Predicting the Corn Basis in the Texas Triangle Area

Vardan Mkrtchyan[1], J. Mark Welch[1,2] and Gabriel J. Power[1]
Texas A&M University
[1] Department of Agricultural Economics and [2] Texas Agri-Life Extension

Selected Paper prepared for presentation at the Southern Agricultural Economics Association Annual Meeting, Atlanta, Georgia, January 31-February 3, 2009

Abstract

The basis is a vital concept in the production, marketing and hedging of many commodities. Concern over basis levels has intensified in corn markets recently because of some significant changes in the corn market place. Corn producers and users would stand to benefit from a new, flexible, and a better performing method to predict the basis. Being able to predict the basis more accurately makes it easier to market corn efficiently and to maximize profit. This study develops a new and straightforward economic model of basis forecasting that outperforms the simple three-year average method suggested in much of the literature. We use monthly data of the corn basis in the Texas Triangle Area from February 1997 to July 2008. The results and the graphs indicate that the new model based on economic fundamentals performs better than basis estimates using a three-year moving average.

Introduction

A central issue for farmers in commodity marketing is forecasting the basis, which is defined as the difference between the cash price and the futures price for a commodity in a specific delivery location and of specific quality grade (Tomek, 1997). In the U.S., corn has long been the crop with the highest total dollar value. The importance of corn increased with the Energy Independence and Security Act of 2007, which mandates the production of at least 36 billion gallons of bio-fuel by the year 2022. It is estimated that 15 billion gallons of the 36 billion gallon mandate will come from corn based ethanol. The U.S. currently has 128 ethanol plants and an additional 85 under construction. Production is concentrated in the grain surplus Midwestern states while Southeast and Southwest states, including Texas, are grain deficit states (see Figure1). The basis is affected by whether a state is in a corn surplus or deficit region. Emerging ethanol production in the Midwest is expected to strengthen the basis in that region meaning importers like the Texas High Plains will need to bid more to get the corn supplies they need. The basis in Texas will be affected as well.

[Figure 1 approximately here]

The focus of this paper is to forecast the corn basis in the Texas Triangle Area, a statistical reporting region defined by the National Agricultural Statistics Service and located in the Texas High Plains (see Figure 2). It includes elevators in an area from Plainview to Canyon to Farwell and is comprised of Castro, Deaf Smith, Parmer, Randall, and Swisher counties in the Texas panhandle. The Triangle Area is a leader of Texas corn production and is at the heart of the Texas cattle feeding industry (TASS, 2008). In addition, White Energy Inc. of Dallas, Texas began operation of a 100 million gallon per year (mgy) corn ethanol plant in Deaf Smith County on January 15, 2008. White Energy also operates a

100 mgy corn ethanol plant in adjacent Hale County, Texas. An additional 100 mgy ethanol plant is currently under construction in Deaf Smith County.

[Figure 2 approximately here]

It is likely that the pattern of corn basis is undergoing changes given the effects of ethanol policies, increased transportation costs, and volatility in the grain markets more generally. The purpose of this paper is to compare forecasts of the basis, given these dynamic conditions, based on estimated models of the determinants of the basis. Two approaches are compared using both in-sample and out-of-sample data: a purely statistical three-year moving average of the basis, and a model that uses as explanatory variables publicly available data on economic fundamentals that are well supported by economic theory. By doing so, the paper makes methodological and policy contributions to understanding the relationship between grain futures markets in Chicago and local cash markets.

Literature Review

Even though predicting the basis and having accurate estimates for local markets is essential, Jiang and Hayenga (1997) note that there have been few basis behavior studies and even fewer basis forecasting studies (not counting simple moving average estimations of historical basis data). The model used by Jiang and Hayenga includes storage cost, transportation cost, and regional supply and demand variables to explain basis behavior. They use a number of forecasting techniques for the corn and soybean basis, including a simple three-year-moving average forecast, a structural econometric model, a modified three-year average model, artificial neural networks, seasonal ARIMA time series models, state-space models, and composite forecasts.

They report in their conclusion that export levels have little to no effect on the local basis. They conclude that three-year-average-plus and seasonal ARIMA models are the most practical, are much easier to use than other alternative models, and slightly outperform the simple three-year-average forecast. Sanders and Manfredo (2006) also find, in the case of the soybean complex, that the gains from using sophisticated time series models rather than a simple moving average to forecast the basis are relatively small.

In their study Taylor, Dhuyvetter, and Kastens (2006) compare practical methods of forecasting the basis. They look at current market information of wheat, soybeans, corn and milo (grain sorghum) in Kansas. They use nine different models to forecast the basis and conclude that, despite not having any rule to define the best forecasting method, using the one-year average basis to forecast the futures basis has worked better than long-term averages with some products. They also state that to forecast the wheat basis at harvest, the five-year average is the best forecast model.

Parcell, Schroeder and Dhuyvetter (2000) look at the live cattle basis in three different states and use a multivariate model to predict the basis. The authors state that the explanatory variables explain 85% of the variation in each state. They also state that corn prices have a significant effect on live cattle basis but the magnitude is lower than was suggested in another study.

Tomek (1997) notes that there has been considerable research done on modeling basis behavior but the number of forecasting analyses is small. Tomek adds that it is often very difficult to obtain the data for all the variables influencing basis behavior, therefore forecasts of the basis have been made from simple time series or naïve models. In his analysis, Tomek looks into two types of basis models. The first is related to inventories carried over from one crop year to the next. This model uses the cash prices pertaining to a period near the end of the current crop year and futures quotes for the first contract in the new crop year. Tomek states that this basis measures how large is the incentive for carrying stocks from one year to the next. The second model is related to inventories within the same year. This model is related to basis changes within a year, that is, changes over a storage interval.

Tomek concludes that existing price forecasting models are generally poor predictors of futures prices but might be valuable to individual enterprises as they develop or obtain information not available to others. He also notes that the effect of small or dwindling inventories on prices is much larger than the effect of large or plentiful inventories. This finding suggests that inventories should be included among the explanatory variables for the basis.

Garcia and Good (1983) examine the factors influencing the corn basis in Illinois. They argue that the supply and demand of storage should be included as explanatory variables for the basis in addition to the cost of storage and transportation. They write that small stocks (inventories) or a strong demand for shipments (exports) could strengthen the basis. They conclude that the three sets of variables that influence the basis are cost, stock, and flow factors. Garcia and Good use cross-section data and time series data for their model. They hypothesize that high levels of corn and soybean stocks create a high demand for storage which in itself creates high price for storage everything else held constant. They also expect that high levels of corn stocks and a high cost for storage create a wider basis. Garcia and Good include barge rates, regional dummy variables, monthly dummy variables, and interest rate to reflect the relationship between cost and the basis. They conclude that the basis patterns are fairly systematic. They find that storage has a strong positive impact on Illinois basis during harvest time and slightly diminishes in other periods. The cost of transportation is important during the off-harvest season but not during the harvest season.

Hranaiova and Tomek (2001) discuss the importance of the timing option on the basis behavior. They look at the basis as a function of interest rate, convenience yield, storage cost, time to maturity and timing option. Their OLS regression estimates show that at day one of the maturity month, the timing option is statistically important and with convenience yield included, represents about 92% of the basis.

Tomek and Peterson (2001) emphasize the importance for hedging of understanding the basis. They discuss different marketing strategies for farmers to maximize profits and argue that getting a good forecast of the basis is a difficult but important task.

Most previous studies conclude that an averaging method to forecast the basis is the most practical. This paper compares an alternative method based on a few relevant variables from readily available data sources to the traditional moving average approach. If the new model is seen as providing better estimates of the cash to futures price relationship, it will be useful to producers and users of corn in the Texas panhandle in formulating price expectations. It may also provide a foundation for corn producers in other areas who seek a better way of forecasting the basis in their region.

Methodology

Based on economic theory, the previous literature, and the goal of keeping the model succinct, we choose seven variables that we anticipate to be significant in predicting the Texas corn basis. These variables and their predicted signs are:

1. Local cash price (+);
2. Futures price, December maturity (-);
3. Estimated marketing year ending stocks (-);
4. Transportation costs (+);
5. The basis in a previous time period (+);
6. Texas Off-Farm Inventories (-); and a
7. Harvest Dummy (-).

The choice of average cash and average futures prices is based on the definition of the basis (basis = cash price minus futures price). The relevant futures contract for corn marketing in this region is the December contract on the Chicago Board of Trade. The ending stocks variable is included following the Kaldor-Working theory of storage because corn is a storable commodity and estimated levels of ending stocks are important measures of supply and demand fundamentals. A transportation cost variable is included since Texas is a corn deficit state and corn is imported into the state from corn-abundant states. This is intended to capture the effect of oil price increases from 2005 to 2008. A lagged basis variable is added to stabilize the data and to account for serial correlation. A Texas Off-Farm inventories variable is added to capture the affect of local inventories on local basis. A harvest-time dummy variable is added to capture the influence of harvest on the Triangle

Area basis. All regressions are run in SAS and predictions are calculated in Excel. The model that we propose is given by:

$$Basis_t = \beta_0 + \beta_1 Basis_{t-1} + \beta_2 Avg.Cash_t + \beta_3 Avg.Dec.Futures_t + \beta_4 EndingStocks_t + \\ + \beta_5 Transportation_t + \beta_6 TexasOffFarm_t + \beta_7 HarvestDummy_t + \varepsilon_t$$

$$\text{for } t = 1, \ldots, 138$$

where:

$Basis_{t-1}$ is the lagged basis one period (monthly);

$Avg.Cash_t$ is the average cash prices in time t in the Texas triangle region. ;

$Avg.Dec.Futures_t$ is the average December Futures Price of corn at time t at the Chicago Board of Trade;

$EndingStocks_t$ is the projected ending stock of corn reported by USDA;

$Transportation_t$ is the transportation index with a base year of 1985;

$TexasOff-Farm_t$ is the inventory data for the Texas Off-farm corn reported quarterly;

$HarvestDummy_t$ is a dummy variable for month of October.

The baseline model chosen is the three-year moving average suggested by the literature to be the simplest and most practical way of calculating the basis:

$$Basis_t = \beta_0 + \beta_1 MA3_t + \varepsilon_t$$

$$\text{for t = 1,…,103}$$

where $MA3_t$ is the three-year moving average of the basis.

Data

The data for the basis model are readily available. The average cash corn price data in the Triangle Region is from the Texas AgriLife Extension website at Texas A&M University's

Department of Agricultural Economics. Futures prices are from the Commodity Research Bureau *Data Xtract*. The average monthly price is a simple average of daily closing prices in the nearest December contract. Corn ending stocks are from the USDA National Agricultural Statistics Service (NASS). Monthly updates of projected ending stocks are collected from the USDA's World Agricultural Supply and Demand Estimates. Transportation data is a monthly producer price index for railroad transportation costs. It is obtained from the Bureau of Labor Statistics in US Department of Labor. Texas Off-Farm inventory levels are from the USDA website. The time period for all the data is from February 1997 to July 2008. Table 1 contains the descriptive statistics for the variables chosen for this study.

[Table 1 approximately here]

Testable Hypotheses

In our model the joint null hypothesis is that: (i) the following set of economic fundamental variables is significant in explaining the basis, and that (ii) the variable coefficients have the signs predicted by economic theory. It is expected that the basis will be:

- Increasing in **average cash price in the Triangle Area** from the identity Basis=Cash-Futures;
- Decreasing in the **average December futures price**, also from the identity;
- Decreasing in the **monthly update of projected ending U.S. stocks (inventories)**, since higher ending inventories are associated with tight storage conditions that may force cash sales thus weakening the basis;

- Increasing in **transportation cost** because higher fuel costs imply it is more expensive to bring corn out of grain surplus regions (i.e. near the par delivery for Chicago Board of Trade futures) to grain deficit regions such as the Triangle Area;
- Increasing in **lagged basis**, because the basis is (weakly) serially correlated; and
- Decreasing in the **Texas off-farm inventories**, because higher regional inventories should depress local cash prices and weaken the basis.

Dummy variable is included for seasonality (harvest). Precisely, the seasonality dummy variable takes the value 1 if it is October and takes the value 0 otherwise. Diagnostic tests are performed on the data to evaluate the presence of heteroskedasticity and serial correlation, with the necessary adjustments being made in the positive case.

Results and Interpretation

This section presents the results obtained from running corrected Ordinary Least Squares (OLS) regressions on the two principal specifications as well as specifications that exclude one or more insignificant independent variables.

Economic Fundamentals Model

The results for our proposed "economic fundamentals" model are summarized in Table 2. All of the results are reported at the 95% confidence level. The coefficient for the Lagged Basis variable is 0.4752 and is significant. The implication is that, all else held constant, if the basis in the previous month is one cent/bushel higher, then the basis in the current month increases by about half a cent. This finding confirms the expectation that the basis is weakly serially correlated. In other words, if the basis for previous month is getting stronger (more positive) the basis for the next month will keep strengthening everything else held constant.

[Table 2 approximately here]

The average cash price variable is also significant with a coefficient of 0.1033. If the local cash price in the Triangle Area region goes up by one cent per bushel, the basis will increase by one tenth of one cent, all else held constant. This result is consistent with the basis formula expressed as cash minus futures.

The average December futures price variable has a negative and significant coefficient of -0.1446. Again, the sign for this variable is consistent with the basis definition as cash price minus futures price. If December futures prices go up by one cent then the basis in the Texas Triangle region will weaken by 0.1446 cents per bushel, all else held constant.

The Projected Ending Stocks variable is statistically significant and negative as expected but the coefficient is very small. The coefficient associated with one million bushels of ending stocks is -0.00002964, implying that it takes a change of one billion bushels in ending stocks to change the basis by 3 cents, *ceteris paribus*. Current estimated U.S. ending stocks for 2008-2009 are 1.154 billion bushels. It would take a change in projected ending stocks of about thirty percent to change the basis one cent. This result is consistent with the theory because higher project year-end inventories suggest declining demand or increasing supplies and lower cash prices.

The transportation index variable has a positive and significant estimated coefficient of 0.00203. This result is consistent with the fact that Texas is a corn deficit state and corn is being imported to Texas from other corn abundant states. If the transportation index goes up by one percentage point, the basis strengthens by 0.2 cents per bushel, all else constant. As it costs more to bring corn from other states to Texas, buyers can afford to pay more to local producers rather than transport it from out of state, strengthening the basis.

Some variables are not statistically significant and are excluded from the final regression specification. These are the Texas Off-Farm Inventory levels variable and the harvest dummy variable. Exclusion of these two variables does not substantially affect the RMSE, although both R^2 and goodness-of-fit decrease. The parameter associated with the Texas Off-Farm inventories variable is negative but not significant. The sign indicates that the basis weakens as local grain inventories increase. Increasing inventories could be a sign of weakening demand which could weaken the basis. Increasing inventories might also reflect large grain production in the area or difficulty arranging transportation to move grain out of inventory. Elevators with full bins would not offer price incentives to encourage producers to bring in more grain. They are more likely rather to weaken basis bids to discourage short term grain deliveries. The Texas Off-Farm inventory variable may not be significant because the data are measured quarterly which is a lower frequency than the monthly data collected for the other variables or because local storage capacity relative to total local demand is small.

The harvest dummy variable has a coefficient of -0.00746 and is not significant. It is dropped from the final regression model. The negative sign of the parameter is consistent with the theoretical prediction that at harvest, the local increase in corn supply depresses the cash price and weakens the basis.

Moving-Average Model

Our comparison model is a three-year moving average of the basis. The results for this model are presented in Table 3. The coefficient for the three-year moving average is 0.4. This model has less explanatory power than does the economic fundamentals model. The R^2 is lower (0.1062) and the Root Mean Squared Error is higher (0.082). The results show that the economic fundamentals model has greater explanatory power for the Triangle Area corn basis. The R^2 of 0.6738 is much greater than the moving-average model R^2 of 0.1062, and the economic model RMSE of 0.0524 is much smaller than the moving-average model RMSE of 0.082.

[Table 3 approximately here]

The improved accuracy of the economic fundamentals model also provides economically significant gains. Consider the problem of a grain merchant who owns an inventory of 100,000 bushels of corn stored for future sale and who estimates the basis to implement his marketing strategy. If he chooses the economic fundamentals model instead of the baseline model he will save $0.02963/bu or $2963[*] for the sale of his inventory. Even though our model is more complex than a straightforward three-year moving average model, the results clearly suggest that the added difficulty is worthwhile. The superiority of the model is illustrated by Figures 3 and 4.

[Figures 3 and 4 approximately here]

[*] This result is obtained by multiplying the quantity of corn (100,000 bu) by the difference between the RMSE for the economic model and the RMSE for the moving-average model, that is, (0.082-0.0524)x100.000=$2963

Conclusion

Understanding the behavior of the basis is essential in grain marketing. It is the means by which the price discovery function of the futures exchange is expressed to producers and users of commodities in specific locations. Recent changes in the fundamentals of corn demand due to ethanol production may have altered the cash-futures relationship in many areas. Specifically, the construction of an ethanol plant in the Texas panhandle may change these market dynamics.

This paper shows that a traditional three-year moving average model of the basis does not track changes in the basis as well as a relatively simple economic model is able to do. We created a model that uses a few significant variables from easily obtained data sets to explain the basis in the Triangle Area better than a three-year moving average. Additional research is needed to improve the basis predictions to make them more responsive to changes in market fundamentals and the other factors that drive the basis levels. It is a challenge to balance potential gains from using more sophisticated methods against the cost of collecting extra data and estimating more complicated models. Although this paper considers a wide range of economically meaningful variables, there remain some explanatory variables that could be further studied to evaluate their contribution to the basis forecasting. One example of a potentially useful explanatory variable is the level of export activity from the ports of Texas.

Our economic fundamentals model includes limited data on the impact of new corn ethanol production capacity in the Texas panhandle. New estimates of the basis after plants under construction have come on line and been in operation longer will provide insight into whether there has been a fundamental shift in the basis due to ethanol manufacture in the area. All of these efforts are designed to give regional farmers and corn users more accurate predictions and guidance for future marketing decisions.

References

Chicago Board of Trade. 2000. *Understanding Basis*. Chicago, IL: Chicago Board of Trade. Available at < www.cbot.com/cbot/docs/46577.pdf>.

Dhuyvetter C. K. and T. L. Kastens. 1998. "Forecasting Crop Basis: Practical Alternatives." In: *Proceedings of the NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management.* Available at <http://www.farmdoc.uiuc.edu/nccc134>.

Garcia, P. and L. D. Good. 1983. "An Analysis of the Factors Influencing the Illinois Corn Basis, 1971-1981." In: *Proceedings of the NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management*. Available at <http://www.farmdoc.uiuc.edu/nccc134>.

Hranaiova, J. and W.G. Tomek. 2001. "Role of Delivery Options in Basis Convergence." *Journal of Futures Markets* 22(8): 783-809.

Jiang, B., and M. Hayenga. 1997. "Corn and Soybean Basis Behavior and Forecasting: Fundamental and Alternative Approaches." In: *Proceedings of the NCR-134 Conference of Applied Commodity Price Analysis, Forecasting, and Market Risk Management*. Available at <http://www.farmdoc.uiuc.edu/nccc134>.

Mintert J., K. Dhuyvetter, E. E. Davis, and S. Bevers. 2002 "Understanding and using Feeder and Slaughter Cattle Basis." *Managing for Today's Cattle Market and Beyond.* Department of Agricultural Economics, Kansas State University. March. Available at <http://www.agecon.ksu.edu/livestock/Master%20Web%20Page%20Folder/Extension%20Bullentins.Research/ManageForTodaysCattleMkt/feederslghtrcattlebasis2002.pdf>.

Outlaw, L. Joe, D. P. Anderson, S. L. Klose, J. W. Richardson, B. K. Herbst, M. L. Waller, M. J. Raulston, S. L. Sneary, R. C. Gill. 2003. "An Economic Examination of Potential Ethanol Production in Texas." Department of Agricultural Economics, Texas A&M University, College Station, Texas, February.

Parcell L. Joe, T. C. Schroeder, and K. C. Dhuyvetter. 2000 “Factors Affecting Live Cattle Basis.” *Journal of Agricultural and Applied Economics* 32(December): 531-541.

Sanders, R. D., 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 (December): 513-523.

Seamon, V. F., C. H. Kahl, and C. E. Curtis Jr. 1997. “A Regional Comparison of U.S. Cotton Basis Patterns.” Department of Agricultural and Applied Economics, Clemson University, Working Paper WP123197, December.

Taylor, R.M., K. C. Dhuyvetter, and T. L. Kastens. 2006. “Forecasting Crop Basis Using Historical Averages Supplemented with Current Market Information.” *Journal of Agricultural and Resource Economics* 31(December): 549-567.

Tomek, W. G. and H.H. Peterson. 2001. “Risk Management in Agricultural Markets: A Review.” *Journal of Futures Markets* 21(10): 953-985.

Tomek, W. G.. 1997. “Commodity Futures Prices as Forecasts.” *Review of Agricultural Economics*, 19(1): 23-44.

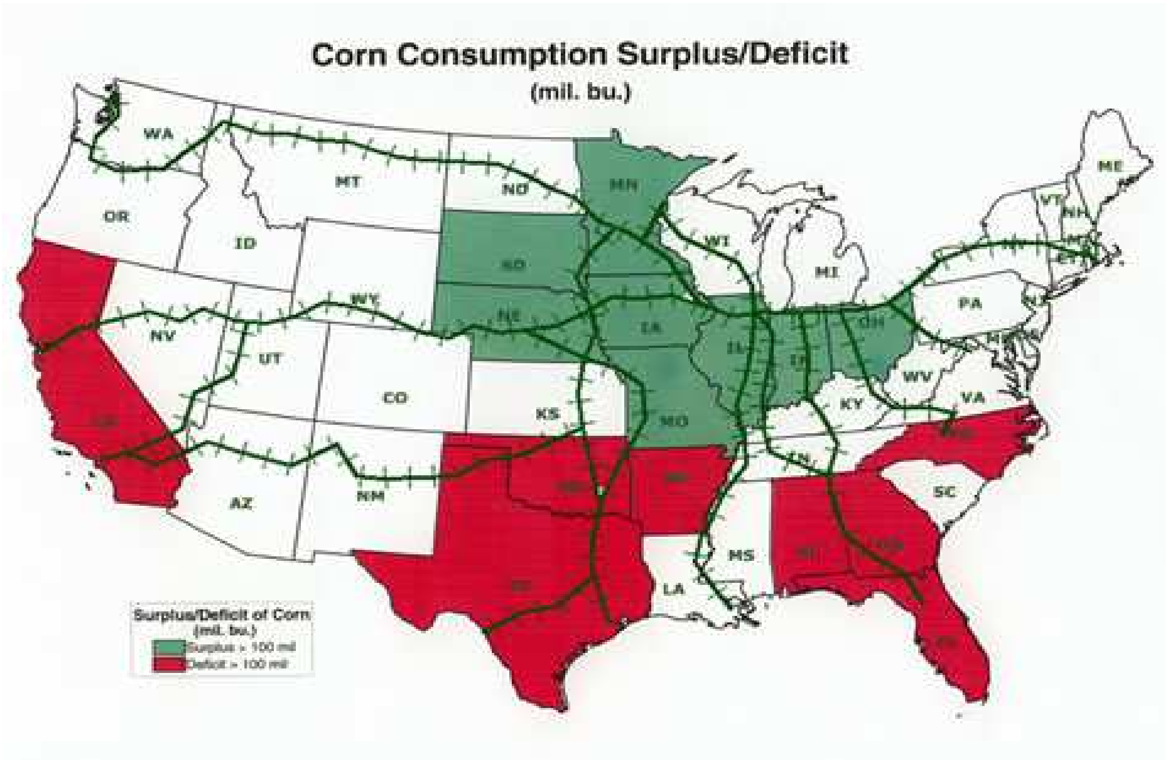

Figure 1: Corn Consumption Surplus/Deficit in the United States.

Figure 2: Texas Triangle Region
Source: http://agecoext.tamu.edu/files/images/maps/Triangle.jpg

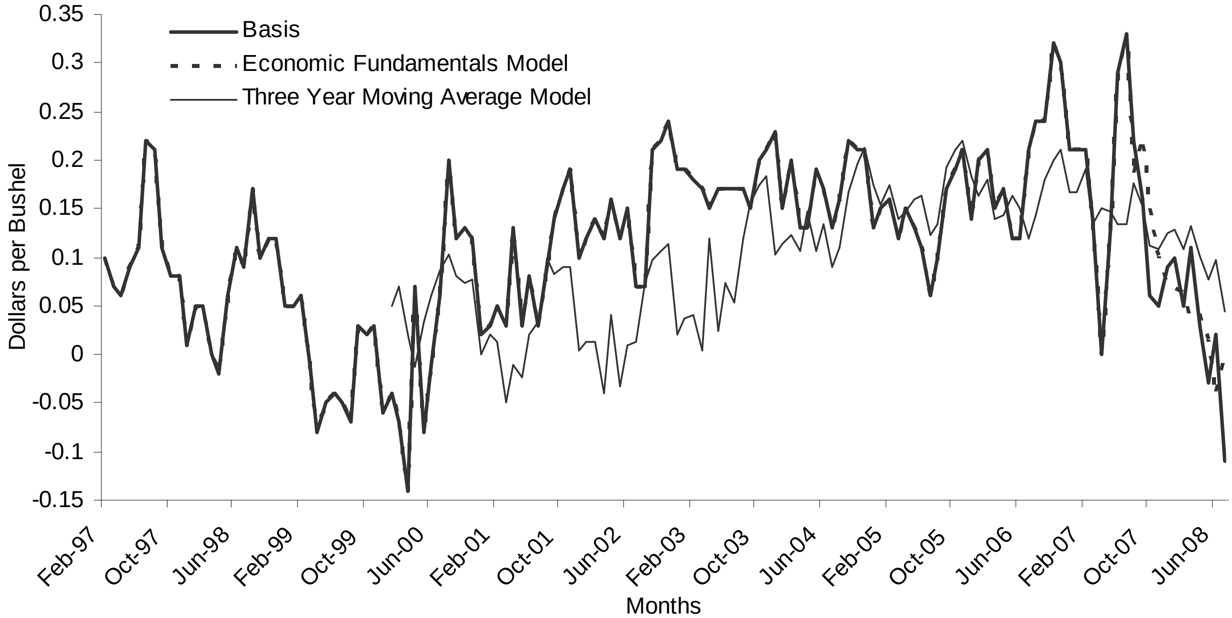

Figure 3: Actual Basis, Basis Prediction from the Economic Fundamentals Model and Basis Prediction from the Three Year Moving-Average Model, using the Complete Sample from Feb. 1997 to Jul. 2008

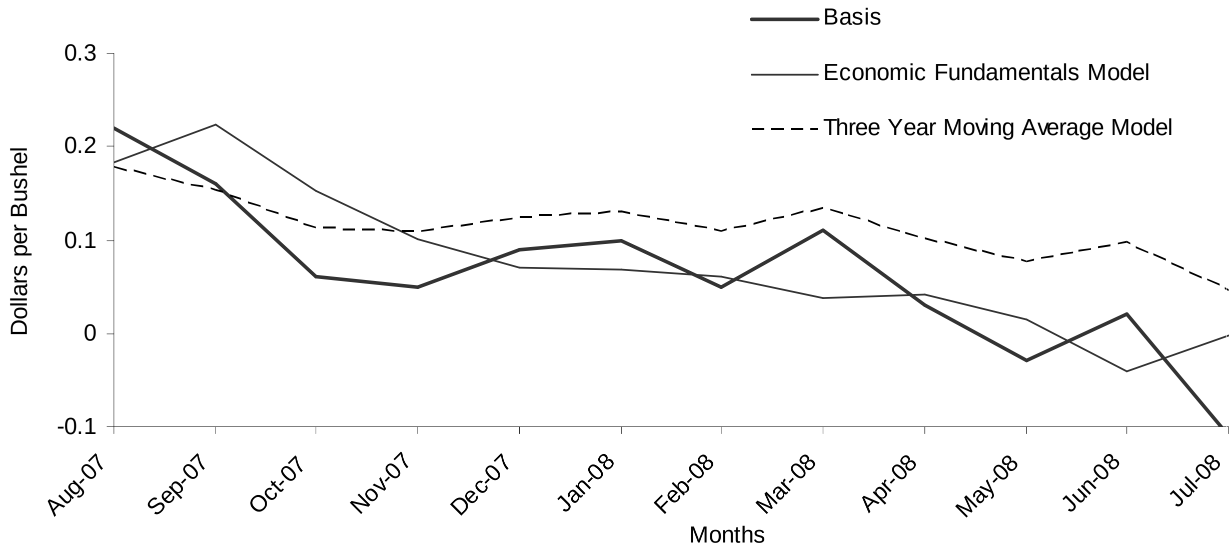

Figure 4: Actual Basis, Basis Prediction from the Economic Fundamentals Model and Basis Prediction from the Three Year Moving-Average Model, from Aug. 2007 to Jul. 2008

Table 1: Descriptive Statistics of the Variables

Variables	Units	Mean	Standard Dev.	Kurtosis	Skewness	Min	Max
Basis	dollars/bu	0.113	0.09	0.020	-0.353	-0.140	0.330
Basis Lagged	dollars/bu	0.115	0.088	0.025	-0.320	-0.140	0.330
Average Cash	dollars/bu	2.752	0.924	6.570	2.447	1.913	7.110
Average Dec. Futures	dollars/bu	2.728	0.941	8.191	2.727	1.890	7.304
Texas Off-Farm	in 1000bu	57523.435	32090.190	-1.113	0.069	6032	115256
Ending Stocks	in million bu	1489.130	495.565	-1.042	0.107	673	2540
Transportation	index	128.402	18.551	-0.048	1.101	111.5	180.3

Sample size: T=138

Table 2: Economic Fundamentals Model Parameter Estimates, Standard Errors and t-Statistics

Variables	Parameter Estimates	Standard Error	t-statistic	*p*-value
Intercept	-0.048	0.03716	-1.29	0.1987
Basis, lagged	0.47525**	0.07489	6.35	<0.0001
Average Cash Price	0.10327**	0.03343	3.09	0.0024
Avg. Dec. Futures Price	-0.14456**	0.03221	-4.49	<0.0001
Ending Stocks	-0.00002964*	0.0000134	-2.21	0.0287
Transportation	0.00203**	0.00049761	4.08	<0.0001
Dropped (Insignificant) Variables				
Texas Off-Farm Inventories	-0.21041	0.15286	-1.38	0.1710
Harvest Dummy	-0.00746	0.01844	-0.4	0.6867

*significant at the 5% level, ** significant at the 1% level

Model 1 Root MSE 0.05239 **R^2 =** 0.6738

Table 3: Three-Year Moving Average Parameter Estimates, Standard Errors and t-Statistics

Variables	Parameter Estimates	Standard Error	t-Value	Confidence Level Pr>|t|
Intercept	0.08879**	0.01508	5.89	<0.0001
Three-Year Moving Average	0.4002**	0.1155	3.46	0.0008

** significant at the 1% level

Model 2 Root MSE 0.08202 **R^2 =** 0.1062