A Follow-Up to Benirschka & Binkley's *"Land Price Volatility in a Geographically Dispersed Market"*: Updates to Data and Methodology

Selected Paper prepared for presentation at the 2016 Agricultural & Applied Economics Association Annual Meeting, Boston, Massachusetts, July 31-August 2

Draft: May 23^{rd}, 2016

Jeffrey S. Young[i], James K. Binkley[i], Raymond J.G.M. Florax[i]

[i] Department of Agricultural Economics, Purdue University (young247@purdue.edu)

A Follow-Up to Benirschka & Binkley's *"Land Price Volatility in a Geographically Dispersed Market"*: Updates to Data and Methodology (DRAFT)

Jeffrey S. Young, James K. Binkley, & Raymond J.G.M. Florax

I. Introduction

I(a): Background

Farmland, the most fundamental of assets in the agricultural economy, is prone to cyclical valuation patterns through time. From the 1960's to the early 1980's, farmland greatly appreciated in value, as commodity prices were on the rise and agriculture was more profitable relative to previous years. However, when the commodity market crashed, so did land prices, leading to a farm financial crisis in the 1980's.

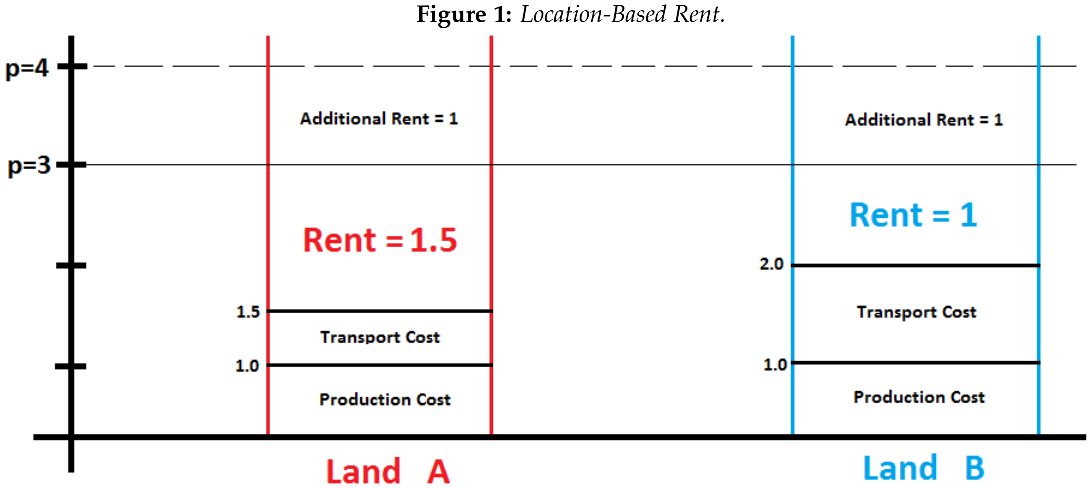
The worst of the impact was experienced in Iowa and the Great Plains states, due to the large debt incurred by the farmers in that area. Farmland prices rose the most in these states, which allowed the farmers to incur large debts using their land as collateral. When the land values crashed, so did the value of the farmers' collateral on their large debts, which led to many bankruptcies. The key question is why was the boom and bust of land prices so much more extreme in these geographically removed states?

I(b): Theory of Ricardian Rent

The answer can be found in the work of Ricardo. While in his essay on rent, Ricardo used the example of farmland of differing qualities, exactly the same applies to land of differing location: land closer to market has lower transport costs, just as the land of better quality has lower production costs. Thus, the land closer to market earns more rent and has higher value. He begins by noting that the more productive land will have lower costs and is therefore brought into production first. As demand increases, less favorable land with higher costs must be brought into production, which requires prices to increase. This generates rent and, hence, value on the more productive land. As this continues, even less favorable land is needed, generating more rent on all previous, superior qualities.

What is important for this paper is that the rent goes up much faster if it is initially small, ie, if the land is in production but of disadvantageous location. Consider two land parcels equivalent in all qualities - including agricultural productivity - but located at different distances from market where the products of the land will be sold. These are depicted in Figure (1). Land B, the land further from market, incurs greater transportation cost, and therefore earns lower rent. Land A, better located, has lower transport cost and hence higher rent. At $p = 3$, and with the costs as indicated, the rent to A is 1.5, the rent to B is 1. Now suppose the price increases to $p = 4$. Both rents rise by the same absolute amount, 1. However, the percentage increase is larger for B, 100%, than for A, 75%. Hence, the value of land B increases more than land A. It follows that with rising output prices, the value of land further from market rises more; with falling prices, it falls more. Hence, distant lands suffer the most from agricultural booms and busts.

Figure 1: *Location-Based Rent.*



I(c): Benirschka & Binkley (1994)

In their study, Benirschka & Binkley evaluated this concept of Ricardian Rent with regards to farmland in the Corn Belt, which experienced unusually large changes in land value from 1969-1982 and from 1982-1987, a period of boom and bust of farmland prices. The states of interest are Ohio, Indiana, Illinois, Missouri, and Iowa as they are arguably homogeneous in terms of agricultural production and are geographically dispersed across the Corn Belt.

The authors collected data for every county in the Corn Belt on farmland values, population dynamics, basic county loan and purchase rate for corn (positively correlated with distance to market, as explained in detail below), and agricultural factors. With these data, they estimated two models ((i) a straightfoward OLS model, and (ii) a spatial autoregressive error model using MLE[1]) to examine how distance to market affects land value changes. Both models were estimated twice - once for the time period of value growth, and once for the period of value decline. In their data, the most locationally advantageous state, with respect to terminal markets, was Ohio (which had the highest average loan rate), and the state that was the least advantageous was Iowa (which had the lowest average loan rate), which is on the opposite end of the Corn Belt.

The basic county loan rate is set by the Price Support Division in the USDA's Farm Service Agency. The loan rate itself refers to a set standard of grain quality and is announced annually for each county in the US, and is adjusted for yearly changes in the national average loan rate and differences in farm location. Pre-1987 (the last year in the original paper's data), the rates were determined using rail freight rates and prices received by producers, but since then, they are based on differentials between local grain elevator prices and the prices at terminal grain markets. In either case, it is clearly a function of transportation costs and distance to market. The corn loan rate is used since the primary crop produced in the region of interest and will thus reflect the price differential between local grain elevators and terminal corn markets.

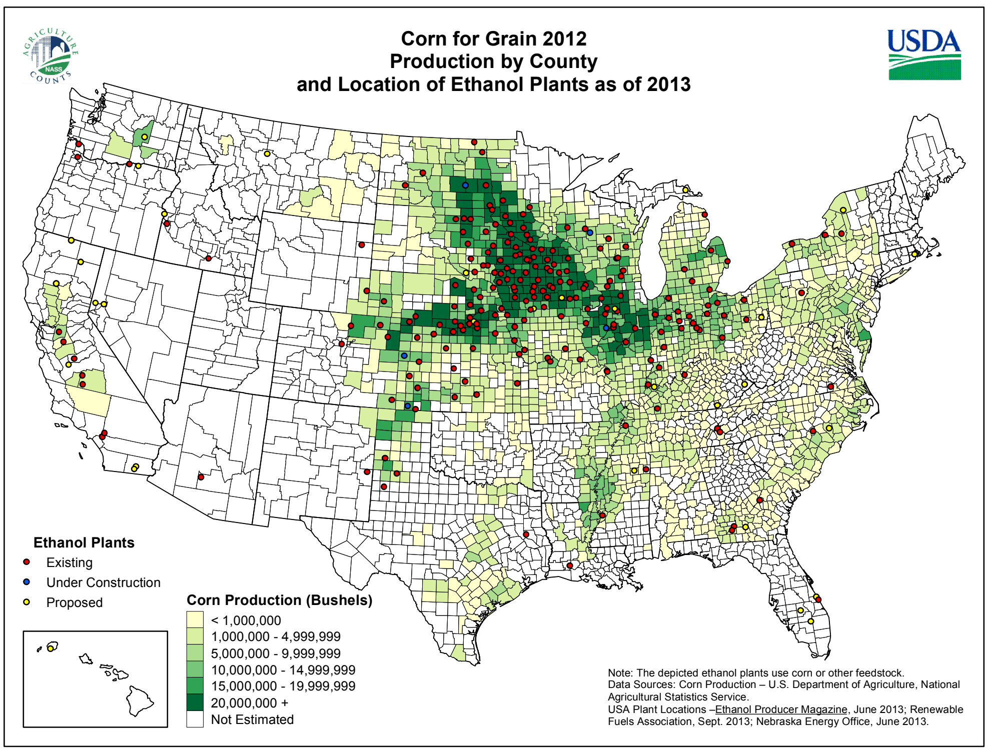
It was found that the changes in farmland value were the largest in absolute value in Iowa, for both periods. Furthermore, the county loan rate for corn (their measure for market proximity) was lowest in Iowa for both periods. As an empirical illustration, the authors took average values of all variables for Iowa in the period of value growth, and subsequently calculated the predicted average increase in land value. Then they replaced the average loan rate for the state of Iowa (the lowest) with the average loan rate for Ohio (the highest), and predicted the average value growth again, but this time artificially improving Iowa's location. As expected, the increase in farmland value was smaller for Iowa in the case of an artificially better location. Similarly, they repeated the process for the period of value decline. Again, the average fall in land values in Iowa were less severe when improving the location with Ohio's loan rate.

I(d): Goals of this study

[1]Their study was, perhaps, the very first published Agricultural Economics article to apply the technique of spatial econometrics.

This study will first estimate the effect of distance to market on land price volatility using the methodology of the original study, but with the most recent data available. Since 1994, the agricultural land market has experienced an ethanol and biofuels boom. Many ethanol plants have been constructed across the Corn Belt. According to a 2012 report by the USDA seen in Figure (2), the majority of ethanol plants were constructed in the Corn Belt, particularly in Iowa where loan rates (and, hence, corn prices) are lower than in, for example, Indiana and Ohio. This is shown in Figure (2):

Figure 2: *Ethanol Production Status, 2012.*



One hypothesis from this is that distance to market matters less now than it did in the 1969-1982 period. The role of space in land markets is less pronounced because the market has moved closer to the production. In addition, transport deregulation and changes such as unit trains are likely also to have diminished the role of space.

More elaborate and sophisticated spatial econometric techniques have been developed since the time of that paper. These newer models allow for a richer estimation of the effect of distance to market on the variation of farmland value. In the preliminary phases of this paper, some of these new cross-sectional techniques will be applied and the results will be compared to those of the previous econometric models. An eventual goal is first to test for structural change (once the historical data is acquired), and subsequently extend the methodology to spatial panel models.

II. Data

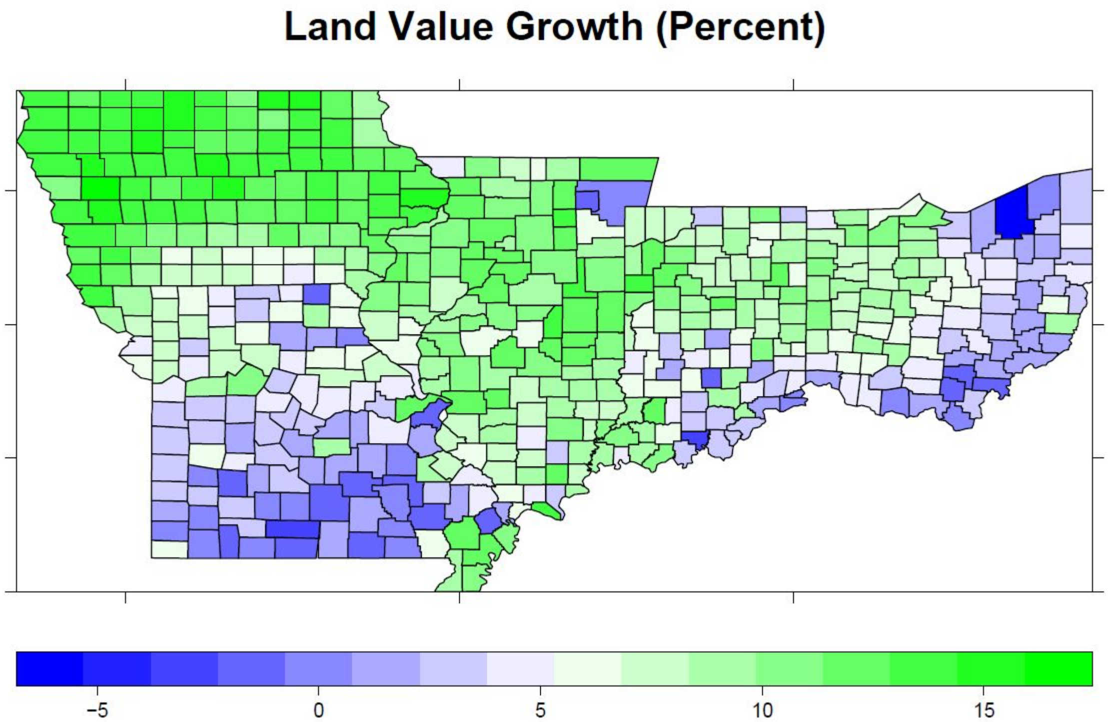
II(a): Current collection of data

The data range from 2007-2012, the two most recent Ag Census years, at the county level, across the five Corn Belt states: Ohio, Indiana, Illinois, Missouri, and Iowa. The total number of observations is 495 counties[2], excluding the city of St. Louis which was marked as a county in some datasets but not in others.

The dependent variable is the percentage change in land value in the county. This was computed by taking the natural logarithm of the land price in 2012 and subtacting the natural logarithm of the land price in 2007, divided by the number of years involved. The price levels were obtained from the USDA-ERS National Agricultural Statistics Service (NASS) in the Ag Census data. Given that this variable is likely to be spatially correlated, Moran's I was calculated, yielding an estimate of 0.70 with a zero p-value. This would indicate that space does, indeed, play a role in determining the behavior of land value changes. Figure (3) shows that, moving from east to west through the Corn Belt, land prices increase by much more.

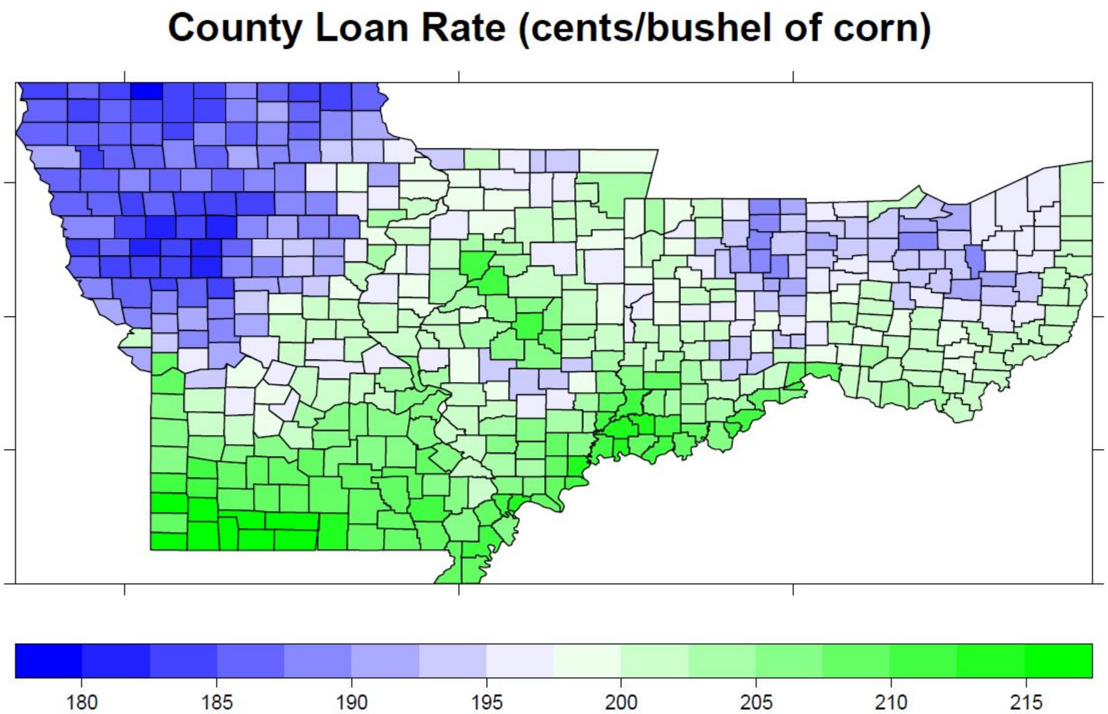
[2]The spatial polygon shapefiles are available for free on the internet from multiple sources at the aggregated national level. A loop was constructed in R to separate out the counties and states specified for this study using the universal FIPS codes indicating both county and state, merging the states and their respective counties, creating a new spatial polygon data frame of the 5 Corn Belt states.

Figure 3: *2007-2012 Land Value Growth Rates.*



The county loan rate for each state was obtained from the USDA Farm Service Agency. This acts as a price floor for corn in dollars per bushel. As previously stated, as the distance to market increases, the loan rate declines as the transportation costs increase with the distance, making it a reasonable proxy for market distance. Furthermore, as seen in Figure (4), it tends to decrease moving from east to west in the Corn Belt for the 2012 rates.

Figure 4: *2012 Basic County Loan and Purchase Rates.*



Another variable which is likely to have influence on the value of farmland is the average corn yield for the county. The Ag Census from NASS also provided this data for 2012. In addition to this, data for the average farm size per county in 2012 was also obtained from NASS, which was computed by dividing the total number of farmland acres in a county by the number of farms in that county. In the original study by Benirschka & Binkley (1994), it is argued that as the farm size increases and soil quality decreases, farmland prices are likely to vary more, under usual conditions.

To further account for geographic farmland heterogeneities across counties, the original article contained four additional variables based on the soil quality evaluations conducted in the Soil Survey by the USDA's Natural Resource Conservation Service's. The first four classifications are used in the model. Class I soil is the least production-impeded soil, and Class IV soil is the most impeded. Shares of classes V-VIII are not used, as they are not well-suited for crop production (e.g. pasture land, woodland). These data were obtained from the USDA's Natural Resource Conservation Service Soil Survey and provided to the authors by David Helmers at the University of Wisconsin. The data were then converted to county shapefile form with assistance from Alexander Young.

Other variables that are likely to influence the change in value of farmland are any population dynamics within a county. Including these types of variables incorporates the effects of being in a metropolitan region on value of land. These data were obtained from the US Census Bureau, and were used to compute the two population variables - percentage change in population from 2007 to 2012 and the population density[3]. The population density affects the supply of farmland as more people require more space, and the population growth affects the demand for farmland for purposes other than agricultural production.

Table (1) below shows basic summary statistics of the current collection of data for the 2007-2012 period.

Table 1: *Summary Statistics*

Variable	Mean	Std. Dev.	Min	Max
Land Value Growth$\star$ (*Percent*)	7.33	4.32	-5.33	15.96
Loan Rate (*cents/bu*)	198.29	7.78	180.00	215.00
Average Farm Size (*Acres*)	300.98	141.48	9.66	1342.00
Population Density (*persons/*mi^2)	139.47	304.19	7.65	3196.39
Population Change (*Percent*)	0.21	0.74	-2.05	4.59
Average Corn Yield (*bu/acre*)	99.40	37.21	19.00	180.10
Share of Class I soil (*Percent*)	4.99	7.44	0.00	41.05
Share of Class II soil (*Percent*)	43.54	24.14	0.99	98.19
Share of Class III soil (*Percent*)	25.11	15.01	1.29	75.03
Share of Class IV soil (*Percent*)	10.72	9.42	0.00	48.37

II(b): Continuing collection of data

Historical data for the variables are available from the NASS Quickstats database, the Census Bureau, the Farm Service Agency Price Support Division, and from Cornell University's archive

[3] The first is computed by taking the difference between the natural logarithm of the population in 2012 and the natural logarithm of the population in 2007 and dividing by 5 for each county; the second is computed by dividing the 2012 population by the total number of square miles in the county.

of USDA Ag Census publications. While most of the data have been obtained (those which have yet to be obtained are awaiting response to special requests for historical data), the numbers are in PDF format and must be either copied into a database software console such as SAS, or in most cases, typed in manually. This process is currently underway.

III. Methodology

This paper will employ the specific-to-general strategy of model specification, beginning with OLS, and performing various specification tests to refine the model choice. For models other than OLS, MLE will be used as it is asymptotically efficient and was the estimator used by Benirschka & Binkley (1994). Moreover, the software packages used assume normally distributed errors, which was confirmed with a Jarque-Bera test for normality. Moving forward in the study, other estimators such as GMM and spatial 2-state least squares will also be explored.

In the original study, the two models (OLS and MLE-spatial error) were the only models estimated, and we will choose these two as our starting point. Formally, we have the model

$$y = \alpha + X\beta + \epsilon. \tag{1}$$

where y is the average annual growth rate of land prices, and X contains (by county) the basic county corn loan rate (cents per bushel), the average number of farms (acres per farm), the percentage share of ClassI-ClassIV soils (percent), the population density (persons per square mile), the percentage change in population (percent), and the average corn yield (bushels per acre).

In the case of the spatial error model, we need to include a more detailed error specification:

$$y = \alpha + X\beta + \epsilon, \ where\ \epsilon = \lambda W\epsilon + \mu.^{4} \tag{2}$$

W is the spatial weighting matrix. It accounts for this spatial autocorrelation between neighboring counties by spatially lagging the series that exhibit spatially correlated patterns. The definition of "neighbors" is counties sharing immediately contiguous borders, consistent with that of Benirschka & Binkley (1994) where the (i, j) element of the spatial weights matrix W is

[4] The variance-covariance matrix in question is no longer concerning ϵ because the errors $\epsilon = \lambda W\epsilon + \mu = (\mathcal{I} - \lambda W)^{-1}\mu$ are not iid; thus, we have $Var[\epsilon] = \sigma^2[(\mathcal{I} - \lambda W)^T(\mathcal{I} - \lambda W)]^{-1}$ which is neither diagonal nor homoskedastic, ie, no heteroskedasticity and no serial correlation are no longer valid assumptions, both of which are necessary to use OLS. Hence, MLE is used.

positive for county i if it is a neighbor of county j, and 0 otherwise[5]. W is row-standardized so that $W\epsilon$ becomes the weighted average of the unobserved effects from neighboring counties. Thus, the interpretation of λ is the correlation between the error term in a county and the errors of the neighboring counties.

As the field of spatial econometrics has developed substantially since the time of the original study, more techniques are at our disposal to more rigorously analyze this problem. Consider, for instance the spatial autoregressive lag model,

$$y = \alpha + \rho W y + X\beta + \epsilon \; = (\mathcal{I} - \rho W)^{-1}(\alpha + X\beta + \epsilon) \tag{3}$$

where we control for spatial processes in the dependent variable, or the spatial ARAR (doubly autoregressive) model,

$$y = \alpha + \rho W y + X\beta + \epsilon, \; \epsilon = \lambda W \epsilon + \mu. \tag{4}$$

Appendix B shows statistical tests for model specification, indicating that the use of spatial models is warranted. If we use these more advanced spatial models, then marginal effects must be computed beyond the coefficients. For example, the ARAR model has the form $y = \alpha + \rho W y + X\beta + \epsilon, \; \epsilon = \lambda W \epsilon + \mu$ which is mathematically equivalent to
$y = (I - \rho W)^{-1}(\alpha + X\beta + \epsilon), \; \epsilon = \lambda W \epsilon + \mu$. This, in fact, creates a matrix of marginal effects - for the k^{th} regressor, $(I - \rho W)^{-1}\beta_k$. The k^{th} diagonal element is exactly the k^{th} marginal effect on a given county's y from a unit change in that county's x_k. Thus, the trace (the sum of the elements on the diagonal) is the sum of the direct effects, ie, the cumulation of the effects of a change in a given county's X. The off-diagonals, therefore, are the indirect effects from changes in the neighboring counties' X. For a total effect on a given county's x_k, the sum of the direct and indirect effects provides this estimate.

Finally, loan rates have uniformly risen since the period of the original study, simply due to inflation. We regressed a sample[6] of loan rates from Benirscha's dissertation on our 2012 loan rates and found that the rates have approximately doubled. Thus, to permit comparability of the

[5]Commonly known as "queen" contiguity, referring to the game of chess.

[6]Benirschka includes in his PhD dissertation a table of average loan rates by crop reporting district, each district being a collection of counties in a state. We averaged our county loan rates over crop reporting district to match his unit of analysis and carry out this computation.

studies, current loan rates were multiplied by 0.5.

IV. Results

Table (2) shows the results[7] from Benirshcka & Binkley (1994) for the period of land value growth. The negative and significant coefficient on the loan rate indicates that land price volatility does, indeed, decrease as distance to market shrinks.

Table 2: *Estimation Results*
Benirschka & Binkley (1994)
Dependent variable: Percent change in land price

Period of value growth 1969-1982 **(t-statistics listed below in parentheses)**		
	OLS	Spatial Error⋆
Intercept	25.5040 * ** (10.42)	21.2385 * ** (5.11)
Loan Rate (*cents/bu*)	−0.0747 * ** (−5.43)	−0.0648 * ** (−2.60)
Average Farm Size (*Acres*)	−0.0014 (−1.57)	0.0023 * * (1.99)
Population Density (*persons/mi*^{2})	−0.0015 * ** (−11.70)	−0.0009 * ** (−7.18)
Population Change (*Percent*)	−0.0329 (−0.65)	0.0716 (1.43)
Average Corn Yield (*bu/acre*)	0.0080 (1.39)	0.0079 (0.96)
Share of Class I soil (*Percent*)	−0.0388 * ** (−4.67)	−0.0328 * ** (−3.52)
Share of Class II soil (*Percent*)	−0.0286 * ** (−7.56)	−0.0212 * ** (−4.48)
Share of Class III soil (*Percent*)	−0.0332 * ** (−8.85)	−0.0193 * ** (−4.08)
Share of Class IV soil (*Percent*)	−0.0003 (−0.03)	0.0077 (0.86)
λ	_	0.60***
R^2 †	0.43	0.24
AIC	_	_
BIC	_	_
Log Likelihood	_	_

We find, in Table (3), results similar to those in the original paper. Regardless of model specification, we find that, as a county's distance to market decreases, the average value of land in that county is significantly less variable. Moreover, following the model specifications of the

[7] The symbols *,**, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.
†R^2 is calculated as the squared correlation between y and $\hat{y}$.
⋆Estimated using Maximum Likelihood Estimation - the $W\epsilon$ term violates the OLS assumption of homoskedasticity and no serial correlation.

original study, we see the coefficient on county loan rate to be larger in absolute value than before. This is contrary to the hypothesis stated earlier in this paper that the role of distance has diminished over time. The results suggest that, if anything, the role of space has actually increased.

Table 3: *Estimation Results*
Young, Binkley, & Florax (2015)
Dependent variable: Percent change in land price

Period of value growth 2007-2012 **(t-statistics listed below in parentheses)**		
	OLS	Spatial Error⋆
Intercept	21.9498 ∗ ∗∗ (5.68)	13.7584 ∗ ∗ (2.48)
Loan Rate (*cents/bu*)	−0.2439 ∗ ∗∗ (−6.86)	−0.1657 ∗ ∗∗ (−3.08)
Average Farm Size (*Acres*)	0.0106 ∗ ∗∗ (11.39)	0.0094 ∗ ∗∗ (9.21)
Population Density (*persons/mi*^{2})	−0.0015 ∗ ∗∗ (−3.66)	−0.0013 ∗ ∗∗ (−3.41)
Population Change (*Percent*)	0.1939 (1.26)	0.3405 ∗ ∗∗ (2.59)
Average Corn Yield (*bu/acre*)	0.0185 ∗ ∗∗ (5.11)	0.0191 ∗ ∗∗ (4.39)
Share of Class I soil (*Percent*)	0.1395 ∗ ∗∗ (7.03)	0.0912 ∗ ∗∗ (4.00)
Share of Class II soil (*Percent*)	0.0599 ∗ ∗∗ (6.89)	0.0689 ∗ ∗∗ (6.89)
Share of Class III soil (*Percent*)	0.0506 ∗ ∗∗ (5.51)	0.0668 ∗ ∗∗ (5.23)
Share of Class IV soil (*Percent*)	−0.0241 (−1.29)	0.0003 (0.02)
λ	_	0.63***
R^2 †	0.67	0.78
AIC	5.04	4.69
BIC	5.13	4.77
Log Likelihood	-1143.95	-1876.03

Table (4) shows results from the more advanced models, spatial lag and spatial ARAR. Both of these are estimated using MLE, and marginal effects were computed along with boostrapped standard errors in R.

These findings indicate that the effect of distance to market is stronger in recent periods than in the earlier period. However, until the historical data are obtained, statistical testing of the significance of this increase is not possible. Testing whether or not the effect actually increased is one of the subsequent steps of this study. Another investigatory aim of this paper is to further address this increase, after testing for its significance, and attempt to relate it to market changes,

Table 4: *Estimation Results, Calculated Marginal Effects of Modern Spatial Econometric Models (standard errors bootstrapped with 20,000 iterations)*
Young , Binkley, & Florax (2015)
Dependent variable: Percent change in land price

Period of value growth 2007-2012 **(t-statistics listed below in parentheses)**			
	Spatial Lag	Spatial ARAR	Manski
Loan Rate $(cents/bu)$	$-0.1614***$ (-2.66)	-0.0243 (-0.62)	$-0.3653***$ (-2.58)
λ	_	0.93***	-0.80***
ρ	0.49***	0.69***	0.86***
R^2 †	0.77	0.84	0.85
AIC	1.51	1.16	1.11
BIC	1.60	1.24	1.19
Log Likelihood	-1079.51	-1069.81	-1047.14

specifically the ethanol boom of the early 2000s.

References

1. Benirschka, M., and James K. Binkley. 1994. "Land price volatility in a geographically dispersed market." *American Journal of Agricultural Economics* 76(2): 185-195.

2. Boots, B. 1999. "A variance-stabilizing coding scheme for spatial link matrices." *Environment and Planning A* 31: 165-180.

3. Iowa Department of Transportation. 2012. "Iowa's Rail Freight System Trends", www.iowadot.gov/iowainmotion/files/pp6.pdf

Appendix A — Spatial Weights Matrix Standardization & Spatial Models

For a given spatial weights matrix

$$W = \begin{bmatrix} w_{11} & w_{12} & ... & w_{1n} \\ w_{21} & w_{22} & ... & w_{2n} \\ . & . & & \\ . & & . & \\ . & & & . \\ w_{n1} & w_{n2} & ... & w_{nn} \end{bmatrix}$$

The standardization method referred to as "W-coding" in Boots (1999) row-standardizes the weights matrix, W. Formally defined, we have the vector of local linkage degrees (by the row-sums of W),

$$\mathbf{d} = [\sum_{j=1}^{n} w_{1j}, \sum_{j=1}^{n} w_{2j}, ..., \sum_{j=1}^{n} w_{nj}]^T$$

Then our row-standardized "queen contiguity" matrix is $W^* = [diag(\mathbf{d})]^{-\mathbf{1}}W$.

The models estimated in this study are listed below:

- **OLS**: $y = \alpha + \beta X + \epsilon$
- **Spatial Error**: $y = \alpha + \beta X + \epsilon,\ \epsilon = \lambda W\epsilon + \mu$
- **Spatial Lag**: $y = \alpha + \rho Wy + \beta X + \epsilon$
- **Spatial ARAR**: $y = \alpha + \rho Wy + \beta X + \epsilon,\ \epsilon = \lambda W\epsilon + \mu$
- **Manski**: $y = \alpha + \rho Wy + \beta X + WX\gamma + \epsilon,\ \epsilon = \lambda W\epsilon + \mu$

Appendix B — Lagrange Multiplier Test Results, Wald Test Results, & Likelihood Ratio Test Results

Table 5: *Lagrange Multiplier Tests for Functional Form*

Model	LM statistic	p-value ¶
Spatial Error	161.47	0.00
Spatial Lag	166.04	0.00
Spatial Error (Robust)	18.87	$1.4(10^{-5})$
Spatial Lag (Robust)	23.44	$1.3(10^{-6})$
Spatial ARAR	194.81	0.00

Table 6: *Wald Tests for Functional Form*

Model	$\chi^2(k)$	p-value ¶
OLS	114.10	0.00
Cross Regressive	64.55	0.00
Spatial Error	396.83	0.00
Spatial Lag	237.74	0.00
Spatial ARAR	239.16	0.00
Manski	276.12	0.00

Table 7: *Likelihood Ratio Tests for Functional Form*

Model	$\chi^2(k)$	p-value ¶
OLS	563.69	0.00
Cross Regressive	611.67	0.00
Spatial Error	251.40	0.00
Spatial Lag	251.11	0.00
Spatial ARAR	210.88	0.00
Manski	181.93	0.00

¶Displayed digits in R ended at the 16^{th} decimal and, thus, p-values are displayed as 0.00.