NCCC-134

APPLIED COMMODITY PRICE ANALYSIS, FORECASTING AND MARKET RISK MANAGEMENT

A Spatial Approach to Estimating Factors that Influence the Corn Basis

by

Michael K. Adjemian, Todd Kuethe, Vince Breneman, Ryan Williams, Mark Manfredo, and Dwight Sanders

Suggested citation format:

Adjemian, M. K., T. Kuethe, V. Breneman, R. Williams, M. Manfredo, and D. Sanders. 2011. "A Spatial Approach to Estimating Factors that Influence the Corn Basis." Proceedings of the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management. St. Louis, MO. [http://www.farmdoc.illinois.edu/nccc134].

A Spatial Approach to Estimating Factors that Influence the Corn Basis

Michael K. Adjemian[1], Todd Kuethe[1], Vince Breneman[1], Ryan Williams[1]
Mark Manfredo[2], and Dwight Sanders[3]

[1]USDA Economic Research Service
[2]Arizona State University
[3]Southern Illinois University

Paper presented at the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management St. Louis, Missouri, April 25-26, 2011.

Abstract: It is well known that supply and demand fundamentals at any location affect the local basis. Because grain markets are tied together by spatial arbitrage, the local basis may also be affected by the supply and demand factors at neighboring locations. Whether or not this is the case, the corn basis is highly clustered across the United States; as such, OLS estimates of basis determinants may be inconsistent. We apply a spatial econometrics framework to adequately control for spatial effects, and find that the county-level corn basis is characterized by spatial spillovers: supply and demand factors in a given county affect its own basis, but also radiate out over space affecting the basis at neighboring counties. We find that unobserved basis determinants are also spatially correlated.

Keywords: Corn, Basis, Determinants, Spatial Econometrics, Spillover

In the local market, prices for grains like corn and wheat are quoted using the basis, defined as the cash price minus the nearest-to-expiration commodity futures contract price. The basis represents the price of storage at a particular location, which is influenced by local commodity supply and demand conditions. Grain market participants use their knowledge of the basis to make important decisions that affect the production and marketing of a commodity: which crop to plant, when and how to hedge, when to sell or buy, and whether to store or not. Conventional empirical studies of the grain basis are generally concerned with estimating the determinants of the basis or forecasting its future levels. This literature typically describes a time series of basis observations at a single location, or a small group of locations. To identify the importance of factors involved in basis determination, econometricians regress basis observations over time on variables that represent local supply and demand of a commodity.

Less explored in this literature is the information provided by the spatial distribution of the basis. CashGrainBids.com collects hundreds of bids each day from different locations across the United States, each of them facing its own supply and demand factors. A more comprehensive portrait of basis behavior would take full advantage of the information provided by the cross-section. Moreover, the basis at a given location is likely to be affected by spillovers. Grain markets are tied together by spatial arbitrage. If markets are integrated and efficient, the equilibrium price at a given location is tied to the price at neighboring locations by transportation costs; otherwise, arbitrage profit opportunities exist. It is well known that changing supply and demand fundamentals at any location affects the local basis. Our aim is to discover whether these changes are transmitted to nearby basis levels, as well.

As shown in figure 1, the corn basis is distributed smoothly over space; areas with a similar basis level tend to group together. Data that are clustered in this way present an important problem: observations over space may not be independent. If observed explanatory variables cannot account for these clusters, the basis at different locations may be mutually determined, and conventional techniques to estimate factors that affect the basis will suffer from omitted variable bias. Furthermore, unobserved factors that are represented by regression residuals may themselves be spatially correlated. As a result, the Gauss-Markov assumptions are violated, and Ordinary Least Squares (OLS) regression—the standard estimation method in the basis literature—will not suffice; due to simultaneity bias, OLS is still inappropriate even if the disturbance terms are i.i.d. Spatial econometrics provides a useful framework to test whether observations are clustered spatially, and if so, to recover consistent estimates.

[FIGURE 1 ABOUT HERE]

In this paper, we use a spatial approach to estimate the determinants of the local corn basis. We confirm that the corn basis is strongly clustered for 785 counties throughout the United States. Our estimates show that the corn basis is characterized by spatial spillovers: local supply and demand factors affect the local basis but also radiate out over space, affecting the basis in neighboring regions. Our results have important implications for the study of the basis. First, researchers should consider accounting for spatial effects: because the basis is mutually determined over space, the spatial filter can generate consistent estimates for the supply and demand factors that affect the local basis. Next, the spatial framework can improve understanding of the price discovery process by determining how regional shocks affect basis level at locations across the United States. Finally, because the basis is characterized by mutual determination, the spatial framework may be used to improve the forecasting process. In future work, we intend to introduce a forecasting model that incorporates time-space basis data, and assess its empirical usefulness.

The remainder of the paper is organized as follows: Section I draws on earlier basis work to explain the context for our research, Section II describes the data we use, Section III presents our modeling and estimation framework, Section IV documents our results and an explanation of their implications, Section V adds concluding thoughts.

I. Background

Unhedged corn market participants are exposed to commodity price fluctuations, which follow a martingale. One advantage of hedging is that it substitutes basis risk for price risk. The basis, which is the local cash minus the futures price, is less volatile than the commodity price level and it follows a known trend, since it theoretically converges to zero at futures contract maturity. Understanding the determinants of the basis is vital to a successful risk management strategy. Numerous studies have estimated the empirical significance of supply and demand factors in grain basis determination (Garcia and Good 1983; Kahl and Curtis 1986; Tilley and Campbell 1988; Jiang and Hanyenga 1997); to accomplish this, researchers specify a reduced form to explain the basis values at a particular location, or a small set of locations. This literature shows the while changes in the basis size tend to vary from month-to-month, the importance of local factors themselves is highly seasonal in nature. For example, current crop year production may influence harvest-time basis, but does not have much to say during the post-harvest period. Overall, local production and consumption, stocks, storage capacity and cost, and transportation costs make up the set of factors consistently found to affect the basis.

Several recent studies demonstrate that basis levels at spatially separated markets are linked. McKenzie (McKenzie 2005) shows that shocks to the soybean basis at the Gulf of Mexico are transmitted to the Little Rock and Memphis markets. Because the Gulf basis is not much affected by shocks at the two interior markets, the former dominates the spatial relationship. Using bivariate Granger causality, Manfredo and Sanders (2006) describe the connection between the corn basis at seven locations, finding evidence of both dominant-satellite as well as simultaneously determined pairs. Building on that work, Lewis et al. (Lewis, et al. 2010) use a similar method to establish directional ties between 13 corn markets. In addition, they specify a purely spatial autoregressive model to determine the average spillover effect, and find that the basis is simultaneously determined in these markets.

II. Data

Local cash price and basis data are drawn from CashGrainBids.com, and represent the monthly average corn bid for 785 counties across the United States. Since the focus of this paper is on a single cross-section, we choose to model the factors that affect the August 2007 basis, since that period matches up well with the contemporaneous Census of Agriculture, and occurs just before the harvest begins, simplifying the collection of animal unit and production figures. We average the basis observations up to the county level, and model the basis for all 785 counties for that period.

To test for the importance of factors that may affect the basis, we model the relationship between the county basis level and explanatory variables that represent transportation cost, grain production and consumption, and stocks. Transportation cost is represented by county-level infrastructure variables: highway, rail, and commercial port access, as well as distance to the nearest delivery (Chicago or Peoria) and the Gulf grain markets. County access variables are constructed by measuring the gridded Euclidean distance to the nearest highway, rail line, and water ports based on data layers provided by National Transportation Atlas Database (NTAD), while the respective market distance variables are measured as the Euclidean distance from the county centroid. Grain production is represented by the average county-level corn yield from 2001-2006 according to the National Agricultural Statistics Service (NASS), while consumption is signified by grain consuming animal units and ethanol plant access. Animal units are calculated at the county level based on the Maryland Department of Agriculture's animal unit equivalencies and the 2007 Census of Agriculture, while ethanol plant access is based on information provided by the Renewable Fuels Association (Breneman and Nulph 2010). Variables used to represent commodity storage include elevator access, as well as state-level on- and off-farm stocks. County-level elevator access is calculated in a similar way to the transportation access variables, while on- and off-farm stocks are interpolated based on quarterly USDA Grain Stocks reports.

III. The Model

In this paper, we consider a single-period model of basis determination to concentrate on the importance of its spatial distribution. In the literature, the location i basis b_i is generally estimated as a function of factors X_i that affect the local market, i.e. $b_i = f(X_i)$. The basis is similarly composed at location j. At any time, if the cash price for corn follows the Law of One Price, the equilibrium relationship between the basis in the two locations is

$$b_i \geq b_j - s_{ij}$$

for $b_i \leq b_j$. The basis at location i can be no weaker than the basis at location $j \in J$ less transportation costs, or riskless profits are earned by buying corn at i and shipping it to j. Therefore, if transportation costs are constant, any local shock that sufficiently increases b_j, or decreases b_i, causes the neighboring basis to adjust in the same direction. In effect, a shock to X_j is transmitted to b_i directly through b_j. To verify whether this is the case empirically, we modify the basis relationship to properly account for potential neighborhood effects, using $b_i = f\left(X_i, b_j\right)$. Of course, this relationship also runs in the other direction—the basis is simultaneously determined at the two locations. Generalizing to the multiple neighbor case, we use a standardized dummy variable w to identify each location in J that we consider to be a neighbor of location i. In addition, w is weighted by inverse of the distance between locations, d_{ij}, which has the effect of making closer neighbors more likely to influence the local basis. To make it clear that the basis is the dependent variable in our analysis, we apply the standard notation, i.e., y_i. In scalar terms, the deterministic portion of the regression relationship is

$$y_i = \rho \sum_{j=1}^{J} w y_j + \sum_{k=1}^{K} \beta_k X_{i,k} \ , for$$

$$w = \frac{1}{d_{ij}} \times \frac{C_{ij}}{\sum_{j=1}^{J} C_{ij}} \ where \ \ C_{ij} = \begin{cases} 1 \ if \ j \ is \ a \ neighbor \ to \ i \\ 0 \ otherwise \end{cases} .$$

The standardization procedure averages the neighboring basis levels. As a result, the ρ coefficient measures the degree to which the average basis at neighboring locations influences the location i basis. As usual, the relationship between the factors X_i and the local basis is provided by the coefficients in β_k. The spatial model can be written more conveniently using matrix notation. Stacking basis observations in $J \times 1$ vector y, and local factors in $J \times k$ matrix X, we have

$$y = \rho W y + X\beta + u \, .$$

This model is commonly referred to as the "spatial autoregressive" or "spatial lag", because lagged values of the dependent variable enter into the right-hand side, although, Anselin (1992) terms this model the mixed spatial autoregressive model because it contains additional explanatory variables besides the autoregressive term. The weighting matrix W is the matrix analog of the dummy w, serving to identify neighbors and standardize their basis values. Standardization is accomplished by forcing the diagonal elements of W to zero, and using w as the off-diagonal for all neighboring locations. By construction, the rows of W sum to unity. The random term u follows the normal assumptions. Because they are jointly determined, the parameters of the spatial lag model are usually estimated by either maximum likelihood or instrumental variables methods.

If u is also correlated over space, the spatial model can be enhanced by adding an autoregressive error term. Let

$$u = \lambda W u + \varepsilon\,.$$

The resulting specification is called the spatial ARAR model, since it includes two autoregressive terms. The ARAR reduces to the spatial error model when the error term alone is lagged over space. We verify that the basis is spatially autocorrelated using a Moran's I test, and choose between OLS and these three potential spatial models using Anselin's selection criteria (Anselin, et al., 1996).

We specify W so that a county's neighbors include the nearest twenty counties. Although this is an arbitrary step, using different numbers of neighbors to create W did not meaningfully alter our results. W is distance-weighted, because the local basis is more likely to be affected by the basis at a closer rather than a more distant neighbor. Including W in the basis models allows the researcher to test whether the basis is simultaneously determined over space, accounts for spatially correlated unobserved characteristics, and filters out spatial effects to produce consistent estimates for the way the other explanatory variables affect the local basis. We define X to include the explanatory variables in table 1. In the next section, we show whether these explanatory variables are associated with a "strengthening" or "weakening" basis. When the local cash price increases (decreases) relative to the futures price, the local basis is said to strengthen (weaken). Figure 2 depicts these situations, graphically.

[FIGURE 2 ABOUT HERE]

IV. Results and Discussion

Descriptive Statistics

Table 1 shows the mean and standard deviation for all variables in the analysis. For August 2007, the monthly average basis is –16.4 cents/bushel. County-level corn demand is measured by ethanol capacity with a mean of roughly 4,000 gallons per year and livestock consumption with a mean of 86,000 animal units. The mean state-level storage for on-farm and off-farm facilities is approximately 100 thousand bushels, and the county-level historical yield is approximately 134 bushels per acre.

[TABLE 1 ABOUT HERE]

Table 1 also includes a number of distance variables that measure county-level market access. The access measures are calculated as the mean distance in kilometers to each amenity across the county. For example, in the average county, the mean distance to the nearest grain elevator is 40.25 kilometers. Likewise, the access value is 8.9 kilometers for state highways, 6.8 kilometers for rail lines, 32.11 kilometers for interstate exchanges, and 155 kilometers for ports. In addition, the mean distance to the delivery and the Gulf of Mexico markets, as measured as distance from county centroid, is 5.96 and 11.69 decimal degrees, respectively.

Spatial Autocorrelation and OLS

The Moran's I test in figure 3 shows that the basis is highly spatially autocorrelated, confirming the apparent clusters in figure 1. As shown in figure 3, strong basis counties are very likely to have neighbors with similarly strong basis values, while weak basis counties are most likely surrounded by likewise weak basis areas. Spatially clustering, however, does not necessarily indicate simultaneous determination. We use regression analysis to estimate the determinants of the basis, and discover whether the basis is characterized by spatial spillovers.

[FIGURE 3 ABOUT HERE]

Table 2 shows our estimates for the different basis model specifications. Coefficients in the table represent the linear correlative relationship between the dependent variable and the basis. Clearly, OLS contrasts with the spatial models, reporting county-level ethanol capacity, distance to delivery and gulf markets, and stocks associated with the basis level. However, the sign on the OLS distance to market coefficients is dubious: the basis rises as the distance increases, which conflicts with our intuition. All else equal, the basis should grow weaker the further a county lies from a terminal market, due to transportation costs. Also, OLS fails to discern a relationship between corn yield and the basis. Finally, a Moran's I test on the OLS regression residuals indicates that the explanatory variables do not sufficiently account for the spatial clustering: because OLS is biased, the basis is more appropriately modeled with a spatial specification.

[TABLE 2 ABOUT HERE]

Spatial Model Selection and Results

The lower portion of table 2 displays our model selection criteria. According to Anselin, et al. (1996), the proper spatial model should be selected via a series of Lagrange Multiplier (LM) tests: choose either lag or error if it's basic LM test is significant, while the other is not; if both basic tests are significant, choose the model with the highest significant robust test value; finally, select the ARAR if the SARMA test is significant. The LM tests in table 2 indicate that the spatial lag model is preferred to the spatial error, although the ARAR is favored by specification tests since it captures the spatially clustered nature of both the basis as well as unobservable factors.

According to the ARAR model, the basis is simultaneously determined at the 7% significance level. We interpret the positive lagged basis coefficient to mean that factors impacting the local basis are transmitted to neighboring basis levels, too. The positive lagged error term confirms that unobservable factors have a similar impact on the basis among neighboring counties. Once the appropriate spatial filters are applied, the degree to which other factors affect the local basis can be estimated consistently.

Beyond the spatial terms, the ARAR model reports that elevator access, port access, grain-consuming animal units, and corn yield are correlated with the local basis. Better county-level access to a grain elevator may be associated with a lower basis for two reasons: elevators may locate in areas with cheaper corn, and they may be able to exercise market power as a monopsonist. The positive sign on the port access term is puzzling, because it indicates that counties further away from a port facility are more likely to have a stronger basis. Although, looking at figure 1, we see that the outlying counties across the plains and the mid-Atlantic do appear to demonstrate a stronger basis in August 2007. As expected, more grain consuming animal units are associated with a stronger average county-level basis: higher demand for grain leads to a higher cash price, narrowing the basis. Likewise, a higher corn yield is inversely correlated with the basis: as the supply curve shifts out, the local price falls.

V. Conclusion

Spatial tests confirm that the county-level corn basis data is highly clustered and best described by a spatial framework. We estimate a spatial model and find that the corn basis is characterized by spillovers: local supply and demand factors affect the local basis, but also impact the basis in neighboring counties. We theorize that this is observed due to the spatial arbitrage relationship between grain prices at neighboring locations. In addition, we find that unobserved basis determinants are also spatially correlated.

Although our results are novel, our approach suffers from some limitations. Most notably, our data represent a single cross-section. In reality, the basis changes over time;

a richer, dynamic model would be more appropriate to uncover causal basis determinants. Indeed, prior research has found that the importance of certain factors changes over the crop year. In future work, we intend to use a spatial model to study the basis panel, to better measure elements that affect the basis, and to determine if such an approach can improve basis forecasting.

Tables

Table 1: August 2007 County Data Summary

	Mean	**Std. Dev.**
Basis (futures - cash, in cents/bushel)	– 16.40	20.91
Ethanol (county capacity in thousands of gallons)	4.08	5.86
Animal Units (in thousands)	85.73	122.98
State On-farm Stocks (hund. thous. bushels)	0.92	0.92
State Off-farm Stocks (hund. thous. bushels)	1.06	1.05
Average Yield over 2001-2006 (in bushels per acre)	134.42	31.54
Distance Measures		
Elevator (mean county dist to amenity, km)	40.25	34.45
Highways (mean county dist to amenity, km)	8.89	5.27
Rail (mean county dist to amenity, km)	6.84	3.36
Interstate Exchange (mean county dist to amenity, km)	32.11	29.92
Ports (mean county dist to amenity, km)	154.66	162.65
Delivery (decimal degrees)	5.96	3.60
Gulf (decimal degrees)	11.69	3.15

Source: Varies; see Data section.

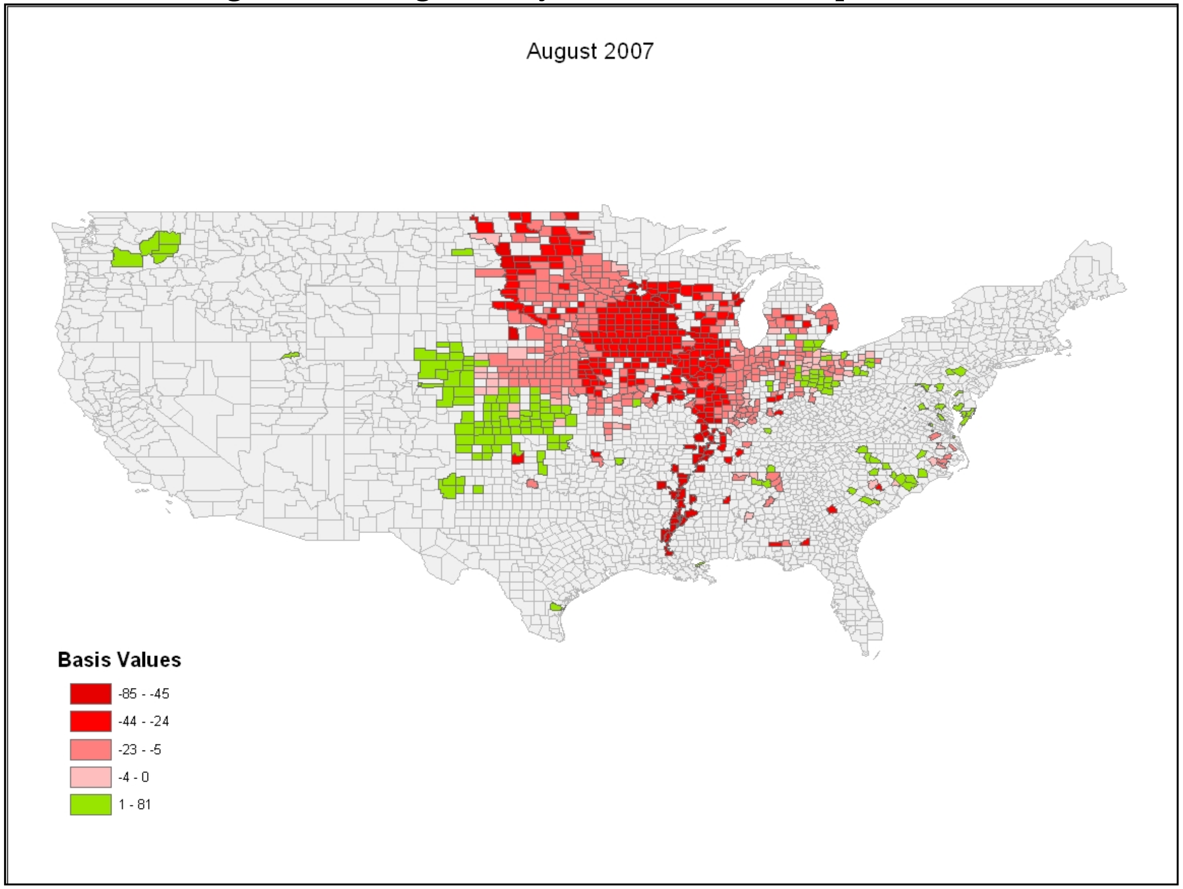
Table 2: County-Level Basis Models for August 2007

	OLS	Spatial Lag	Spatial Error	ARAR
Intercept	-33.78***	6.86**	-1.13	-0.08
	(4.99)	(3.25)	(12.66)	(9.95)
Elevator Access	-4.14**	-2.42*	-5.42***	-5.4***
	(2.1)	(1.33)	(1.89)	(1.86)
Ethanol Capacity	-30.24**	9.52	24.95	20.38
	(14.64)	(9.29)	(18.98)	(18.09)
Highway Access	-86.14***	-25.62***	-12.16	-12.9
	(13.58)	(8.63)	(9.58)	(9.62)
Rail Access	-37.94	-5.79	1.43	2.29
	(23.51)	(14.9)	(16.77)	(16.83)
Port Access	1.05*	0.8**	5.13***	3.99***
	(0.64)	(0.4)	(1.36)	(1.3)
Animal Units	2.49***	1.15***	1.19***	1.19***
	(0.55)	(0.35)	(0.36)	(0.37)
Yield	18.29	-48.8***	-76.06***	-75.06***
	(25.05)	(15.87)	(19.75)	(19.72)
Dist. to Delivery Market	1.7***	0.07	0.16	0.28
	(0.3)	(0.2)	(0.96)	(0.79)
Dist. To Gulf	1.74***	0.02	-0.74	-0.45
	(0.25)	(0.16)	(0.87)	(0.69)
On-farm Stocks	-16.23***	-1.61	-0.08	-0.7
	(1.64)	(1.07)	(2.43)	(2.34)
Off-farm Stocks	8.18***	0.16	-1.05	-0.75
	(1.37)	(0.87)	(1.85)	(1.8)
Lagged Basis	-	0.9***	-	0.29*
		(0.02)		(0.16)
Lagged Error	-	-	0.94***	0.88***
			(0.02)	(0.05)
Observations	785	785	785	785
R-Squared	38%	-	-	-
Residual Moran's I	0.13***	-	-	-
Log Likelihood	-	-2997	-2994	-
LM Test	-	1698***	1568***	-
Robust LM Test	-	192***	62***	-
SARMA	-	-	-	1761***

***p<0.01, **p<0.05, *p<0.1

Figures

Figure 1: Average County-level Basis in Cents per Bushel



Source: CashGrainBids.com

Figure 2: Basis behavior

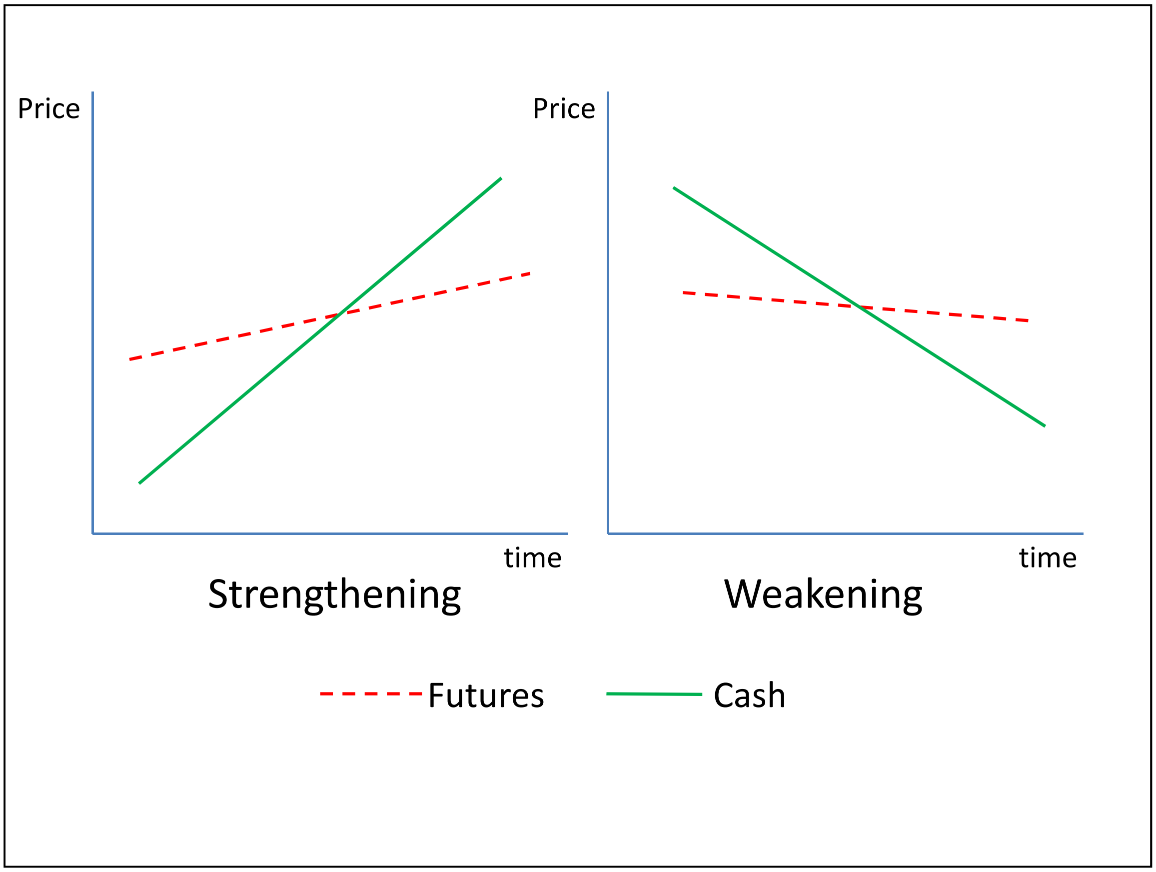
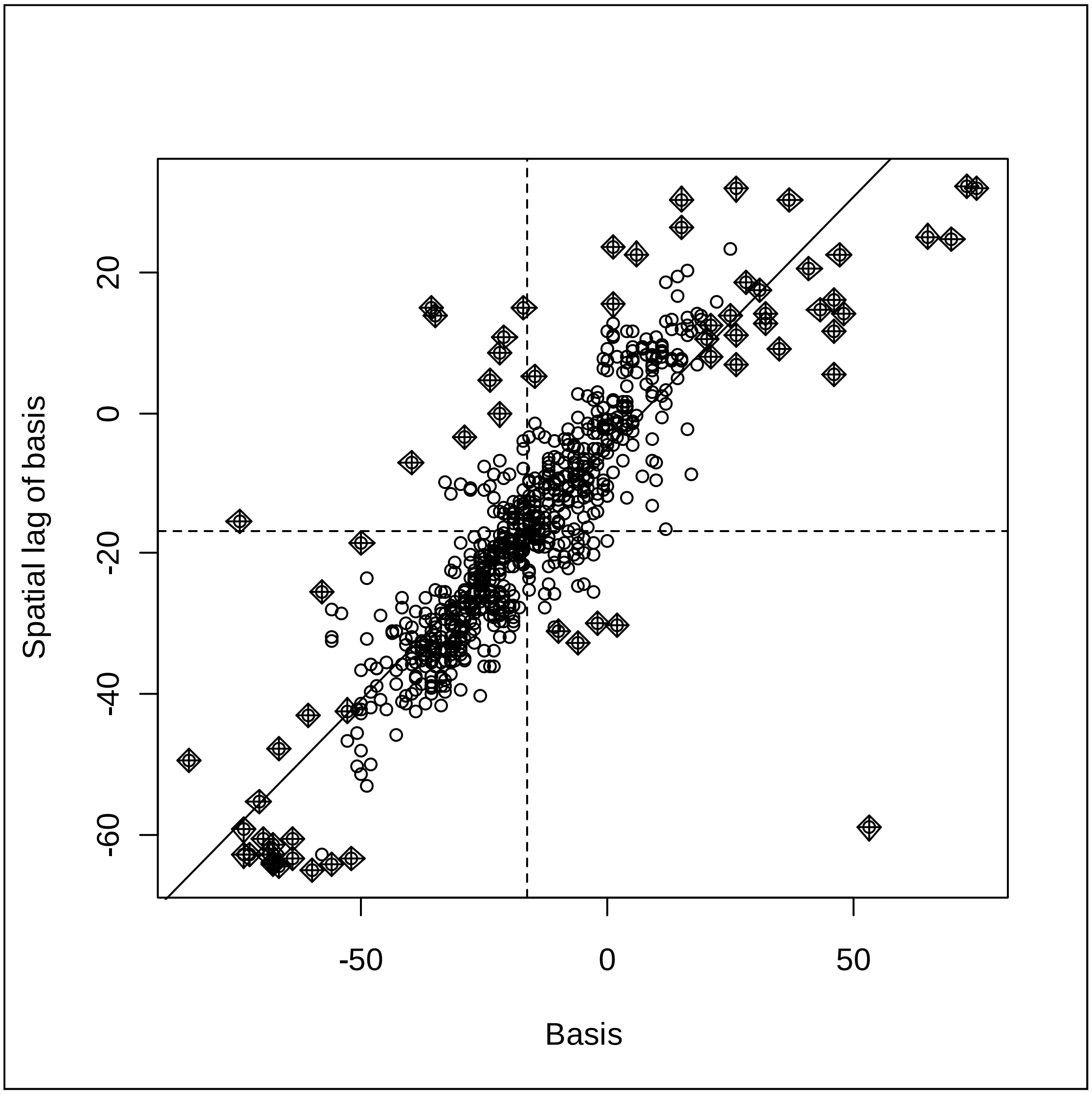


Figure 3: Moran's I for County-Level Basis: August 2007 Avg.



Moran's I of Basis: 0.716***

References

Anselin, L. "Spatial Data Analysis with GIS: An Introduction to Application in the Social Sciences." National Center for Geographic Information and Analysis (NCGIA). Technical Report 92-10. (1992).

Anselin, L., A. Bera, R. Florax, and M. Yoon (1996) "Simple Diagnostic Tests for Spatial Dependence" *Regional Science and Urban Economics* 26: 77-104.

Breneman, V., and D. Nulph. "Analysis of Renewable Fuels Association Plant Data." USDA Economic Research Service. Washington, D.C., (2010).

Garcia, P., and D. Good. "An Analysis of Factors Influencing the Illinois Corn Basis, 1971-1981." *NCR-134 Conference on Applied Commodity Analysis, Forecasting and Market Risk Management*, (1983), Des Moines, IA.

Jiang, B., and M. Hanyenga (1997) Corn and Soybean Basis Behavior and Forecasting: Fundamental and Alternative Approaches. Chicago, IL.

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

Lewis, D. A., T. H. Kuethe, M. R. Manfredo, and D. R. Sanders (2010) Uncovering Dominant-Satellite Relationships in the U.S. Soybean Basis: A Spatio-Temporal Analysis. St. Louis, MO.

Manfredo, M. R., and D. R. Sanders (2006) Is the Local Basis Really Local? St. Louis, MO.

McKenzie, A. M. "The Effect of Barge Shocks on Soybean Basis Levels in Arkansas: A Study of Market Integration." *Agribusiness* 21, no. 1(2005): 37-52.

Tilley, D. S., and S. K. Campbell. "Performance of the Weekly Gulf-Kansas City Hard-Red Winter Wheat Basis." *American Journal of Agricultural Economics* 70, no. 4(1988): 929-935.