate to analyze basis convergence over the nearby basis.

simple naive forecast model (Hauser et al. 1990). The naive model basis forecast has typically been the previous year's basis, the average of last three years' basis, or the basis is forecast as a function of price spread between the two nearest futures contracts or the time to expiration of the futures contract (Hauser et al. 1990). Since the naive model based on previous three years' average basis is very popular, it will serve as the standard for comparison for the more sophisticated forecast models outlined below.

Table 3. Number of significant coefficients with expected sign in each contract model for seven soybean markets

	Contracts					
Variables	January	March	May	July	Aug&Sep	November
	Total number of significant coefficients					
Storage cost	7	5	2		3	4
Transport cost	1		4	6	3	3
Production	1	4	4		2	
Export						
Crushing			2	2		1
AUC_G	1	1				
TTM	6	2	2	1		2
	Number of significant coefficient at different significance levels					
Storage cost	5,6,7	5,5,5	1,2,2		0,2,3	4,4,4
Transport cost	0,0,1		2,3,4	6,6,6	2,2,3	1,3,3
Production	0,0,1	0,1,4	2,4,4		1,2,2	
Export						
Crushing			1,2,2	1,2,2		1,1,1
AUC_G	0,0,1	0,1,1				
TTM	6,6,6	2,2,2	0,2,2	0,1,1		0,1,2

Note: The three numbers represents the number of coefficients significant at 1%, 5% and 10% significance levels, respectively.

Three-year-average plus model

If you assume that basis is seasonal and stable over years, the last three year's basis for a particular month will be a good forecast of basis for the same month this year. Though it is easy to use, this approach does not take into account any current market information. Therefore, it is hypothesized that this naive model forecast could be improved if current market information is added to that model. That is the three-year-average plus model:

(EQ.2) $$BS_t = f(3AVG_t, Total\ Supply_t, Exports_t, DM_i), \quad i = 1, \ldots, 9, 11, \text{ and } 12,$$

where *3AVG* is average basis of previous three years at the same month, DM_i are monthly dummies variables from January through December (except October), *Total Supply* and *Exports* are USDA supply and export estimates.

Fundamental forecast model

The traditional approach is to forecast the independent variables individually and then insert the forecasted independent variables into the estimated model to derive the basis forecast, EQ.3 is an example for Northeast Iowa corn basis forecasting model:

$$BS\_nia_j^f = \alpha_j + \beta_{1j}*RP\_nia_j^{\,f} + \beta_{2j}*BR\_nia_j^{\,f} + \beta_{3j}*CN\_nia_j^{\,f} + \beta_{4j}*EX\_nia_j^{\,f} + \beta_{5j}*AUC\_G_j^{\,f} + \beta_{6j}*TTM_j.^{\,f}, \tag{EQ.3}$$

$_j$ = Mar., May, Jul., Sep., and Dec.

where *i* represents market, the superscript ^{f} stand for forecasted value, and the Greek characters are estimated coefficients from the behavior model. Appropriate forecast models (usually simple time series techniques such as moving average model) for these independent variables have to be determined first, or forecasts available at the time such as USDA production projections can be used as the forecasted values (see Table 4). These are the ancillary forecasts which usually introduce additional errors into the forecasts.

Table 4. Methods of ancillary forecasting

Variables	Methods used to obtain ancillary forecasts
Prime interest rate	Naive model (equals t-1 PIR)
Barge rates	ARIMA model
Production	Regression model and USDA WASDE (the relationship between state and US productions is identified by simple regression, then the coefficients from the regression and WASDE of US production are used to calculate the state corn and soybean forecasts
Storage capacity	Naive model (equals to last year's capacity)
Soybean crushing	Previous three year average
AUC_G	It is known for the same crop year and naive forecast for next crop year
Exports	Previous three year average
Cash prices	Previous three year average

Seasonal ARIMA model

Though the fundamental approach to basis forecasting has appeal, ARIMA time series methods may produce better forecasts. Theoretically, time series techniques have some advantages over econometrics modeling (Jenkins 1979). Others, such as Strobl et al. (1996), cited practical reasons for preferring time series forecasting methods over econometrics forecasting model. The general form of the seasonal ARIMA model, SARIMA(p,d,q)(P,D,Q)$_s$, is:

(EQ.4) $$\Phi_P(B^s)\phi_p(B)(1-B)^d(1-B)^D BS_t = \Theta_Q(B^s)\theta_q(B)a_t$$

where B is the backshift notation, s is the seasonal period (observations per period, here it is 12 months), the (p,d,q) are orders of autoregressive, differencing and moving average terms, respectively, and (P,D,Q) are the orders of seasonal counterparts corresponding to (p,d,q). The Greeks, ϕ, θ, Φ, and Θ, are the coefficients for regular autoregressive terms, moving average terms, seasonal regular autoregressive terms, and seasonal moving average terms, respectively.

State Space Model

In this paper, the approach proposed by Akaike (1976) will be used. One reason for this selection is that a procedure called STATESPACE, which follows Akaike's method, is readily available in the SAS program. This approach utilizes Kalman filters to compute the optimal estimates. It also allows for the maximum likelihood estimation of the unknown parameters in the model, which is done by prediction error decomposition.

The general state space form (SSF) applies to a multivariate time series Y_t.

(EQ.5)
$$Y_t = G_t X_t + W_t + d_t \quad \text{(Observation or measurement equation)}$$
$$X_{t+1} = F_t X_t + V_t + c_t \quad \text{(State or transition equation)}$$

where all the variables are in matrices or vector forms. Y_t, W_t, d_t are w x 1 vectors; X_t, V_t, c_t are v x 1 vectors; G_t is a matrix with dimension of w x v and F_t is v x v. W_t, and V_t have means of zero and variances of R_t, and Q_t, respectively, and covariance of S_t. W_t, d_t, V_t, c_t, R_t, Q_t, S_t are system matrices. If these matrices do not depend on time, the model is said to be time-invariant or time-homogeneous.

Artificial Neural Networks (ANN or NN)

Artificial neural networks have been applied in various scientific fields with successes. However, it is not until recently that it began to be applied in financial and economic studies. Several authors tried to relate neural networks to standard statistical approaches (Azoff 1994; Cheng and Titterington 1994; Kohzadi et al. 1995; Hill et al. 1994). Others have compared its performance with statistical approaches (Hill et al. 1994; Kohzadi et al. 1995; Dasgupta et al. 1994; Kuan and Liu 1995; Kaastra and Boyd 1995; Grundnitski and Osburn 1993; Uhrig et al. 1992; Hamm et al. 1993; Claussen and Uhrig 1994; Kohzadi et al. 94).

The ANN could be an appropriate model to forecast local grain basis. The true model generating the observed price series is uncertain. The data driven property of the ANN could find the hidden patterns in the series. The interactions among several markets across the country and between cash and futures markets are probably nonlinear in nature, which ANN can handle with its nonlinear mapping. As Kohzadi et al. (1995) stated, many price series were found to be non-random and nonlinear, not the typically assumed random and linear.

Forecast Performance Comparison

All the forecasts by various forecasting models discussed above are compared to forecasts of the simple bench mark 3-year-average model, and to each. Four different criteria measuring the accuracy of forecasts are used[6]: mean absolute error (MAE), root mean squared errors (RMSE), two Theil's U statistics (or Theil's inequality coefficients. Mean absolute error (MAE) calculates the average of absolute values of the forecast errors, root mean square error (RMSE) is the square of the average if the squared values of forecasts errors, and the U statistics are defined as the square root of ratio of the mean square error of the predicted (percentage) change to the average squared actual (percentage) change. The formulas for these criteria are:

$$MAE = \frac{1}{n^0}\Sigma_i |A_i - P_i|$$

$$RMSE = \sqrt{\frac{1}{n^0}\Sigma_i (A_i - P_i)^2}$$

$$U_\Delta = \sqrt{\frac{(1/n^0)\Sigma_i(\Delta A_i - \Delta P_i)^2}{(1/n^0)\Sigma_i \Delta A_i^2}}$$

where P and A represent predicted and actual values, respectively, $\Delta A_i = A_i - A_{i-1}$ and $\Delta P_i = P_i - A_{i-1}$ or $\Delta A_i = (A_i - A_{i-1})/A_{i-1}$ and $\Delta P_i = (P_i - A_{i-1})/A_{i-1}$, and n^0 is the number of periods being forecasted.

All the measures will be zero for perfect forecasts, larger values indicate poor forecasts. There are some differences among these forecasting accuracy measures. The RMSE penalizes model with large prediction error more than MAE does. Therefore, MAE is more appropriate when the cost of forecast errors is proportional to the absolute size of the forecast error, while RMSE is more appropriate to situation in which the cost of the error increases in accord with the square of the forecast error. The U statistic, either calculated in absolute changes or percentage changes, will reflect the model's ability to track turning points in the data (Greene). When U = 1, the forecast is as good as no-change forecast ($\Delta P = 0$). For U

6 The Henriksson-Merton test is also calculated, the results are consistent with other measures.

> 1, the forecast is less accurate than the simple forecast of no change. The following discussion focuses on the RMSE since it is the most popular measure.

Table 4.5 shows the number of times the particular forecasting method has the lowest RMSE[7]. There are 5 sets of 1-12 months ahead forecasts for each market and 8 markets for corn models (7 for soybean models), therefore there will be total of 40 RMSEs for corn and 35 for soybean. The RMSE averages are calculated for forecast period of 1-12 month ahead, and also for three shorter periods: 1-4, 5-8, and 9-12 months ahead (defined roughly as short term, intermediate term, and long term forecasts, respectively).

Table.5 Number of times the forecasting models have the lowest RMSE

Forecast	Forecasting Models [a]									
Periods	3YR	3YR+	ARIMA	USS	MSS2	MSS7	SBBM	NFD	SCG	COMP
					Corn					
1-12	6	4	6	2	4	4	1	8	0	5
1-4	7	8	0	3	0	2	8	4	1	7
5-8	6	5	1	10	3.5[b]	1.5	2	5	0	6
9-12	8	0	3	4	3	3	1	7	6	5
					Soybean					
1-12	3	11	9	1	1	2	1	1	1	5
1-4	5	5	7	1	3	1	3	1	2	7
5-8	6	12	2	1	2	0	4	2	3	3
9-12	2	9	8	1	2	4	0	3	1	5

[a] The forecasting models are simple 3-year-average, 3-year-average-plus, seasonal ARIMA, univariate state space, 2-crop multivariate state space, 7-market multivariate state space, structural basis behavior, neural network with NFD algorithm, neural network with SCG algorithm, and a composite forecast, respectively.

[b] The number with 0.5 indicates the lowest RMSE is shared by two models.

In general, in forecasting corn basis for eight locations, the three-year-average-plus (3YR+), artificial neural network with noise-feedback descent (NFD) learning algorithm, and composite forecasts (COMP) perform as well or slightly better than the three-year-average forecasts (3YR). For 1-12 month ahead corn basis forecasts, the artificial neural network with NFD has the largest number of lowest RMSEs; the second best models are the simple three-year-average and seasonal ARIMA model (each has the lowest RMSE 6 times), closely followed by the composite model. The structural basis behavior forecasting model (SBBM) and three-year-average-plus models outperform other models (each with 8 times of the lowest RMSE) in the short term (1-4 months) forecasts. The univariate state space model is best for the intermediate term forecasts, and no models beat simple three-year-average model in long term

[7] Detailed RMSE and other statistics are available up request.

forecasts (though both neural network models come close). The multivariate state space models, both 2-crop (MSS2) and 7-markets (MSS7), do not perform well in corn basis forecasting. Improved forecasts have a fraction of a cent up to less than 2 cents per bushel off the RMSE, which is approximately equivalent to the standard error of the forecast at the means. For some large volume grain merchandisers, that may be worth the added effort to develop, and update their forecast procedures, but for many it will not be worth while.

For soybean basis forecasts, the three-year-average-plus model, seasonal ARIMA model and the composite forecasts generally outperform simple three-year-average forecasts. The three-year-average-plus model is the best according this criterion for 1-12, 5-8 and 9-12 month ahead forecasts, while seasonal ARIMA and composite forecasting models outperform all other models in the short-term soybean basis forecasts, and also perform well for other forecasting periods.

Conclusions

The results of basis behavior models for both corn and soybeans generally consistent with previous studies. Nearby basis patterns exhibit seasonal patterns, and the relative importance of the variables affecting basis variation varies seasonally. For example, storage cost primarily is important in the early storage season. Barge rates significantly affect corn and soybean basis in the spring and fall. When upper Mississippi river is closed during the winter, the grain basis in St. Louis and Gulf Ports respond differently to barge rate changes compared to other times of the year. Production is an important factor only for months during and immediately after harvest. Demand factors have mixed results in this study. Export volumes usually have little effect on local basis levels. Basis behavior also differs between production locations and port locations. Significant basis convergence is found in this nearby basis from December to February for corn and November to December for soybean. No evidence suggest that basis behavioral factors are different between delivery and non-delivery markets in the primary production areas.

Basis forecasting results shows that the simple three-years-average forecast method can be outperformed by alternative models, even though it is still a reasonably good forecasts when basis variation and seasonal pattern are stable over years (such as in recent years). Overall comparisons show 3-year-average-plus and seasonal ARIMA models fare the best in our out of sample tests. Other more complicated approaches, like the state space and structural econometric models, have inconsistent performance. For example, the structural behavior model performs very well in 1- 4 month ahead corn basis forecasting, but not for longer term forecasts. But there is room for improvement in the simple ancillary forecast procedures we used. In conclusion, 3-year-average-plus and seasonal ARIMA models are most practical and much easier to use than other alternative models, and they slightly outperform simple 3-year-average forecasts. But, do the costs of developing and maintaining the improved forecast procedures outweigh the potential benefits? Perhaps large scale grain merchandisers, grain marketing consultants with multiple clients, or extension grain marketing specialists serving many clients may find such efforts worthwhile, while others will not.

References

Akaike, Hirotugu. "Canonical Correlation Analysis of Time Series and Use of the Information Criterion." in Mehra, R.K. and D.G.Laintiotis, eds. *System Identification: Advances and Case Studies.* Academic Press, NY (1976):27-96.

Azoff, E. Michael. *Neural Network Time Series Forecasting of Financial Markets.* John Wiley & Son, Chichester, England 1994.

Benirschka, Martin and James K. Binkley. "Optimal Storage and Marketing Over Space and Time." *America Journal of Agricultural Economics* 77(August 1995):512-524.

Brennan, Donna, Jeffrey Williams, and Brian D. Wright. Brennan, Donna, Jeffrey Williams, and Brian D. Wright. "Convenience Yield without the Convenience: A Spatial-Temporal Interpretation of Storage under Backwardation." Forthcoming in *The Economic Journal* (July 1997).

Brorsen, B. Wade, Jean-Paul Chavas, Warren R. Grant, and A. W. Ngenge. "Spatial and Temporal Relationships among Selected U.S. Grain Markets." *North Central Journal of Agricultural Economics.* 7(1985):1-10.

Cheng, Bing and D.M. Titterington. "Neural Networks: A review from a statistical perspective." *Statistical Science.* 9(1994):2-54.

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

Claussen, Kent L. and J.William Uhrig. "Cash Soybean Price Prediction with Neural Networks." *Applied Commodity Price Analysis, Forecasting, and Market Risk Management,* Proceedings of the NCR-134 conference (1994): 56-65.

Dasgupta, Chanda Ghose, Gary S. Dispensa, and Sanjoy Ghose. "Comparing the Predictive Performance of a Neural Network Model with Some Traditional market Response Models." *Journal of Forecasting* 10(1994):235-244.

Garcia, Philip and Darrel L. Good. "An Analysis if the Factors Influencing the Illinois Corn Basis, 1971-1981." *Applied Commodity Price Analysis, Forecasting, and Market Risk Management, Proceedings of the NCR-134 conference* (1983):306-326.

Grundnitski, Gary and Larry Osburn. "Forecasting S&P and Gold Futures Prices: An Application of Neural Networks." *Journal of Futures Markets* 13(1993):631-643.

Hamm, Lennie, B. Wade Brorson, and Ramesh Sharda. "Futures Trading with a Neural Network." *Applied Commodity Price Analysis, Forecasting, and Market Risk Management,* Proceedings of the NCR-134 conference (1993):286-296.

Hauser J. Robert, Philip Garcia and Alan D. Tumblin. "Basis Expectations and soybean Hedging effectiveness." *North Central Journal of Agricultural Economics.* 12(1990):127-136.

Heifner, Richard G.. "The Gains from Basing Grain Storage Decisions on Cash-Future Spreads." *Journal of Farm Economics.* 48(1966):1490-1495.

Hieronymus, T. A. "Basis Patterns." *Views from the Trade: Readings in Futures Marketings.* Edited by A.E. Peck. Book III(1978):45-56. This paper was originally presented in 1963 at the Hedging Symposium for Country Grain Elevators at the Chicago Board of Trade.

Hill, Tim and Leorey Marquez, Marcus O'Connor, and William Remus. "Artificial Neural Network Model for Forecasting and Decision Making." *Journal of Forecasting* 10(1994):5-15.

Jenkins, Gwilym M. *Practical Experiences with Modeling and Forecasting Time Series.* Gwilym Jenkins & Partners (Overseas) Ltd. 1979.

Kaastra, T and M.S. Boyd. "Forecasting Futures Trading Volume Using Neural Networks." *Journal of Futures Markets* 15(1995): 953-970.

Kahl, Kandice H.. "Changes in the Chicago Corn Basis, 1960-1975." *Agricultural Economics Research,* Vol. 34, No. 1 (1982): 25-29.

Kahl, Kandice H. and Charles E, Curtis, Jr. " A Comparative Analysis of the Corn Basis in Feed Grain Deficit and Surplus Areas." *Review of Research in futures Markets.* 5(1986):220-232.

Kaldor, Nicholas. "Speculation and Economic Stability." *Review of Economic Studies* 7(1939):1-27.

Kaldor, Nicholas. "A Note on the Theory of the Forward Market." *Review of Economic Studies* 7(1940):196-201.

Karlson, Nicholas; Brad Anderson and Reynold Dahl. *Cash-Futures Price Relationships: Guide to Corn Marketing.* Dept. of Agricultural and Applied Economics, University of Minnesota, Staff Paper Series P93-1 (1993).

Kenyon, David E. and Steven E. Kingsley. "An Analysis of Anticipatory Short Hedging Using Predicted Harvest Basis." *Southern Journal of Agricultural Economics* 5 (1973): 199-203.

Keynes, J. M. *A Treatise on Money,* vol. 2: *The Applied Theory of Money.* London: Macmillan & Co., 1930.

Kohzadi, N., M.S. Boyd, I. Kaastra, B.S. Kermanshahi, and Iebeling Kaastra. "Forecasting Livestock Prices with an Artificial Neural Network versus Linear time Series Models." *Applied Commodity Price Analysis, Forecasting, and Market Risk Management,* Proceedings of the NCR-134 conference (1994): 131-143.

Kohzadi, N., M.S. Boyd, I. Kaastra, B.S. Kermanshahi, and D. Scuse. "Neural Networks for Forecasting: an Introduction." *Canadian Journal of Agricultural Economics* 43(1995):463-474.

Kuan, C.M. and T. Liu. "Forecasting Exchange Rates Using Feedforward and Recurrent Neural Networks." *Journal of Applied Econometrics* 10(1995):347-364.

Leuthold, Raymond M., Joan C. Junkus and Jean E. Cordier. *The Theory and Practice of Futures Markets.* Lexington Books (1989): chapter 3 and chapter 7.

Martin, Larry, John L. Groenewegen and Edward Pidgeon. "Factors Affecting Corn Basis in Southwestern Ontario." *American Journal of Agricultural Economics.* 62(1980):107-112.

Naik, Gopal and Raymond M. Leuthold. "A Note on the Factors Affecting Corn Basis Relationships." *Southern Journal of Agricultural Economics.* (1991):147-153.

Powers, Nicholas and Aaron Johnson, Jr. "Forecasting the Storage-season Wisconsin Basis for Corn." *Applied Commodity Price Analysis, Forecasting, and Market Risk Management, Proceedings of the NCR-134 conference* (1983):327-341.

Strobl, Maximilian, T. Randall Fortenbery, and Paul L. Fackler. *Forecasting Futures Basis: Model Interpretation and Evaluation*. Unpublished paper. North Carolina State University, 1996.

Taylor, Patricia D. and William G. Tomek. "Forecasting the Basis for corn in Western New York." *Journal of Northeast Agricultural Economic Council,* 13(1984):97-102.

Telser, Lester G. "Futures Trading and the Storage of Corn and Wheat." *Journal of Political Economy* 66(1958):1-22.

Tomek, William G.. *Commodity Futures Prices as Forecasts.* Working Paper, Department of Agricultural, Resource, and Managerial Economics, Cornell University. WP96-07. August 1996.

Thompson, S. R., J. S. Eales, and R. J. Hauser. "An Empirical Analysis of Cash and Futures Grain Price Relationships in the North Central Region." *North Central Journal of Agricultural Economics.* 12(1990):241-254.

Uhrig, J. William, Bernard A. Engel and W. Lance Baker. "An Application of Neural Networks: Predicting Corn Yields." *Applied Commodity Price Analysis, Forecasting, and Market Risk Management*, Proceedings of the NCR-134 conference (1992): 407-417.

Working, Holbrook. "Theory if the Inverse Carrying Charge in Futures Markets." *Journal of Farm Econ.* 30(1948):1-28.

_______. "The theory of Price of Storage." *American Economic Review.* 39(1949):1254-1262.

Wright, B. D., and J.C. Williams. "A Theory of Negative Prices for Storage." *Journal of Futures Markets.* 9(1989):1-13.