Spatial Heterogeneity in U.S. Agricultural Productivity

Christopher Shultz

The Hartford Financial Services Group, c.shultz@live.com

# Abstract

@TODO: Rewrite this after the entire paper is finished.

This study uses U.S. county-level agricultural data for 2017 to estimate a spatially explicit production function and analyze the impact of spatial heteroegeneity on the corresponding input-output relationships. This an important consideration given that traditional analyses are usually temporal in nature. Results indicate that significant spatial heterogeneity is present which in turn impacts productivity by reducing the return to inputs in certain areas. Clustering is apparent, with high productivity areas concentrated in the Midwest and Southeast, and low productivity areas in the Southwest and Northeast. This is likely due to factors including land use, geographic characteristics, and water availability. Spillover effects from one region’s input use on the net income of a neighboring region are not significant. However, the presence of a spatial lag is clear regarding net income. The meaning of this finding is not as clear, as it’s possible that counties are like those around them, thus engaging in similar activity under similar conditions.

*Keywords:* agricultural productivity, spatial heterogeneity, spillover effects.

*JEL: C21, Q10, Q18*

# Introduction

The United States is one of the largest agricultural producers in the world and has more arable land than any other nation on earth (World Bank, 2022). The global agricultural system is complex, and food is more affordable in the U.S. than any other developed country. The business of agriculture is characterized by high startup costs and a commodity output with low value per unit. Even though the per-unit value of most crops is low, the 2017 Census of Agriculture showed that the market value of U.S. agricultural products sold was approximately $388.5 billion.

Farming operations are subject to input price fluctuations, including machinery, fertilizer, chemicals, seed, fuel, energy, and livestock. In addition, inflation has had a quantifiable impact. The EPA notes that using 1910-1914 as a base, prices received by today’s farmers have increased at least 6x, while costs of production have increased at least 16x (EPA, 2015a).

Historically low per-unit value of agricultural commodities and changing price conditions have led to a push for increased efficiency in agricultural production. This resulted in a shift toward large-scale operations producing in high-volume. The table below reveals that small farms, while high in numbers, comprise a small percentage of overall farm land and that farms in the highest economic sales class are much smaller in number, but use the highest percentage of farmland (EPA, 2015b).

The productivity of U.S. agriculture is growing faster than domestic food and fiber demand. Thus, U.S. farmers rely heavily on export markets to sustain income. According to data from the USDA (USDA, 2022), in the months since October 1976, U.S. agricultural imports have exceeded exports only 36 times (~6% of months in the past 46 years). Torgerson and Shane (2014) show that the value added to GDP by agriculture and related industries is approximately 5% of GDP (~$776 billion).

Despite evident shifts in the structure of U.S. agriculture, the sector is still dominated by small family farms and relies almost entirely on inputs such as rainfall, sunlight, and temperature (Alston et al., 2010). As a result, variation in these inputs can negatively impact agricultural producers. This provides the motivation for our study, which adapts the framework set forth in a recent paper by Yu et al. (2014), in which the authors attempt to evaluate the impact of spatial heterogeneity on the variation of input-output relationships in Turkey’s agricultural sector. The idea is that spatial heterogeneity may lead to variation in agricultural productivity.

In the U.S., different regions are characterized by different conditions important for agricultural production, including rainfall, soil, sunlight, temperature, etc. This variation can have a significant impact on agricultural production and productivity. For example, Lkupitiya et al. (2012) find substantial spatial and temporal variation in the in organic matter stocks in U.S. croplands.

Agricultural activity in the U.S. is also affected by policy. The primary agricultural policy tool of the U.S. government is the Farm Bill, which is renewed and revised every five years. The most recent update is the 2018 Farm Bill, which largely continues the farm and nutrition policy of prior years, with some details and caveats. Farm Bills are highly impactful to the agricultural sector and may involve direct payment programs, crop insurance, risk management, and Supplemental Nutrition programs, such as SNAP (Stabenow, 2014). Agricultural policies also exist at the state-level and may address local or regional concerns (Hamilton, 2013).

It is important for agricultural policies to be created and implemented in a way that most effectively allocates resources. Therefore, the purpose of this study is to provide additional empirical evidence on the degree to which U.S. agricultural productivity is affected by spatial heterogeneity. This provides a resource for agricultural policymakers and resources by painting a more complete picture of U.S. agriculture, especially when combined with other analyses.

Section 2 provides a detailed overview of the existing literature, beginning with a discussion of agricultural production economics and moving toward an exposition of agricultural productivity, production function estimation, and spatial heterogeneity. Sections 3 and 4 describe the methodological procedures employed in our spatial analysis. Section 5 is an overview of the dataset, and sections 6 and 7 present results, conclusions, and recommendations for future research.

# Background

The last century resulted in significant progress in agricultural economics and farm management strategies. Debertin (2012) describes agriculture as the closest real-world example of a traditional purely competitive market. There are clearly some caveats in the case of the U.S. market, as subsidies, risk programs, and other publicly funded or mandated actions impact the agricultural sector at all levels. Despite this consideration, the purely competitive model appears to represent farming better than any other existing model of economic behavior.

In a 2010 study, Chavas et al. provide a “reflection on the path taken by production economics and farm management over the past century.” In doing so, they identify 16 major contributions in production analysis, agricultural productivity, risk management, and dynamics. Some relevant findings include the identification of the role of diminishing returns in agricultural production and an assessment of economies of scale in agriculture (Spillman, 1923; 1924; 1933; Spillman and Lang, 1924).

Lyson and Welsh (1993) compare the current agricultural system to that of the mass production model of manufacturing in which efficiency and profit maximization is the ultimate aim. The level of U.S. farm output and input use from 1948-2011 grew at an average rate of 1.49 and 0.07 %, respectively. The significant difference between these numbers implies serious growth in productivity over time. Despite this, state-level analyses show considerable variation in growth rates across space (Ball et al., 2014). This variation in space provides additional motivation for the examination of spatial heterogeneity in agricultural production.

Many studies exist regarding the estimation of production functions, and that discussion could be a paper in and of itself. Rather, this review focuses on production function estimation with a spatial component. Shaik (2014) shows that traditional measures of return to scale and technology are under- and over-estimated in states at the upper and lower quantiles of the distribution, respectively. Weiss (1996) provides an overview of the emerging role of spatial analysis in agricultural economics and defines the core opportunity of economists as the need to quantify the costs and benefits of detecting and exploring spatial variation. Yu et al. (2014) refer to multiple other studies examining the role of space in production functions (Anselin et al., 1997; Cavailhes and Wavresky, 2003; Fingleton and Mccombie, 2006; Lambert and Cho, 2008; Vaya et al., 2004). Despite a fairly wide application of the more general methods of “spatial analysis”, the use of spatial econometrics in agricultural production function modeling to date is limited.

Cho et al. (2007) estimate an agricultural production function using Chinese county-level data and geographically weighted regression (GWR). They also compute county-specific input-output elasticities and create a visual representation using GIS. Their results are obtained by comparing GWR and OLS estimates to confirm that allowing for spatial variation in the regression paradigm significantly improves model performance.

Yu et al. (2014) provide an evaluation of the spatial variation that exists in the Turkish agricultural sector and how it impacts input-output elasticities across the country. Their results indicate that disparities in agricultural activities and geographic conditions affected the return from input factors, and that policy makers should consider this regional heterogeneity and potential comparative advantage when creating new legislation. The authors break this process down into four steps: [1] identifying spatial dependence; [2] generalizing a spatial production function; [3] comparing and selecting appropriate models; and [4] estimation and results. These four steps provide the generalized framework for this study. Additional papers of interest include Bille et al. (2015), Pedersen et al. (2017), Koc et al. (2017), Xu et al. (2020), and Ma et al. (2021).

# Identifying Spatial Dependence

Before considering the use of a spatial econometric model, it is important to explore the concept of spatial dependence. Generally, spatial dependence is “a situation in which values observed at one location depend on the values observed at nearby locations” (LeSage and Pace, 2009). Whether or not a location is nearby can be determined through the use of a various mathematical criteria, e.g. Euclidean distance, connected borders, etc. The authors provide a generalized example for which observations represent neighbors as defined by some criteria. In this case, the data generating process may take the form:

where

Where , the value of is a function of , and vice versa. This system represents a generalized example of spatial dependence, but does not necessarily define all cases or situations.

There are three forms of spatial autocorrelation: [1] positive spatial autocorrelation is the situation in which nearby observations are likely to be similar to one another; [2] negative spatial autocorrelation is the situation in which nearby observations are likely to be opposite one another; and [3] the case of no spatial autocorrelation is when no spatial effect is discernible. In most any dataset with a spatial component, everything is related to everything else. However, things that are closer together are “more” related than those that are farther away (Tobler, 1970). Yu at al. (2014) cite a number of other examples that have been discussed in the spatial econometrics literature (Anselin, 1995; Can, 1990; Can, 1992; Cliff and Ord, 1973; Dubin, 1992; Kilkenny and Thisse, 1999; LeSage, 1997; Leung et al., 2000; McMillen, 1992; McMillen, 2003).

In testing for the presence of spatial autocorrelation, the Moran’s I statistic is a generally accepted starting point. Moran’s I is a global statistic that provides an indication of the presence (or lack) of a pattern of spatial dependence that is true for the entire dataset (Anselin, 1996). The Moran’s I statistic is defined as follows, where N is the number of spatial units, X is the variable of interest,   is the mean of X, and is the spatial weights matrix (Moran, 1948).

Values of the Moran’s I statistic can range from -1 to +1 in which a value of -1 corresponds to perfect negative autocorrelation, 0 corresponds to no autocorrelation, and +1 represents perfect positive autocorrelation.

While it is important to consider the values of Moran’s I, given its global nature, it is not without flaws. When datasets become large, the degree of spatial autocorrelation between observations becomes more likely to show instability in the form of nonstationarity at the local level, spatial regimes, or spatial drift (Anselin, 1996). As a potential solution, Anselin (1995) proposed a general class of “local indicators of spatial association (LISA)” which allow for the decomposition of global indicators such as Moran’s I into the contribution of each observation. This technique lends itself readily to visualization and can provide insight into the spatial distribution of spatially associated effects. We apply both the Moran’s I and LISA procedures to our data to ensure that spatial autocorrelation is detectable.

# Empirical Model

The framework employed in this paper draws heavily from Yu et al. (2014), but differs in its method of model selection, estimation, and interpretation. We begin with a Cobb-Douglas functional form, a tool widely used in agricultural studies since 1928 (Debertin, 2012). The generalized form is represented below, in which represents net agricultural income for county , is total factor productivity (TFP), is input for state , and is the return of each input to output.

If production is assumed to be stochastic, a random shock can be incorporated, also referred to as a Solow residual.

The Solow residual is the regional variation in technological efficiency (LeSage and Pace, 2009). This is the productivity of input factors by county not included in the aggregated total factor productivity (TFP) (). This provides a clear window to the degree of spatial heterogeneity that may exist in U.S. agricultural production. In order to linearize the relationship, the natural log of both sides can be taken and converted to matrix form.

We consider multiple spatial econometric models for updating the production function. The most common spatial models are the Spatial Autoregressive Model (SAR), Spatial Durbin Model (SDM), and the Spatial Error Model (SEM) (LeSage and Pace, 2009). The SAR model’s only deviation from our formulation above is its inclusion of a spatial lag on the dependent variable. The SAR model is stated below, in which is the spatial lag coefficient value and is the weighted average of the dependent variable values of the surrounding counties:

It is important to consider the choice of a spatial weights matrix. There has been much debate in the econometrics literature regarding the specification of weights matrices, with concern that various specifications may produce varying or biased estimates. LeSage and Pace (2014) find no theoretical basis for this belief and conclude that weight-matrix specification is not likely to significantly impact results.

The Spatial Durbin Model (SDM) expands the framework further by including spatial lags of explanatory variables as well as the dependent variable.

Finally, the Spatial Error Model (SEM) uses lags to reflect dependence in the error process.

where .

Various specifications of these three models are considered through a rigorous model selection process. The use of OLS in spatial models can lead to inconsistent estimates, incorrect standard errors, and more. Maximum Likelihood, however, is consistent for spatial models and provides a reliable estimation method (LeSage and Pace, 2009). It should be noted that the interpretation of estimation results in spatial models is not as straightforward as that of a linear process. E.g., in the simple linear regression model, the interpretation of is that for a one-unit increase in X, y should increase by . This does not hold true for models such as the SAR or SDM.

This discrepancy comes as a result of the functional form, whereas the partial derivative is not simply equal to as a result of the spatially lagged dependent variable. A change in an explanatory variable in one region will affect the region itself (direct effect) and potentially all other regions (indirect effects). These effects are combined to yield the “total effect” of a marginal change. To determine the value of the derivative for an equation with a spatially lagged dependent variable, the equation must be algebraically manipulated (i.e. the model must be written in reduced form) (LeSage and Pace, 2009).

In the estimation of such models, the ideal data transformation is that of a log-log specification because of the scaling effect that occurs and also because coefficients of this type of model can be interpreted as the input-output elasticity for the given variable. Given that this data includes various negative net income values which yield complex numbers when logged, the data is transformed via the use of studentization. This process consists of subtracting the mean from each observation and dividing by the standard deviation. Coefficients (in the case of models with spatially lagged dependent variables, effects estimates) should then be interpreted as changes in standard deviation.

Model specification is assessed using the procedure defined by Elhorst (2010) in an article that provides a broad perspective on some key issues regarding the 2009 LeSage and Pace text. Elhorst begins with the testing procedure defined by Florax et al. (2003), which allows for testing the significance of the spatially lagged dependent variable and spatially lagged error term. Using this as a starting point, he proposes a generalized testing procedure that can help the econometrician determine the best model specification given the data.

The main problem with the procedure provided by Florax is that it provides only a limited model space in which we have . However, this does not allow for the consideration of the alternative specifications . The Spatial Lag of X (SLX) model is simply a standard normal linear model with an additional term consisting of spatially lagged independent variables. In this model, OLS can be used and coefficient interpretation is straightforward.

The Spatial Durbin Error Model (SDEM) augments the SEM model with a spatial lag of the explanatory variables.

where . By implementing a likelihood-ratio test and Wald test, Elhorst is able to develop a procedure that expands the model space to , providing expanded possibilities for global, local, and OLS models. A step-by-step overview of the Elhorst testing procedure can be found in Appendix 2. For a more thorough review, see the Elhorst (2010) article.

# Data

Data for this analysis is derived from the USDA NASS Quick Stats 2.0 web portal and the USDA Census of Agriculture for 3,079 counties in 2017. Net income ($) serves as a proxy for output. Input variables include labor expense ($), fertilizer expense ($), fuel expense ($), acres of land harvested, machinery asset value ($), and inventory of tractors and trucks. Many studies are fairly limited in their choice of input variables for the agricultural production function. The inclusion of this wide array of variables allows a vast overview of relationships and spatial dependence. Description of raw data values is provided in the table below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Avg** | **SD** | **Min** | **Max** |
| Net Income ($M) | 28.9 | 55.8 | -46.4 | 1,162.9 |
| Fertilizer ($M) | 7.7 | 12.8 | 0.001 | 269.8 |
| Fuel ($M) | 4.4 | 7.3 | 0.001 | 210.3 |
| Labor ($M) | 10.4 | 32.5 | 0.003 | 648.3 |
| Land (Acres, thousands) | 129.5 | 151.7 | 0.007 | 1,309.7 |
| Machinery ($M) | 88.7 | 83.7 | 0.024 | 1,120.9 |
| Tractors (Inventory, hundreds) | 13.2 | 10.6 | 0.04 | 145.2 |
| Trucks (Inventory, hundreds) | 10.9 | 8.2 | 0.04 | 103.4 |

The dataset includes missing values in a small number of counties/variables (~2.5% of counties have at least one missing value). When this occurs, for variable , state , we impute the average value of variable for all other counties in state .

# Results

|  |  |
| --- | --- |
| **Variable** | **Moran’s I** |
| Net Income | 0.474 |
| Labor | 0.534 |
| Land | 0.726 |
| Machinery | 0.579 |
| Fertilizer | 0.517 |
| Fuel | 0.455 |
| Tractors | 0.513 |
| Trucks | 0.474 |

Considering net income and various input values for U.S. agriculture, a Moran’s I statistic is displayed above. These results were computing in GeoDa and a rook contiguity matrix in which the value of if states share a common border and otherwise. The degree of spatial autocorrelation is positive and relatively high for all variables, ranging from approximately 0.455 to 0.726.

This provides sufficient justification for further exploration of the impact of space on agricultural production. In examination of local spatial autocorrelation, the LISA method of analysis is employed via GeoDa, mapping results across 3,079 mainland counties for each variable. This creates a visual overview of the presence of local spatial association in the dataset. Within each section, shading represents significance of at least 95%. Darker portions represent higher significance. It’s clear that significant clustering exists, with the location of that clustering varying depending on input variable. Given that this represents positive spatial autocorrelation, counties that are highly significant tend to be surrounded by other similarly valued counties.

A picture containing background pattern

Description automatically generated

Specification testing results are displayed below. The Florax et al. (2003) procedure indicates a rejection of the null hypothesis that in the generalized test, but fails to reject in the robust variety. Also, the procedure rejects the null that in both the generalized and robust varieties. Using the steps laid out by Elhorst in 2020, this points us toward the SDM model. Moving forward with the likelihood ratio test to determine whether the SDM can be reduced to the SAR model results in a rejection of SAR. The same is true for the restricted SEM. This implies that spatially lagged dependent variables and spatially lagged error terms are likely to play a role in the DGP. Consequently, OLS can be rejected in favor of the SDM. In support of this conclusion, LeSage and Pace often argue that the SDM is the best place to start when considering model specification.

|  |  |  |
| --- | --- | --- |
|  | **Marginal Probability** | **Result** |
| LM Lag | 0.0000 | Reject the null that rho = 0. |
| LM Error | 0.0000 | Reject the null that lambda = 0. |
| LM Lag Robust | 0.7473 | Fail to reject the null that rho = 0. |
| LM Error Robust | 0.0000 | Reject the null that lambda = 0. |
| LR (SDM vs. SAR) | 0.0000 | Reject the restricted (SAR) model. |
| LR (SDM vs. SEM) | 0.0000 | Reject the restricted (SEM) model. |
| Wald (SLX vs. OLS) | n/a | Testing points to SDM model. |

Table 9 presents the results of the SDM model for U.S. mainland counties in 2012. Most notably, is significant at the 99% level and suggests a high degree of positive spatial autocorrelation. This means that counties with high agricultural net income are likely surrounded by counties with high agricultural net income. The opposite is also true in that low agricultural net income counties are located near other low agricultural net income counties.

It is important to recall that when interpreting output from a regression model with a spatially lagged dependent variable, attention must be paid to effects estimates rather than coefficients. Effects estimates of an independent variable in county can be interpreted such that a one standard deviation increase (decrease) in an explanatory variable increases (decreases) the dependent variable by standard deviations. Results indicate that fertilizer, fuel, labor, and machinery all have a positive impact with direct effects of 0.1073, 0.5452, 0.0997, and 0.5855, respectively.

The magnitude of these effects is worth noting. As previously discussed, the growth of agricultural productivity and the use of efficient technology has significantly reduced the need for labor. Our results confirm that labor plays a small role when compared with other inputs. E.g., a one standard deviation increase in labor results in only a 0.0997 standard deviation increase in net income. Conversely, a one standard deviation increase in machinery results in a 0.5855 standard deviation increase in net income. This provides an empirical demonstration of the relative importance of machinery to labor in today’s agricultural production system. Land, tractors, and trucks have a negative effect on net income, which may be explained by the hypothesis that these are assets held by the owner that may or may not directly add to the net income, and may involve debt service that takes away from net income.

In consideration of spillover effects, the indirect effect is the change in standard deviations of net income in county that can be attributed to a one standard deviation change in the independent variable for the weighted average of surrounding counties. The presence of spillovers in this analysis is minimal and only shows significance in the land variable. The data represents land as the number of acres of agricultural land harvested in a given county. It is important to consider that spatially significant variables may be explained by the fact that counties near one another have similar characteristics, and thus engage in similar agricultural activity with similar resources.

These results align with those found by Yu et al. (2014) in that indirect effects are not shown to be statistically significant in agricultural production. However, their direction is still worth noting. E.g. bidding up the price of labor in county has a negative effect on the net income of neighboring counties.

Note that the regression included 3068 observations and the R-squared is 0.7929. Note that the map of Solow residuals is a percentile map of their absolute deviation from accuracy. I.e. the raw residual is y – yhat, but we take the absolute value of that to show where areas have the largest distance from the true value.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Coefficient** | **Direct** | **Indirect** | **Total** |
| Constant | 0.557\*\*\* |  |  |  |
| Fertilizer | -0.170 | -0.232\*\* | -0.930\*\* | -1.162\*\* |
| Fuel | 3.314\*\*\* | 3.578\*\*\* | 4.176\*\*\* | 7.754\*\*\* |
| Labor | 0.430\*\* | 0.401\*\*\* | -0.484\*\*\* | -0.082 |
| Land | -0.438\*\*\* | -0.943 | -7.825\*\*\* | -8.768\*\*\* |
| Machinery | 0.331 | 0.341\*\*\* | 0.152\* | 0.493\*\*\* |
| Tractors | -0.0002 | -0.0001 | -0.010\*\* | -0.012\*\* |
| Trucks | -0.012\*\*\* | -0.012\*\*\* | -0.002 | -0.014\*\*\* |
| W-Fertilizer | -0.227\*\*\* |  | | |
| W-Fuel | -0.615\*\*\* |
| W-Labor | -0.458 |
| W-Land | -2.654\*\*\* |
| W-Machinery | -0.158 |
| W-Tractors | -0.004 |
| W-Trucks | 0.007\*\*\* |
| Rho | 0.649\*\*\* |

Solow residuals are mapped for each mainland county in the U.S. in the figure below. Recall that Solow residudals () represent regional variation in the productivity of input factors by county not included in the TFP (). Clusters of high productivity occur in the Midwest and Southeast. Particularly low productivity regions are scattered throughout the Southwest and Northeast. This reveals significant impacts of spatial heterogeneity on the input-output relationships in U.S. agricultural production.

Our findings are intuitive, as states in the Midwest and Southeast have historically exhibited favorable conditions for agricultural production and production is usually focused on crops that perform well in the conditions of the local environment. One important factor is water availability. While irrigation systems exist to mitigate risk, they do not entirely offset water concerns in dry areas such as the Southwest. The plains of the Midwest lend themselves readily to production of fields of crops and the utilization of large machinery which significantly increases productivity. In addition, land in the Northeast is highly forested and this extends to agricultural parcels. Comparing this to a highly productive region such as the Midwest reveals that highly productive regions are typically less forested and more conducive to mass-scale row crops and animal agriculture.

Results reveal some significance in total effects and these values are worth noting. The total effect is the combination of direct + indirect effects and represents the effect of a marginal change in input use on agricultural net income across the country. Total effects for fertilizer, fuel, machinery, and trucks are significant at the 99% level and amount to 0.159754, 0.520459, and -0.228783, respectively. A marginal increase in the use of fertilizer, fuel, or machinery in county leads to an increase in net income not only in county , but also in surrounding counties. It is worth noting that direct effects make up roughly 94% of total effects for these three inputs. Thus, indirect effects are not only statistically insignificant, but small. The significance of the total effects can be derived from the dominance of direct effects, which are highly significant and substantially larger in size.

Map

Description automatically generated

# Conclusion

This study used U.S. county-level agricultural data for 2017 to estimate a spatially explicit production function for U.S. agriculture. This allows for an expanded understanding of the traditionally non-spatial and largely temporal framework of estimation. Results indicate that significant spatial heterogeneity is present and impacts productivity substantially by reducing the degree of return to inputs in certain areas. Significant clustering is apparent, with high productivity areas concentrated in the Midwest and Southeast, and low productivity areas in the Southwest and Northeast. This is likely a result of a combination of various factors including land use, geography, water availability, and local policy actions.

The use of this spatial econometric estimation technique provides value by allowing for the separation of the effects of generalized productivity increases and those derived from spatial heterogeneity, which has been clearly demonstrated through the use of a Solow residual. Spillover effects from one region’s input use on the net income of a neighboring region are not significant, confirming the results of a similar study by Yu et al. (2014). The presence of a spatial lag is significant with regard to net income. The meaning of this finding is not well-defined, as it’s possible that counties are similar to those around them with regard to economics, geographic characteristics, and climate, resulting in similar activity under similar conditions.

Policymakers should consider the costs and benefits of subsidizing agricultural activity in low-productivity areas, as a more efficient scenario likely exists. This may result in more fragmented policies that encourage agricultural activity in productive areas and discourage it in less productive areas. Alternatively, policymakers could focus on ways to improve productivity in low-productivity zones.

# Appendix: Elhorst Testing Procedure

The Elhorst testing procedure is defined as follows. We consider a model space , which contains three global spatial models, two local spatial models, and one simple OLS specification. The procedure is outlined as follows:

1. Estimate model via OLS and use the various LM tests to determine the proper model (use the Florax et al. process).
2. If OLS is rejected in favor of SAR, SEM, or both, estimate the SDM.
3. Use the LR test to examine the following hypotheses:
   1. H0: tests whether SDM can be reduced to SAR.
   2. HO: tests whether SDM can be reduced to SEM.
4. If H0(a) can’t be rejected, use SAR as long as robust LM test statistics point to SAR.
5. If H0(b) can’t be rejected, use SEM as long as the LM test points to SEM.
6. If one of the conditions in 4 and 5 are not satisfied, use the SDM.
7. If OLS is estimated and not rejected in favor of both SAR and SEM, the OLS should be re-estimated with spatially lagged explanatory variables and the hypothesis H0: tested.
8. If H0: can’t be rejected, use OLS.
9. If H0: is rejected, the SDM model should be estimated and H0: should be tested.
10. If H0: is rejected, use SDM. Else, use SLX.

# References

Alston, J., James, J., Andersen, M. and Pardey, P. 2010. A Brief History of U.S. Agriculture. In *Persistence Pays*. Springer, New York, NY.

Anselin, L. 1995. Local Indicators of Spatial Association-LISA. *Geographical Analysis.* 27(2): 93-115.

Anselin, L. 1996. The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association. *Spatial Analytical Perspectives on GIS.* 111-125. Routledge, London, UK.

Anselin, L. 1998. *Spatial Econometrics: Methods*

Anselin, L. 1998. *Spatial Econometrics: Methods and Models.* Vol. 4. Springer Science & Business Media.

Anselin, L., Varga, A. and Acs, Z. 1997. Local Geographic Spillovers Between University Research and High Technology Innovations. *Journal of Urban Economics.* 42: 422-448.

Bauer, A. and Black, A. 1994. Quantification of the Effect of Soil Organic Matter Content on Soil Productivity. *Soil Science Society of America Journal.* 58: 185-193.

Ball, E., Wang, S. and Nehring, R. 2014. *Agricultural Productivity in the U.S.* Economic Research Service. USDA. Washington, DC.

Bille, A., Salvioni, C. and Benedetti, R. 2015. Spatial Heterogeneity in Production Functions Models. No. 212662, 150th Seminar, October 22-23, 2015. Edinburgh, Scotland. European Association of Agricultural Economists.

Can, A. 1990. The Measurement of Neighborhood Dynamics in Urban House Prices. *Economic Geography.* 66: 254-272.

Can, A. 1992. Specification and Estimation of Hedonic Housing Price Models. *Regional Science and Urban Economics.* 22: 453-474.

Cavailhes, J. and Wavresky, P. 2003. Urban Influences in Periurban Farmland Prices. *European Review of Agricultural Economics.* 30: 333-357.

Chavas, J., Chambers, R., and Pope, R. 2010. Production Economics and Farm Management: a Century of Contributions. *American Journal of Agricultural Economics.* 92(2): 356-375.

Cho, S., Chen, Z., Yen, S., and English, B. 2007. Spatial Variation of Output-Input Elasticities: Evidence from Chinese County-Level Agricultural Production Data. *Papers in Regional Science.* 86: 139-157.

Cliff, A. and Ord, J. 1973. *Spatial Autocorrelation.* Pion Limited. London, UK.

Debertin, D. 2012. *Agricultural Production Economics.* 2nd Edition. Macmillan Publishing Company. Upper Saddle River, NJ.

Dettori, B., Marrocu, E. and Paci, R. 2012. Total Factor Productivity, Intangible Assets, and Spatial Dependence in the European Regions. *Regional Studies.* 46: 1401-1416.

Dubin, R. 1992. Spatial Autocorrelation and Neighborhood Quality. *Regional Science and Urban Economics.* 221: 433-452.

Elhorst, J. 2010. Applied Spatial Econometrics: Raising the Bar. *Spatial Economic Analysis.* 5(1): 9-28.

EPA. 2015a. Land Use Overview. *US Environmental Protection Agency.* Washington, DC. <http://www.epa.gov/agricultural/ag101/landuse.html>.

EPA. 2015b. Ag 101. *US Environmental Protection Agency.* Washington, DC. <https://www.epa.gov/agriculture/ag-101-overview-agriculture>

Fingleton, B. and Mccombie, J. 2006. Empirical Growth Models with Spatial Effects. *Papers in Regional Science.* 85: 177-198.

Florax, R., Folmer, H. and Rey, S. 2003. Specification Searches in Spatial Econometrics: the Relevance of Hendry’s Methodology. *Regional Science and Urban Economics.* 33: 557-579.

Kilkenny, M. and Thisse, J. 1999. Economics of Location: A Selective Survey. *Computers & Operations Research.* 26: 1369-1394.

Koc., A., Yu., T., Lambert, D., Cho, S., Boluk, G. and Kim, S. 2017. A Spatial Analysis of the Relationship Between Agricultural Output and Input factors in Turkey. *New Medit.* 16(1): 11-17.

Lambert, D., and Cho, S. 2008. Forecasting Input Demand Shocks on China’s Gross Value of Agricultural Output. *Review of Regional Studies.* 38: 343-359.

LeSage, J. 1997. Regression Analysis of Spatial Data. *Journal of Regional Analysis and Policy.* 27: 83-94.

LeSage, J., and Pace, R. 2009. *Introduction to Spatial Econometrics.* CRC Press. Boca Raton, FL.

LeSage, J. and Pace, R. 2014. The Biggest Myth in Spatial Econometrics. *Econometrics.* 2: 217-249.

Leung, Y., Mei, C. and Zhang, W. 2000. Testing for Spatial Autocorrelation Among the Residuals of the Geographically Weighted Regression. *Environment and Planning A.* 32: 871-890.

Lokupitiya, E., Paustian, K., Easter, M., Williams, S., Andren, O., and Katterer, T. 2012. Carbon Balances in US Croplands During the Last Two Decades of the Twentieth Century.” *Biogeochemistry.* 107: 207-225.

Lyson, T. and Welsh, R. 1993. The Production Function, Crop Diversity, and the Debate Between Conventional and Sustainable Agriculture.” *Rural Sociology.* 58(3): 424-439.

Ma., L., Long, H., Tang, L., Tu, S., Zhang, Y. and Qu, Y. 2021. Analysis of the Spatial Variations of Determinants of Agricultural Production Efficiency in China. *Computers and Electronics in Agriculture.* 180: 1-13.

McMillen, D. 1992. Probit with Spatial Autocorrelation. *Journal of Regional Science.* 3: 335-348.

McMillen, D. 2003. Spatial Autocorrelation or Model Misspecification? *International Regional Science Review.* 26: 208-217.

Moran, P. 1948. The Interpretation of Statistical Maps. *Journal of the Royal Statistical Society B.* 10(2): 243-251.

Moran, P. 1950. Notes on Continuous Stochastic Phenomena. *Biometrika.* 37(1/2): 17-23.

Pederson, C. and Krogli, S. 2017. The Effect of Land Type Diversity and Spatial Heterogeneity on Farmland Birds in Norway. *Ecological Indicators.* 75: 155-163.

Shaik, S. 2014. Re-Examination of Production, Cost, and Restricted Profit Functions Using Quantile Regression Approach. *Agricultural Economics.* 45(2): 199-212.

Solow, R. 1957. Technical Change and the Aggregate Production Function. *Review of Economics and Statistics.* 39: 312-320.

Spillman, W. 1923. Applications of the Law of Diminishing Returns to Some Fertilizer and Feed Data. *Journal of Farm Economics.* 5: 36-52.

Spillman, W. 1924. Law of Diminishing Increment in the Fattening of Steers and Hogs. *Journal of Farm Economics.* 6: 166-178.

Spillman, W. 1933. Use of the Exponential Yield Curve in Fertilizer Experiments. United States Department of Agriculture. Technical Bulletin 348.

Spillman, W. and Lang, E. 1924. *The Law of Diminishing Returns.* World Book Company. Yonkers-on-Hudson, NY.

Stabenow, D. 2014. Agriculture Reform, Food and Jobs Act of 2013. United States Senate Committee on Agriculture, Nutrition, and Forestry.

Tobler, W. 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography.* 46: 234-240.

Torgerson, D. and Shane, M. 2014. What is Agriculture’s Share of the Overall U.S. Economy? USDA Economic Research Service.

USDA. 2019. 2017 Census of Agriculture. United States Department of Agriculture.

USDA. 2022. U.S. Agricultural Trade Update. United States Department of Agriculture. Economic Research Service.

Vaya, E., Lopez-Bazo, E., Moreno, R., and Surinach, J. 2004. Growth and Externalities Across Economies: An Empirical Analysis Using Spatial Econometrics. In *Advances in Spatial Econometrics* by Anselin et al. Springer.

Weiss, M. 1996. Precision Farming and Spatial Economic Analysis: Research Challenges and Opportunities. *American Journal of Agricultural Economics.* 78(5): 1275-1280.

World Bank. 2022. Agriculture, Forestry, and Fishing Value Added (Current USD). <http://data.worldbank.org>

Xu, W., Xiaboin, J., Jing, L. and Yinkang, Z. 2020. Impact of Cultivated Land Fragmentation on Spatial Heterogeneity of Agricultural Agglomeration in China. *Journal of Geographic Sciences.* 30(1): 1571-1589.

Yu, T., Cho, S., Koc, A., Boluk, G., Kim, S., and Lambert, D. 2014. Evaluating Spatial and Temporal Variation in Agricultural Output Elasticities in Turkey. *Agricultural Economics.* 45: 279-290.