The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

A Spatial Econometric Analysis of Cotton Yield Response to Nitrogen

Xiaofei Li
Department of Agricultural Economics
Mississippi State University
xiaofei.li@msstate.edu

Jac Varco
Department of Plant and Soil Sciences
Mississippi State University
JVarco@pss.msstate.edu

Amelia Fox
Department of Plant and Soil Sciences
Mississippi State University
aaf103@msstate.edu

Selected Paper prepared for presentation at the 2018 Southern Agricultural Economics Association Annual Meeting, Jacksonville, Florida February 2-6, 2018

Copyright 2018 by Xiaofei Li, Jac Varco, and Amelia Fox. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

Abstract

Cotton lint yield response to nitrogen levels has been studied extensively based on randomized complete block design experiments. In order to estimate the response curve, the most widely used statistical model is the ordinary least squares (OLS) regression model. Yield errors at specific plots conditioning on nitrogen treatments are canceled out by the model. However, statistically OLS estimates are the most efficient only when the yield errors are completely random. In the experiment practice, the yields errors are often spatially correlated across plots, mainly driven by the unobserved (and uncontrolled) soil characteristics in the field. In the presence of spatially non-random errors, spatial econometric models provide more accurate estimates than OLS. This study applies the Spatial Error model to the estimation of cotton yield response to nitrogen. Our data are from field experiments conducted during three crop years from 2012 through 2014 in three separate locations in Mississippi. Results show that the response coefficients estimated by Spatial Error model are significantly different from those of OLS model. Statistical theory and numerical simulation both prove the spatial model outperforms OLS. This study suggests spatial econometric model is more desirable in analyzing cotton field experiment data compared to OLS.

Keywords: Cotton nitrogen response, spatial econometric model, OLS, field experimental data

1. Introduction

Nitrogen management is of particular importance in cotton production. In the agronomy literature cotton yield response to nitrogen levels has been studied extensively. Those studies are typically based on randomized complete block design experiments, and nitrogen response curves are fitted through the ordinary least squares (OLS) regression models. However, an underlying assumption for OLS is that the regression errors are completely random. But in practice those errors are mainly driven by unobserved soil characteristics in the experiment fields (such as soil remaining nutrients, organic matters, water holding capacities, soil textures, etc.) which are highly likely to be spatially correlated. The presence of the spatially non-random errors may considerably affect the accuracy of cotton nitrogen response curve estimates by OLS. In the meanwhile, the spatial econometric models are widely recognized for their capacity to more properly address spatial correlation in data. However, to date they are still rarely used in the field experiment data analysis.

This study applies the Spatial Error model to the estimation of cotton yield response to nitrogen. Our data are from field experiments conducted during three crop years from 2012 through 2014 in three separate locations in Mississippi. The coefficients of cotton nitrogen responses estimated by the Spatial Error model are found to be quite different from those of OLS model. While the spatial model results are not necessarily always better, from a statistical point of view the spatial model estimates have a higher probability to be closer to the true coefficients. In other words, they are more accurate than the OLS estimates.

To verify the improvement in estimation performance by Spatial Error model over OLS, a 250-iteration simulation comparison is carried out. As shown in Figure 3, the OLS estimates are more likely to be deviating from the true coefficients, especially for small sample. On the other hand the spatial model estimates are relatively much more concentrated around the true coefficients. Given that cotton field experiments mostly have quite small sample size, the low accuracy level of OLS model is of a particular concern. Thus, this study calls for the use of spatial econometric models in cotton nitrogen response experiment data analysis, which significantly outperform the conventionally used OLS models.

2. Models and Data

This study applies the spatial econometric models to the estimation of cotton nitrogen responses using field experiment data. The experiments were conducted during three crop years from 2012 through 2014 in three separate locations in Mississippi. In 2012 the field was at south of Natchez, MS, with N rates of 30, 60, 90, and 120 pounds per acre. In 2013 two fields were established north of Schlater, MS, with either N rates of 30, 60, 90, and 120 pounds per acre or 70, 95, 120, and 145 pounds per acre. In 2014, one field at Money, MS, was established with N rates of 30, 60, 90, and 120 pounds per acre. A six row cotton picker equipped with a yield monitor was used to harvest the cotton and yield data was extracted using buffers around points. All soils are alluvial in nature.

The cotton yield to N relation is estimated by using the Spatial Error regression model, where the regression errors are assumed to follow a distance-based spatial autocorrelation process. The Spatial Error model is chosen due to the fact that the unobserved soil quality factors are more properly modeled as a spatial error process (Anselin et al., 2004). OLS model is also estimated for the purpose of comparison. We use the both linear and quadratic functional forms to capture the potential nonlinear effect of nitrogen on cotton growth.

The OLS model specifications are:

$$Y_i = \beta_0 + \beta_1 N_i + \varepsilon_i$$

or

$$Y_i = \beta_0 + \beta_1 N_i + \beta_2 N_i^2 + \varepsilon_i \, .$$

Y_i represents the cotton lint yield in plot *i*. N_i is the nitrogen treatment level in the experiment. For OLS models the error term ε is assumed to be randomly distributed over space.

For Spatial Error models, the error term ε is assumed to follow a spatial autoregressive process:

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

where W is the spatial weights matrix, and u is spatially random errors. In this study we adopt the inverse distance weights matrix. It captures the distance decaying pattern of the correlation between plots.

3. Results

The estimation results by OLS and Spatial Error models for the 3 separate experiments are shown in Table 1 through 3. The coefficients of cotton nitrogen responses estimated by the Spatial Error model are found to be quite different from those of OLS model. While the spatial model results are not necessarily always better, from a statistical point of view the spatial model estimates have a higher probability to be closer to the true coefficients. In other words, they are more accurate than the OLS estimates.

4. Simulation Illustration

It should be first clarified that the OLS estimates are not "wrong". In fact, they are still unbiased even when error terms are spatially correlated. That is, the statistical average of each estimated parameter still equals its true value. But the OLS estimates are no longer the most efficient. The distribution of the OLS estimate is highly dispersed around its mean, which means the accuracy level is low. In contrast, the Spatial Error model estimates have more concentrated distributions around the means. That suggests the Spatial Error model estimates are more accurate than the OLS estimates. This comparison can be demonstrated by a simple simulation work. Suppose there is one cotton experimental field with 16 by 16 square plots as shown in Figure 2 (a). Four nitrogen treatment levels are applied in a randomized complete block design setting: 0, 45, 90, and 135 kg N ha^{-1}. Each plot's cotton lint yield is simulated as a summation of a quadratic N function[1] and an error term ε:

$$yield = \alpha + \beta_1 N + \beta_2 N^2 + \varepsilon . \qquad \text{(Eq.1)}$$

The yield error ε comes from the unobserved soil quality variation in the field. Those plot-level errors are spatially correlated as shown in Figure 2 (b). Based on the simulated data for the 16 by 16 plots we can estimate the yield function (Eq.1) and obtain the estimated parameters β_1 and β_2 using OLS and Spatial Error models, respectively. Due to the existence of error ε, the estimated parameters will not be exactly the same as the true parameters values we imposed.

[1] Quadratic yield function is used to capture the agronomic fact that N deficiency reduces vegetative and reproductive growth and induces premature senescence, thereby potentially reducing yields (Gerik et al., 1994; Tewolde and Fernandez, 1997). On the other hand, high N availability may shift the balance between vegetative and reproductive growth toward excessive vegetative development, thus delaying crop maturity and reducing lint yield (Gaylor et al., 1983; Howard et al., 2001; Kohli and Morrill, 1976).

This simulation and estimation process is repeated for 250 times. Each time a different set of spatially correlated errors are generated. In total we obtain 250 different sets of estimated β_1 and β_2 by both OLS and Spatial Error models. Since the ultimate interest is the optimal N level (N*), we calculate the predicted optimal N level using the estimated model parameters as:

$$\hat{N}^* = -\frac{\hat{\beta}_1}{2\hat{\beta}_2},$$

where $\hat{\beta}$ denotes the estimate of the parameter. We can then compare the 250 predicted $\hat{N}^*$ against the true N^* which is:

$$N^* = -\frac{\beta_1}{2\beta_2},$$

to evaluate the performances of the two models. The results are presented in Figure 2. As can be noticed, the probability density distributions of both models' estimated $\hat{N}^*$ are centered around the true N^*. But the OLS distribution is much more dispersed than that of the Spatial Error model. The probability of Spatial Error model estimated $\hat{N}^*$ falling within ±2.5 kg ha^{-1} of true N^* is 99.1%, while the OLS is only 87.2%. That means the Spatial Error model has a higher probability to obtain an estimate that is closer to the true parameter. In other words, after obtaining an estimated $\hat{N}^*$ by Spatial Error model we are 99.1% sure that the true N^* is within ±2.5 kg ha^{-1} from it, but by OLS model we are only 87.2% sure.

The accuracy of OLS estimates is especially poor with small sample. As shown in Figure 3, as the size of the field experiment decreases to 12 by 12 and 8 by 8 plots, the distribution of OLS estimates becomes even more flat. The chance of estimated $\hat{N}^*$ falling within ±2.5 kg ha^{-1} of true N^* decreases to 77.6% and 42.9%. In contrast, the Spatial Error model estimates still remain fairly concentrated distributions, with the two numbers as 96.3% and 92.7%. In practice, randomized complete block design experiments usually do not have very large size due to the high costs. Thus the poor accuracy level of OLS model estimates is of particular problem, while the Spatial Error model provides more acceptable estimates. Note that the numbers in this simulation work is only for illustration purpose. The actual accuracy levels in real experimental data depend on the actual field conditions. But as long as there exist spatial correlation in unobserved soil quality the Spatial Error model always outperforms OLS model in terms of estimation accuracy. Therefore we highly promote the use of Spatial Error model in the cotton nitrogen response studies.

5. Conclusions

Through cotton experimental data, this study demonstrates that the cotton yield response to nitrogen parameters can be quite different from OLS and Spatial Error model estimations. Given the commonly existing spatial correlation in soil conditions and other unobservable growing factors, the assumptions for OLS models are unlikely to hold. Our results illustrate that OLS may lead to larger deviation from the true parameters through a simulation experiment. This study suggests spatial econometric model is more desirable in analyzing cotton field experiment data compared to OLS.

Table 1. Cotton yield regression on N treatment (Guedon 2012)

	Dependent variable:			
	YIELD_LBAC			
	OLS	Spatial Error	OLS	Spatial Error
N treatment	0.813***	0.748***	1.761	1.741
	(0.266)	(0.231)	(1.518)	(1.106)
N treatment^2			-0.632	-0.683
			(0.996)	(0.743)
Constant	786.474***	787.643***	758.049***	759.684***
	(21.864)	(21.424)	(49.912)	(49.912)
λ		0.337***		758.049***
		(0.095)		(49.912)
Observations	72	72	72	72

Note: Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01

Table 2. Cotton yield regression on N treatment (PDMoney, 2013)

	Dependent variable:			
	LNT_YLD			
	OLS	Spatial Error	OLS	Spatial Error
N treatment	4.770***	5.340***	11.717***	9.717***
	(0.816)	(0.414)	(4.284)	(2.197)
N treatment^2			-4.43	-2.815**
			(2.683)	(1.389)
Constant	1,070.092***	1016.415***	844.491***	876.950***
	(72.756)	(60.083)	(154.542)	(90.791)
λ		0.671***		0.674***
		(0.054)		(0.054)
Observations	120	120	120	120

Note: Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01

Table 3. Cotton yield regression on N treatment (GYMoney, 2014)

	Dependent variable:			
	LINTLBAC			
	OLS	Spatial Error	OLS	Spatial Error
N treatment	5.109***	4.961***	9.009**	10.608***
	(0.790)	(0.364)	(4.486)	(1.962)
N treatment^2			-2.321	-3.314***
			(2.627)	(1.389)
Constant	927.793***	933.562***	798.973***	741.517***
	(75.526)	(60.594)	(164.279)	(89.317)
λ		0.654***		0.684***
		(0.060)		(0.056)
Observations	87	87	87	87

Note: Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01

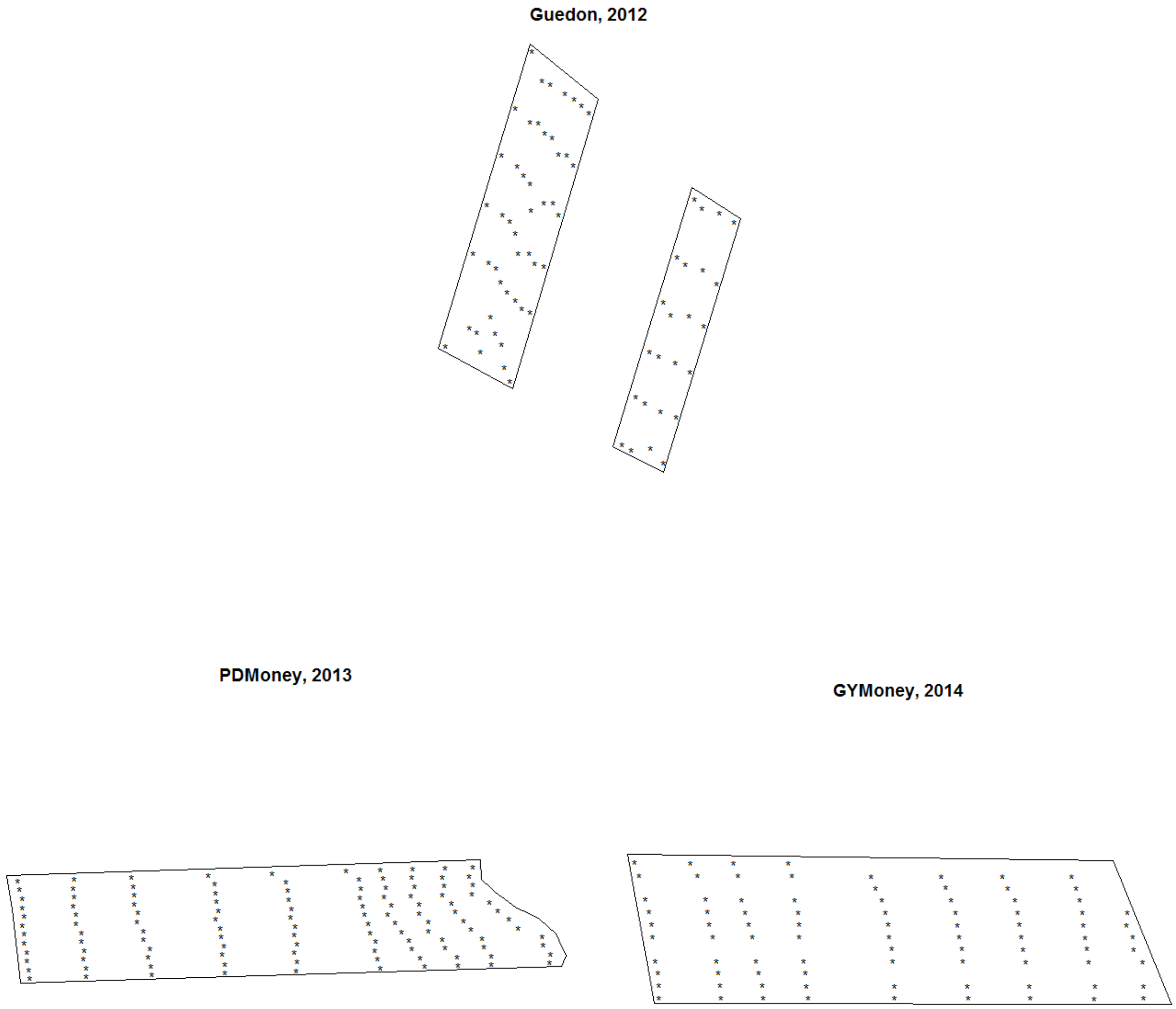

Figure 1. Cotton experiment fields layout in this study.

Randomized Complete Block Design of Nitrogen Treatments

135	90	0	45	45	0	90	135	90	0	135	45	90	45	0	135
0	90	135	45	0	135	90	45	0	90	45	135	90	45	0	135
45	135	0	90	135	0	45	90	135	0	45	90	90	135	45	0
135	90	0	45	45	135	90	0	90	45	0	135	135	45	90	0
135	0	90	45	0	135	45	90	90	135	45	0	0	90	45	135
135	90	45	0	135	45	0	90	45	0	90	135	45	90	135	0
135	45	0	90	90	135	45	0	45	0	90	135	0	90	45	135
45	135	90	0	45	0	90	135	135	90	0	45	45	135	0	90
135	0	45	90	90	0	45	135	90	0	45	135	45	135	0	90
135	45	90	0	90	135	45	0	45	90	135	0	0	90	135	45
0	90	45	135	135	90	45	0	45	0	90	135	135	0	45	90
135	45	90	0	90	135	45	0	135	90	45	0	135	90	0	45
90	45	135	0	0	45	135	90	0	90	135	45	0	45	90	135
0	45	135	90	45	90	135	0	90	0	135	45	45	0	90	135
135	45	0	90	90	0	45	135	135	45	0	90	0	135	90	45
45	135	0	90	0	135	90	45	0	135	45	90	135	45	90	0

Unobserved Soil Quality

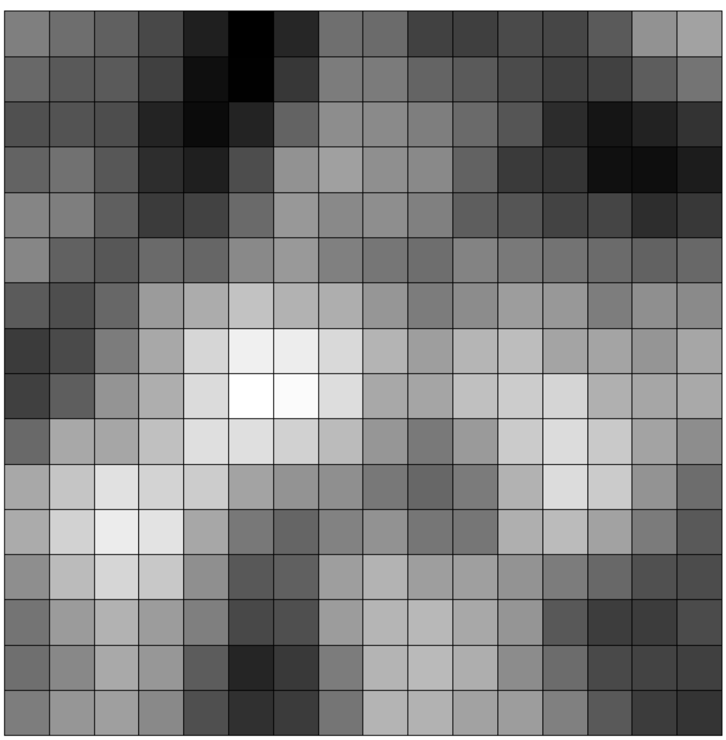

Figure 2. A simulated cotton experiment field with (a)(top) randomized complete block design N treatment kg ha^{-1} and (b)(bottom) spatially distributed unobserved soil quality

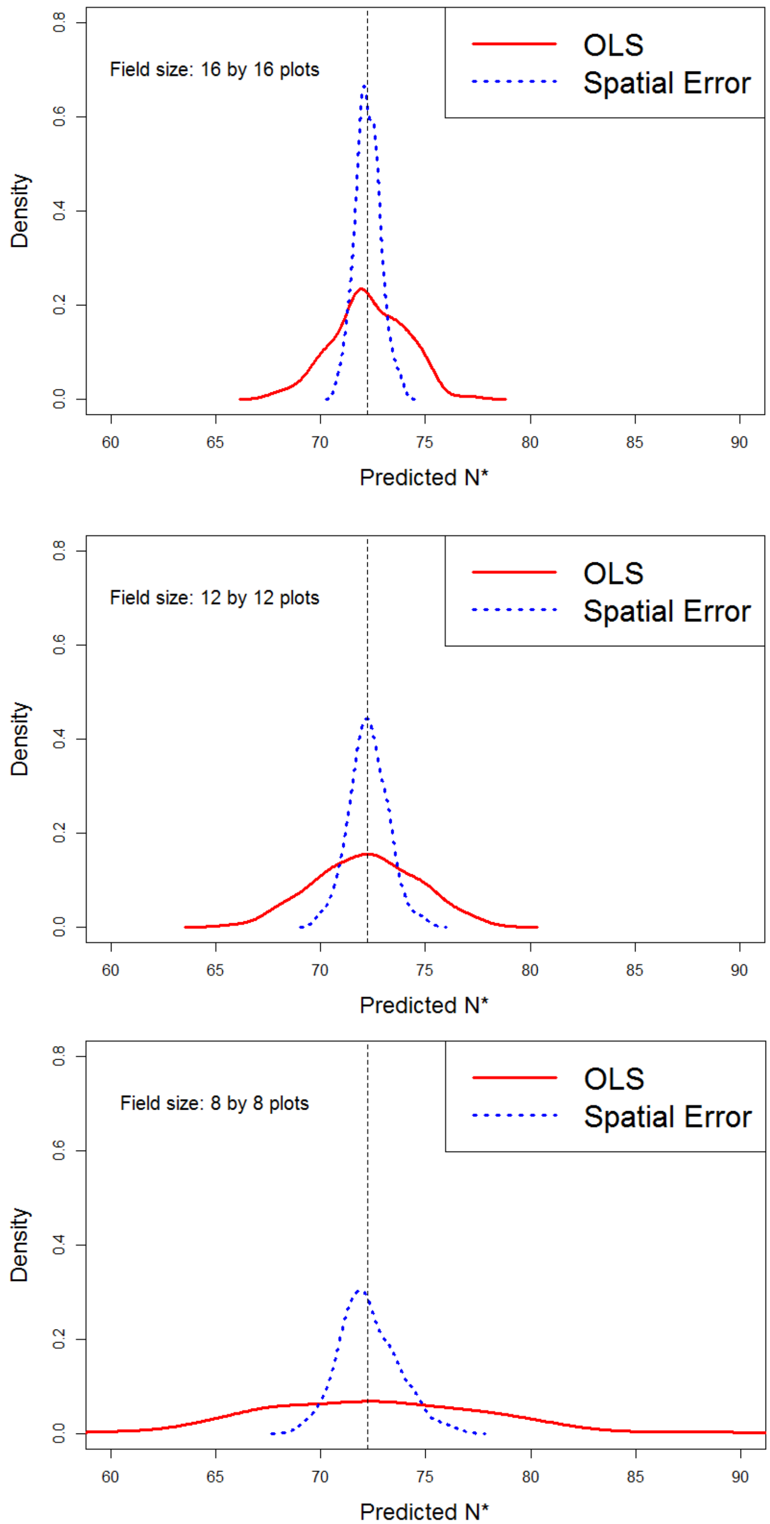

Figure 3. Kernel density distributions of predicted optimal nitrogen level (N*) by OLS and Spatial Error models, based on 250 times of simulation. Sample size of field: 256, 144, and 64 plots. True N*=72.2 (dotted vertical line).