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

Uncovering Dominant-Satellite Relationships in the U.S. Soybean Basis: A Spatio-Temporal Analysis

by

Daniel A. Lewis, Todd H. Kuethe,
Mark R. Manfredo, and Dwight R. Sanders

Suggested citation format:

Lewis, D. A., Todd H. Kuethe, Mark R. Manfredo, and Dwight R. Sanders. 2010. "Uncovering Dominant-Satellite Relationships in the U.S. Soybean Basis: A Spatio-Temporal Analysis." 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].

Uncovering Dominant-Satellite Relationships in the U.S. Soybean Basis: A Spatio-Temporal Analysis

Daniel A. Lewis

Todd H. Kuethe

Mark R. Manfredo

Dwight R. Sanders*

Paper presented at the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management

St. Louis, Missouri, April 19-20, 2010

*Daniel A. Lewis is a graduate student in the Morrison School of Agribusiness and Resource Management at Arizona State University. Todd H. Kuethe is an economist with the USDA's Economic Research Service. Mark R. Manfredo is an Associate Professor in the Morrison School of Agribusiness and Resource Management at Arizona State University. Dwight R. Sanders is a Professor at Southern Illinois University, Carbondale.

The views expressed are those of the author(s) and should not be attributed to ERS or USDA.

Uncovering Dominant-Satellite Relationships in the U.S. Soybean Basis:

A Spatio-Temporal Analysis

Practitioner's Abstract

Time series analysis shows that local soybean basis levels have some tendency to follow or be determined by the basis levels at export locations (Toledo and U.S. Gulf). Processing centers tend to show the most independence in basis discovery. Spatial modeling shows that each local basis produces a "spillover" and impacts neighboring basis levels. The spatial linkages are greatest during the spring and tend to be the lowest during fall. The results suggest that soybean basis discovery may be concentrated at export locations within the U.S. marketing system. Moreover, these dominant-satellite relationships are strongest during the spring season. Market practitioners may utilize this information when forming expectations for basis levels during the marketing year.

Key Words: Soybeans, Basis, Causality, Spatial Relationships

Introduction

Basis values, the difference between cash and futures price, play an important role in guiding commodities through the supply chain (Tiley and Campbell, 1988; Tomek and Robinson, 1990). It has long been believed that the basis is determined, to a large extent, by local factors, such as storage capacity, quality differentials, and transportation costs. However, a number of recent studies indicate that the prices offered at one location may also reflect basis values at other locations. That is, the basis may not be entirely local as some locations provide a source of market information used to determine the basis at other locations. This is often referred to as a dominant-satellite relationship. Previous studies have identified dominant-satellite relationships for basis values in corn (Manfredo and Sanders, 2006) and soybeans (McKenzie, 2005).

This study examines soybean basis relationships between market locations across both time and space. A time series analysis examines to what degree certain markets provide information to other locations. Following previous work by Manfredo and Sanders (2006) and McKenzie (2005), we test for causal relationships through time for a number of market definitions, including export terminals, interior river locations, processing centers, and interior markets. The analysis provides relevant information on market leaders or price discovery points.

We also address the degree to which markets simultaneously share pricing information across space in a spatial econometric framework. The analysis indicates the degree to which the covariation of basis levels between observations at different locations is subject to spatial ordering. The spatial analysis therefore demonstrates to what degree changes in basis levels at one location induce spillover effects, impacting neighboring locations.

Methods

Time Series Analysis
Granger Causality provides one approach to identify whether markets share information. In a Granger Causality framework, market *X* is said to *Granger cause* market *Y* if market *X* provides valuable information when forecasting market *Y*. The method has been used to test corn basis relationships among major export markets and interior locations (Manfredo and Sanders, 2006), as well as the relationship between spot and futures prices for live cattle (Koontz and Hudson, 1990; Oellerman and Farris, 1985). The causality test is based on the equation:

$$(1) \qquad y_t = \alpha + \sum_{i=1}^{m} \lambda_i y_{t-i} + \sum_{j=1}^{n} \theta_j x_{t-j} + \omega$$

where y_t is the basis value at time *t* in market *Y*, and *m* and *n* are the optimal lag lengths for y_t and x_t, respectively. The null hypothesis that *X* does not Granger cause *Y* is examined by a Wald test on the restriction $\theta_j = 0 \; \forall \; j$ (Hamilton, 1994). When the null hypothesis is rejected, the test suggests that market *X* plays a role in the determination of the basis at market *Y*.

Spatial Analysis
The contemporaneous spatial spillover for each period is estimated through spatial regression analysis. The spatial lag model accounts for observed spatial dependence in the dependent variable by including spatially weighted values of the dependent variable on the right hand side of the equation (Anselin, 1988). The model takes the form:

$$(2) \qquad \begin{gathered} y = \alpha + \rho W y + \varepsilon \\ \varepsilon \sim N(0, \sigma^2 I) \end{gathered}$$

where *y* is an N x 1 vector of basis values at each location n=1,...,N. The neighbor relationships are defined by the exogenous N x N matrix *W*. The spatial weights matrix defines a relevant neighborhood for each observation with zeroes on the diagonal and non-zero off-diagonal elements w_{ij} which indicate that observations *i* and *j* share a spatial relationship. The weights matrix is a row-normalized five nearest neighbors specification. The spatial lag term *Wy* is therefore the average basis level at the five nearest neighbors, with each neighbor receiving an equal weight. Following standard regression procedures, the disturbance term ε is expected to follow a mean zero, constant variance i.i.d. process.

The sign of the estimated spatial parameter indicates the nature of the spatial process. For example, when $\rho > 0$, the model suggests that the basis is, on average, positively impacted by neighboring values. Unlike temporal lags in time series analysis, the spatial lag characterizes simultaneous feedback in space such that a home is impacted by the value of its neighbor while it simultaneously impacts the value of its neighbors. Because of this, the spatial lag term *Wy* is correlated with the disturbance term introducing biased estimates of regression parameters using ordinary least squares. We therefore estimate the spatial lag coefficient via maximum likelihood (Anselin, 1988).

A distinct feature of the spatial lag model is that it captures "global spillovers" attributed to what Anselin (2003) calls the *spatial multiplier*. The spatial multiplier refers to the fact that all locations are related through higher-order neighbors. Thus, a change in the basis level at one location directly impacts its neighbors, as well as its neighbor's neighbors, and so on. The relationship can be seen when the spatial lag model is expressed in the reduced form:

(3)
$$(I-\rho W)y = \alpha + \varepsilon$$
$$y = (I-\rho W)^{-1}\alpha + (I-\rho W)^{-1}\varepsilon$$

When the assumptions of disturbance term hold, the spatial spillover process reduces to $(I - \rho W)^{-1}\alpha$, where $(I-\rho W)^{-1}$ is the spatial multiplier. Kim et al. (2003) show that when the weights matrix is row-normalized, as we have done, the expected value reduces to:

(4)
$$E[y] = \frac{1}{(1-\rho)}\alpha$$

In sum, the spatial regression analysis provides an estimate of the impact of a change in the basis level in one location on the basis at all other locations. The impacts are most pronounced for immediate neighbors, yet the effects "spillover" to all locations through higher-order neighbor relationships.

Data

To keep the Granger Causality analysis tractable, the data is selected for thirteen markets shown in Figure 1. The sample locations include the major export terminals of Louisiana Gulf and Toledo, OH (Lucas County), interior river locations of St. Louis, MO (St. Louis County) and Peoria, IL (Peoria County), a major soybean processing facility in Bellevue, OH, and several interior locations including Omaha, NE, Raleigh, NC, Central Illinois (Champaign County), Northwestern Iowa (Buena Vista County), Central Iowa (Hamilton County), Eastern Iowa (Black Hawk County), Central Kansas (Pawnee County), and Southeastern Indiana (Decatur County). The data is comprised of weekly (Wednesday) nearby basis values obtained from cashgrainbids.com.[1] For the Louisiana Gulf, Omaha, and Raleigh locations, the basis data provided by cashgrainbids.com is USDA-AMS data. For all other locations except for Bellevue, OH, the nearby basis used is an average of the basis reported at individual elevators within the county noted, with anywhere from two to eight elevators from each county composing the average[2]. Each time series spans January 2003 – November 2009, providing 357 weekly observations of the basis for each location.

The data needs for the spatial analysis are slightly different than those for the Granger Causality tests. For each of the thirteen markets defined above, the sample size is expanded to ensure a sufficient number of observations and degree of spatial variability. In particular, the spatial analysis draws individual elevator-level point observations of the basis from the previously

[1] In the event of missing observations for Wednesday, we selected from nearby data with the given priority: Tuesday, Thursday, Monday, or Friday.
[2] For St. Louis, MO, the basis data are drawn from an individual elevator in St. Louis, County.

defined markets on a daily basis, as well as daily nearby basis observations from individual elevators in adjacent counties (see Figure 2). For each individual elevator, these data are then aggregated to mean values for four periods in the crop year over the years 2006 to 2009. The four periods are defined as spring (April – June), summer (July – August), fall (September – November), and winter (November – March).

Results

Time Series Analysis

Augmented Dickey-Fuller test statistics show that each basis series is stationary in levels. Following the procedure of Beveridge and Oickle (1974) the optimal lag length for the Granger Causality test is found by estimating equation (1) for all lag combinations i=1,…,12 and j=1,….,12 and using the lag structure that minimizes Akaike Information Criterion (Akaike, 1974). In addition, we test for heteroskedasticity using White's test and apply White's consistent covariance estimator where necessary.

The Granger Causality results are reported in Table 1. Table 1 shows the information flow from the row to the column and vice-versa. For example, considering Toledo (row) and Omaha (column), the → symbol signifies that Toledo (x) leads Omaha (y), with the rejection of the null hypothesis that $\theta_j = 0 \;\forall\; j$ at the 1% level of confidence (equation 1). However, when the relationship is reversed, Omaha (x) does not lead Toledo (y) since there is a failure to reject the null hypothesis of $\theta_j = 0 \;\forall\; j$ at the 1% level. Therefore, it can be said that the direction of causality is from Toledo to Omaha. Considering Omaha (row) and C IL (column), the ← symbol signifies that Omaha (x) does not lead C IL (a failure to reject the null hypothesis at the 1% level), but C IL (x) does lead Omaha (y) since the null hypothesis of $\theta_j = 0 \;\forall\; j$ is rejected at the 1% level. Hence the direction of causality is from C IL to Omaha. The ↔ symbol signifies two-way or simultaneous causality significant at the 1% level. For example, in the case of Peoria (row) and Raleigh (column), the null hypothesis that Peoria (x) does not cause Raleigh (y) is rejected at the 1% level. When the relationship is reversed the null hypothesis that Raleigh (x) does not cause Peoria (y) is also rejected, thus suggesting two-way or simultaneous causality. A zero (0) in any of the row/column combinations suggests that there is a failure to reject the null hypothesis in both directions, hence neither market leads the other.

Table 2 summarizes the Granger Causality results that are presented in Table 1 and provides an indication of the connectivity of the individual markets. St. Louis, Toledo, and the Gulf have the greatest amount of leading information (6, 5, and 4 respectively), while the Omaha market (6) exhibits the largest degree of lagging information. In addition, Raleigh and Eastern Iowa (E IA) demonstrate the greatest number of two-way information flows (↔) with 9 each. The Bellevue market (site of a major soybean processing facility), Northwest Iowa (NW IA), Central Illinois (C IL), and Central Kansas (C KS) appear to show the least amount of connectivity as indicated by the largest number of "0's" at 3 each.

Indeed, one can also observe market dynamics across the four market categories (major export terminal, interior river, processing facility, and interior locations) by examining relationships based on these average linkages. The export markets appear to have the greatest amount of influence on average. That is, the export markets (Gulf and Toledo) have the greatest amount of forward linkages considering both leading and simultaneous relationships. Further, Omaha exhibits a lagging relationship, as demonstrated by the greatest number of lagging relationships, as well as independent relationships. Finally, the interior river and interior markets exhibit the greatest amount of combined forward and backward price transmissions. In sum, the results suggest that export markets tend to display dominate relationships with other markets displaying satellite behavior.

Spatial Analysis

Spatial dependence is the result of a systematic pattern in observed basis values or price bids over space. Spatial dependence may be the result of shared information among neighbors as well as similar geographic properties at each location, such as access to transportation or natural geographic features (McNew, 1996). The classic measure of spatial dependence is Moran's I (Moran, 1950). Moran's I measures the degree of spatially weighted deviations from the global mean. Moran's I is expressed:

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2} \quad (5)$$

where N is the number of observations, w_{ij} is a neighbor definition which is strictly positive when observations i and j share a meaningful relationship in space, and $\bar{X}$ is global mean value of the random variable X. Moran's I is weakly bounded by – 1 and + 1, where – 1 indicates perfect negative spatial dependence and + 1 indicates perfect positive spatial dependence. The measure takes the value of 0 for spatially independent variables. Hypothesis testing is conducted using a bootstrap procedure based on random draws with replacement. We define the spatial weights as equally weighted values at the five nearest neighbors.

The Moran's I test results are reported in Table 3. The results demonstrate that the basis levels are positively spatially autocorrelated at a statistically significant level. The Moran's I follows a seasonal pattern with peak levels during the spring months (April – June) and minimum values during the fall months (September – November) with the exception of 2007. The nearby basis during the spring months are for either the May or July futures contract which reflects the old crop. Thus these results indicate that the degree of spatial association is positively related to uncertainty throughout the drawdown period of the old crop that is stored. As well, this is a time that new crop is being planted, thus a considerable amount of uncertainty exists in the market. Indeed, the results indicate that the degree of spatial association is positively related to uncertainty throughout the growing season.

The spatial multiplier estimates are presented in Table 4 and in Figure 3. The vertical axis shows the estimated spatial multiplier in each period, and the figure includes a two-period rolling average and a linear trend. The estimated spatial spillover in each year is the highest during the spring months. This would suggest that price changes at each location carry the greatest global effect during the periods of the greatest uncertainty. However, the period with the minimum spillover is not consistent across years. In addition, the two-period moving average does not contain a discernable trend. In sum, the results suggest a strong degree of simultaneous information sharing over space.

Summary and Conclusions

Our analysis demonstrates the degree to which soybean markets share information in determining basis levels over both time and space. We define four market categories: major export terminals, interior river locations, soybean processing facilities, and interior locations. The evaluation is conducted through time series analysis and spatial econometrics. The time series analysis consists of Granger causality tests of weekly basis values in each market over the period January 2003 – November 2009. The results suggest that export markets lead the price discovery at the other locations

The spatial econometric analysis, on the other hand, examines the degree of simultaneous spatial spillovers in observed basis over the period January 2006 – December 2009. The spatial lag specification addressed four periods in the soybean crop year: planting, summer, harvest, and winter. The model specification directly ties each location to its neighbors (via the spatial weights matrix), yet all observations are linked through higher order neighbor relationships (i.e., neighbors of neighbors and so on). The spatial multiplier therefore models the expected change in the basis at all locations given a change in the basis at a single location. The maximum value was observed during the spring months, most associated with old crop inventory drawdown as well as planting of new crop bean. Thus, price changes at each location carry the greatest global effect during the periods of the greatest uncertainty.

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dorddrecht: Kluwer Academic Publisher.

Anselin, L.(2003). "Spatial Externalities, Spatial Multipliers, and Spatial Econometrics" *Regional Science Review* 26 (2): 147–152.

Beveridge, S. and Oickle, C.(1994). "A Comparison of Box-Jenkins and Objective Methods for Determining the Order of a Non-Seasonal ARMA Model," *Journal of Forecasting* 13: 419-434.

CME Group.(2010). " Soybean Futures" Retreived May 1, 2010 from: http://www.cmegroup.com/trading/commodities/grain-and-oilseed/soybean_contract_specifications.html

Hamilton, J.D. (1994). *Time Series Analysis*. New Jersey; Princeton University Press,

Haigh, M.S. and Bessler, D.A. (2004). "Causality and Price Discovery: An Application of Directed Acyclic Graphs," *Journal of Business* 77: 1099-1121.

Kim W.K., Phipps T., Anselin L. (2003). "Measuring the Benefits of Air Quality Improvement: A Spatial Hedonic Approach" *Journal of Enviormental Economics and Management* 45: 24-39.

Koontz, S.R., Garcia, P. and Hudson, M.A. (1990). "Dominant-Satellite Relationships Between Live Cattle and Cash Futures Markets," *The Journal of Futures Markets* 10: 123-136.

McKenzie, A. M. (2005). "The Effects of Barge Shocks on Soybean Basis Levels in Arkansas: A Study of Market Integration," *Agribusiness: An International Journal* 21: 37-52.

McNew, K. (1996). "Spatial Market Integration: Definition, Theory and Evidence" *Agricultural and Resource Economics Review* 25: 1–11.
Moran, P. (1950). "Notes on Continuous Stochastic Phenomena" *Biometrika* 37: 17–33.
Manfredo, M. R., and Sanders D. R. (2006). "Is the Local Basis Really Local?" Proceedings of the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management. St. Louis, MO. [http://www.farmdoc.uiuc.edu/nccc134].
Ollermann, C.M. and Farris, P.L. (1985). "Futures or Cash: Which Market Leads Beef Cattle Prices?" *The Journal of Futures Markets* 5: 529-538.
Pindyck, R.S. and Rubinfeld, D.L. (1998). "Econometric Models and Forecasts – Fourth Edition", New York, NY: Irwin McGraw-Hill.
Tilley, D.S. and Campbell, S.K. (1988). "Performance of the Weekly Gulf-Kansas City Hard-Red Winter Wheat Basis", *American Journal of Agricultural Economics* 70: 929-935.
Tomek, W.G. (1993). "Dynamics of Price Changes: Implications for Agricultural Futures Markets," Research Frontiers in Futures and Options Markets: An Exchange of Ideas, *Proceedings from the Symposium in Recognition of Thomas A. Hieronymus,* Office for Futures and Options Research, University of Illinois at Urbana-Champaign, pp. 45-55.
Tomek, W.G. and Robinson, K.L. (1990). "Agricultural Product Prices – Third Editon". N.Y.: Cornell University Press.

Table 1. Granger Causality Results[a,b]

	Gulf	Toledo	St. Louis	Peoria	Bellevue	Omaha	Raleigh	C IL	NW IA	C IA	E IA	C KS	SE IN
Gulf		↔	0	↔	→	→	→	↔	↔	↔	↔	↔	→
Toledo			↔	↔	→	→	↔	→	→	↔	↔	←	→
St. Louis				0	↔	↔	→	→	↔	→	→	→	→
Peoria					↔	→	↔	→	↔	↔	↔	↔	↔
Bellevue						←	↔	0	0	↔	↔	0	←
Omaha							↔	←	0	←	←	→	↔
Raleigh								↔	↔	↔	↔	←	↔
C IL									0	→	↔	0	↔
NW IA										↔	↔	→	↔
C IA											↔	←	→
E IA												↔	→
C KS													0
SE IN													

[a] Results are interpreted from row to column. For example, for Peoria (row) and Raleigh (column), there is a simultaneous causality relationship in that there is a rejection of the null hypothesis that Peoria does not lead Raleigh, and a rejection of the null that Raleigh does not lead Peoria, both at the 1% level (↔). Similarly, for Peoria (row) and Omaha (column), Peoria leads Omaha (→) as there is a rejection of the null that Peoria does not lead Omaha, but a failure to reject the null that Omaha does not lead Peoria. For Omaha (row) and C IA (column), C IA is found to lead Omaha (←) at the 1% level (rejection of the null at the 1% level), but Omaha does not lead C IA (failure to reject null). A zero (0) suggests that there is a failure to reject the null in each direction (no causality).

[b]. C IL is Central Illinois, NW IA is Northwest Iowa, C IA is Central Iowa, E IA is Eastern Iowa, C KS is Central Kansas, and SE IN is Southeast Indiana.

Table 2. Summary of Granger Causality Results[a]

	Simultaneous Causality	Lead	Lag	No Causality
Gulf	7	4	0	1
Toledo	6	5	1	0
St. Louis	4	6	0	2
Peoria	8	2	0	1
Bellevue	5	0	4	3
Omaha	4	2	6	1
Raleigh	9	0	3	0
C IL	4	2	3	3
NW IA	7	1	1	3
C IA	7	2	3	0
E IA	9	2	1	0
C KS	3	3	3	3
SE IN	5	1	5	1

[a] C IL is Central Illinois, NW IA is Northwest Iowa, C IA is Central Iowa, E IA is Eastern Iowa, C KS is Central Kansas, and SE IN is Southeast Indiana.

Table 3. Spatial Dependence Test Results [a]

	2006	2007	2008	2009
Winter	0.726	0.674	0.719	0.807
Spring	0.739	0.784	0.847	0.846
Summer	0.739	0.690	0.841	0.841
Fall	0.653	0.744	0.629	0.722

[a] All results are statistically significant at $\alpha \leq 0.01$.

Table 4. Spatial Multiplier Estimates

	2006	2007	2008	2009
Winter	6.671	5.915	6.301	7.856
Spring	7.321	7.536	9.065	9.718
Summer	7.321	5.656	8.952	5.823
Fall	6.701	6.063	4.505	5.823

Figure 1: Market Definitions and Locations

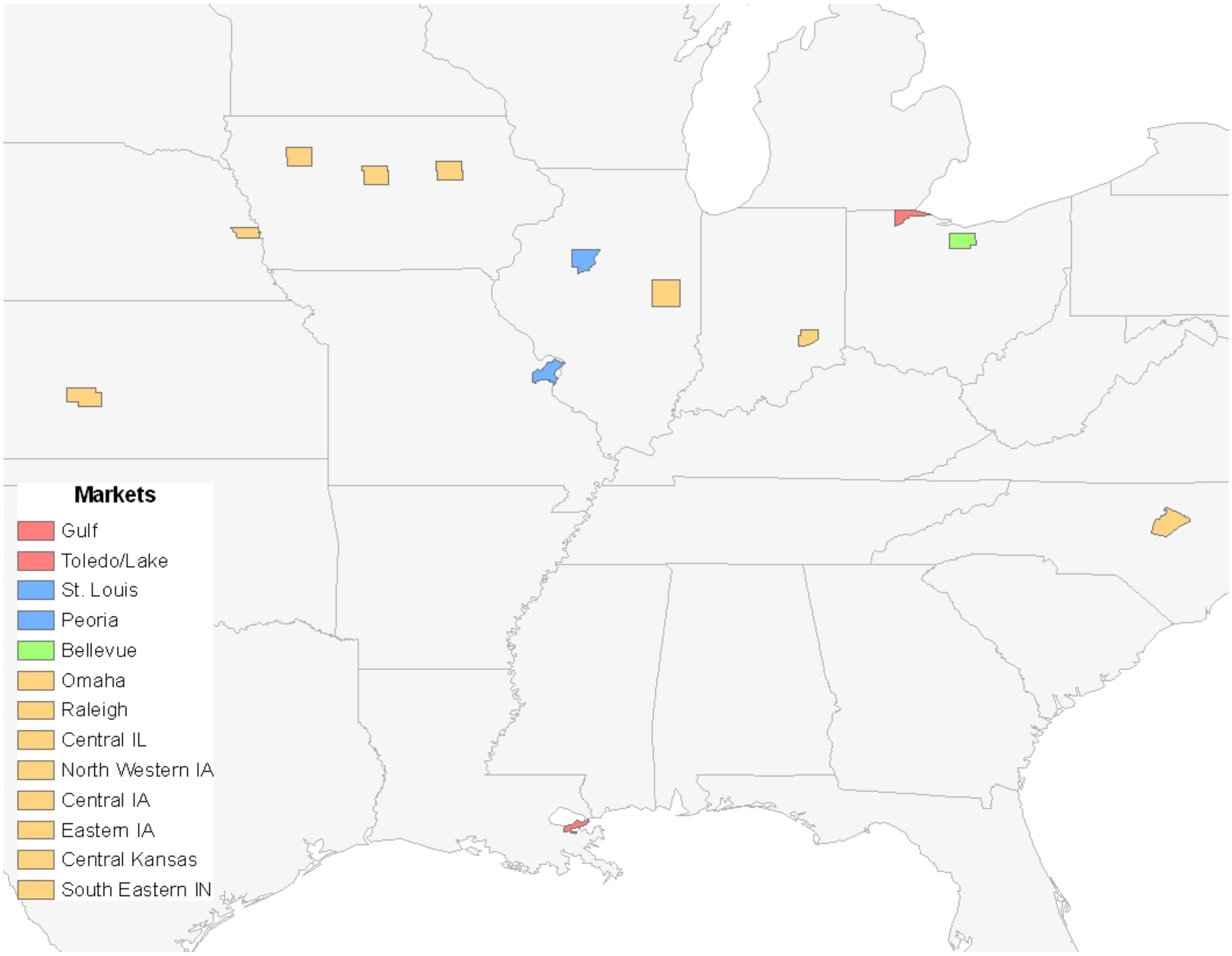

Figure 2: Elevator Locations and Market Definitions

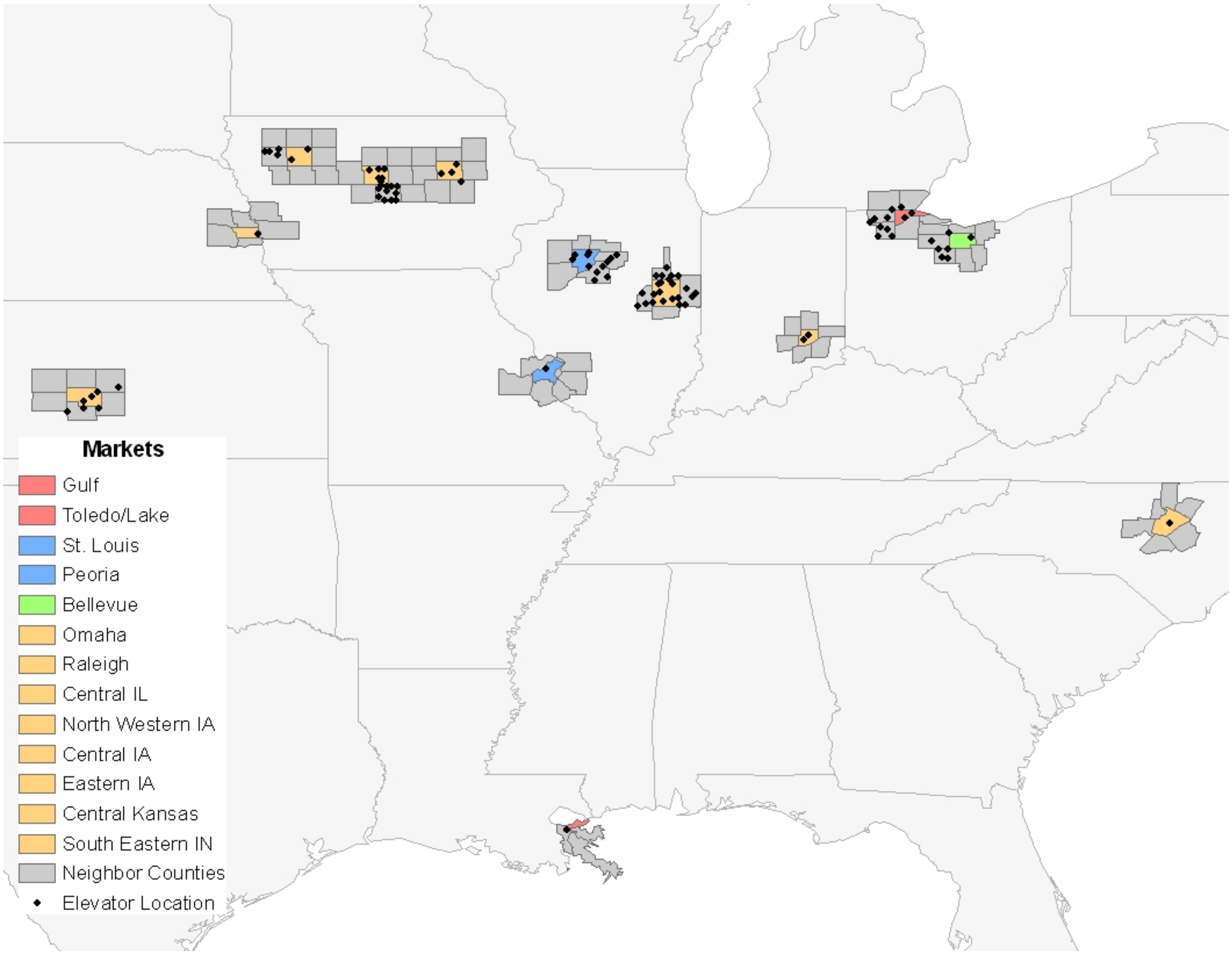

Figure 3 Spatial Multiplier Results