BT COTTON AND SPATIAL AUTOCORRELATION OF YIELD

by

RAN HUO

(Under the Direction of Lynne Seymour)

ABSTRACT

The spatial autocorrelation of crop yield violates the independence assumption of conventional methods such as ordinary least square (OLS). Therefore, in the model of crop yield, spatial structure must be incorporated. More often, crop yields are available across space as well as over time, and the additional dimension allows the estimation of the full spatial covariance matrix, using the time dimension. However, the stationarity of the spatial pattern does not necessarily hold over time. In this study, we present empirical evidence that the spatial autocorrelation pattern of cotton yield can be fundamentally changed by the adoption of Bt cotton seeds. The finding of this study provides a cautionary note that spatial autocorrelation may vary over time due to technological change.

INDEX WORDS: Spatial autocorrelation, Crop yield, Bt cotton, Moran's *I*

THE IMPACT OF BT COTTON ON THE SPATIAL AUTOCORRELATION OF YIELD

by

RAN HUO

M.S., Chinese Academy of Agricultural Sciences, Beijing, China, 2014

A Thesis Submitted to the Graduate Faculty

of The University of Georgia in Partial Fulfillment

of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2019

THE ADOPTION OF BT COTTON AND SPATIAL AUTOCORRELATION

by

RAN HUO

Major Professor: Lynne Seymour
Committee: Jeffery Dorfman
Paul Schliekelman

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2019

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my major advisor Dr. Seymour for her invaluable supports and advice on the development of my dissertation. I highly appreciate all the time, patience, and efforts she devoted to my dissertation.

I also deeply appreciate my committee members Dr. Dorfmand and Dr. Schliekelman, who provided valuable advice and insights into this dissertation. I would like to give my thanks to Dr. Liu and Dr. Park. As coordinators, they gave me support and help when I was studying in the department. Also, I would like to take this chance to express my deepest appreciation to all the faculty members, stuffs, and fellow graduate students in the Department of Statistics at the University of Georgia.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Crop yield are often correlated among spatially neighboring observations. This spatial autocorrelation violates the assumption of independence in classical statistical analysis. Models that fail to consider spatial autocorrelation will result in incorrect inference. Therefore, spatial analysis approaches are needed to model crop yield. Spatial autocorrelation can be introduced into this model in the form of spatial lag process or spatial error autoregressive process. In either of the two methods, the spatial correlation pattern is assumed to be invariant within the study field. Most often, crop yield observations are available as over time. This additional dimension allows the estimation of a spatial autocorrelation structure, using the time dimension to provide the asymptotics (Anselin, 2013). However, as for crop yield, it is possible that the spatial autocorrelation pattern changes over time. Few studies have explicitly considered the variation of spatial autocorrelation of crop yield. The exceptions include Ping et al. (2004) and Maestrini and Basso (2018). The former provides evidence that the spatial autocorrelation pattern of cotton yield is significantly different between dry years and wet years. The latter discusses the predictors of crop yield by categorizing the crop field into two types - stable spatial pattern and variant spatial pattern. They conclude that for spatial patterns that are stable over time the best predictor of the spatial variability is the historical yield, while for the fields that are more sensitive to weather and thus fluctuate over years, the best predictor of the spatial patterns are exogenous variables such as weather, Normalized Difference Vegetation Index (NDVI), etc.

Different from the short-run impact of weather conditions discussed in Ping et al (2004), the advancement of agricultural technology is able to cause fundamental changes in a spatial autocorrelation pattern by reducing the impacts of an adverse natural environment. Therefore, the spatial dependence is expected to decrease with the diffusion of the technology since it can induce stable yield across natural environmental conditions. In this study, we hypothesize that the spatial autocorrelation of crop yield can be decreased with the widespread use of agricultural technology. Taking genetically modified cotton as our example, we examine the change in spatial autocorrelation of the cotton yield before and after the adoption of *Bacillus thuringiensis* (hereafter abbreviated Bt) cotton in the Yellow River valley region of China. This genetically modified cotton produces an insecticidal protein that protects the cotton plant from certain caterpillar insect pests. Since its first introduction into the market in 1996, Bt cotton was rapidly adopted in both developed and developing countries. The Yellow River valley is one of the cotton planting regions that promptly adopted Bt cotton after its official approval by the Chinese government, and provides us an appropriate sample area to explore our research question. To investigate our hypothesis, we use Moran's I as the statistic of spatial autocorrelation and explore its change with the adoption Bt cotton. In order to support the potential causality relationship between the change of spatial autocorrelation and the adoption of Bt cotton, we present a set of robustness checks. We further discuss the application of this finding in precision agriculture and crop insurance. In general, the changes in the spatial autocorrelation challenge the validity of the assumption of stable spatial pattern over time. Cautious examination *ex ante* and certain adjustment of the assumption *ex post* are suggested if fundamental technology advancement took place over the study period. This finding serves as a cautionary note for the use of an underlying assumption of stable second moments (covariance structure). We suggest that statistics such as Moran's *I* provide a tool to examine the stationarity of the spatial autocorrelation pattern of crop yield over time when modeling and predicting crop yields.

The finding of this study provides evidence of an unexplored impact of Bt cotton. While there are concerns that Bt cotton might have limited ability to benefit small farmers and may even intensify social inequality, abundant literature has provided empirical evidence that Bt cotton reduces the use of insecticides, increases productivity and farmers' cotton profit and living standards around the world. (Pray et al., 2001; Huang et al., 2002b; Chen et al. 2013, Gouse et al., 2005; Qaim et al, 2006; Kathage and Qaim, 2015). However, the beneficial effects of Bt cotton are more than the increased economic welfare, and new findings on the broader beneficial impacts are disclosed including improvements in farmer's health status (Huang et al., 2002b), water savings (James, 2002), and gender equality employment (Kouser et al., 2015). Bt cotton can also provide considerable environmental benefits by substantially reducing the number of pesticide sprayings. A number of studies have demonstrated that the adoption of Bt cotton can increase the population of beneficial insects (Head et al., 2001; Smith, 1997; Xia et al., 1999) and is compatible with integrated pest management initiatives (Benedict and Altman, 2001). Many crop scientists have examined the impact of Bt cotton on soil health and biodiversity, and pointed out that the cultivation of Bt cotton produces positive effects on diversity of soil, plants, microbial groups (Cattaneo et al., 2006; Donegan et al., 1995; Velmourougane, 2017; Sarkar, et al., 2009). The quantitative measure of the change in cotton spatial dependence caused by Bt cotton has not been discussed to date. Contributing to the larger literature exploring the impacts of Bt cotton, this study provides evidence that Bt cotton may potentially reduce cotton yield's spatial dependence and increase the resistance of cotton to systemic adverse geographic conditions. To our knowledge, this work is the first attempt to examine the effect of the adoption of Bt cotton on the spatial autocorrelation of cotton yield.

The rest of this thesis is organized as follows. First, we outline the concept of spatial autocorrelation and the statistics used to test the spatial autocorrelation, and review the literature on spatial models for crop yield. Next, we introduce the background of Bt cotton in the study area, the Yellow River valley region in China. Then, the findings and robustness checks are presented. Similar situations in the cotton planted in another province in China and three states of the U.S. are discussed as supplemental evidence of the generalization of the results. Next, we discuss the application of this finding in precision agriculture and crop insurance. The conclusion and discussion section wrap up these findings and give some directions for future work.

CHAPTER 2

THE METHOD AND LITERATURE REVIEW

In this section, we outline the specification of spatial autocorrelation in econometric models, review the diagnostic statistics used to test the presence of spatial autocorrelation, and summarize the literature that models crop yield with spatial statistic methods.

2.1 SPATIAL AUTOCORRELATION

Spatial autocorrelation occurs when the value of a spatially observed random variable is dependent from the observed values at nearby locations. The framework used for the statistical analysis of spatial autocorrelation is a so-called spatial stochastic process, or a random variable y referenced by location i

$$\{y_i, i \in D\} \tag{2.1}$$

where the location set D can be either a continuous surface or a finite set of discrete locations.

Spatial processes can be modeled in two basic approaches. One is geostatistics, or Kriging, which is used to model the process over a continuous location set (Cressie 1992). The fundamental assumption of the Kriging method is in terms of covariance stationarity. That is, the spatial autocorrelation is a function only of distance. The other method uses a discrete location set, and is used to model areal (lattice) data. The stationarity assumption is replaced by spatial autoregressive models that are based on specific choices of spatial weight matrix. The most fundamental spatial autoregressive models are the spatial error model and the spatial lag model. The spatial error model is given as

$$y = X\beta + \varepsilon, \varepsilon = \lambda W \varepsilon + \upsilon. \tag{2.2}$$

where y is an $n \times 1$ vector of spatial random variables,; X is an $n \times k$ matrix of explanatory variables; β is a $k \times 1$ vector of regression coefficients; ε is an $n \times 1$ vector of residuals; λ is a spatial autoregressive parameter, which measures the influence of neighboring values; W is an $n \times n$ spatial weights matrix; and υ is a homoscedastic uncorrelated error term. The spatial lag model is given as

$$y = \rho W y + X\beta + \upsilon, \tag{2.3}$$

where ρ is the spatial autoregressive parameter and the others are as above. Theory and *a priori* information suggest a spatial error regression in Equation 2.2 to model crop yield.

2.2 MORAN'S *I*: GLOBAL AND LOCAL

To detect the existence of spatial autocorrelation, spatial diagnostics such as Moran's I and Geary's C are widely used. Moran's I is an extension of Pearson correlation coefficient in space. However, different from Pearson's correlation coefficient which requires that the observations of the pairs of two variables be mutually independent, Moran's I considers the case in which observations of a variable are correlated with one another.

In 1948, Moran asked if the occurrence of an event (with probability p) can be regarded as statistically independent across counties, or, on the other hand, if the presence of some event in a county makes its presence in neighboring counties more or less likely. In this question, he simplified the spatial relationship as 1 for adjacent neighbors, and 0 otherwise, and proposed a test of significance for the spatially random distribution. In the 1970s, Cliff and Ord (1972, 1981) presented a comprehensive work on spatial autocorrelation, in which they explicated and generalized Moran's work and suggested the calculation of a spatial autocorrelation coefficient they named Global Moran's I. Spatial autocorrelation over the entire area of interest can be described in terms of Global Moran's I, with value that ranges from -1 to 1. Value 1 means complete clustering of similar values, while -1 suggests a checkerboard pattern, and 0 implies no spatial autocorrelation that is, there is no detectable spatial structure in the data.

Another widely used statistic to test spatial randomness is Geary's C (Geary 1954), which measures dissimilarity of a spatial variable within a dataset. The value range of this statistic is between 0 and 2, where 0 implies strong positive spatial autocorrelation, 1 implies no autocorrelation, and 2 implies strong negative spatial autocorrelation. It is inversely related to Moran's I, and the difference in effectiveness of Moran's I and Geary's C to measure the spatial autocorrelation is negligible (Myint, 2003).

In the 1990s, Anselin (1995) introduced local Moran's I and local Geary's C to measure the spatial associations of neighboring observations to a specific observation. He also showed that the local values were proportional to their global values. These statistics were called Local Indicators of Spatial Association (LISA), and used to identify possible centers of statistically significant clustering or "hot spots".

The Global Moran's I statistic (Cliff and Ord, 1981) is given by

$$I = \frac{N}{S_0} * \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij} z_i z_j}{\sum_{i=1}^{N} {z_i}^2} \tag{2.4}$$

where z_i is the deviation of a variable of site i from its spatial mean, i.e. $z_i = x_i - \bar{X}$, and $\bar{X} = \frac{\sum_{i=1}^{N} x_i}{N}$, w_{ij} is the spatial weight between site i and site j, N denotes the total number of sites, and S_0 is the aggregate of all the spatial weights:

$$S_0 = \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} \ . \tag{2.5}$$

Contrary to intuition, the expected value of Moran's *I* under the null hypothesis of no autocorrelation is not equal to zero, but is given by

$$E[I] = \frac{-1}{N-1} \tag{2.6}$$

and its variance is given by

$$V[I] = \frac{N(S_1(N^2-3N+3)-NS_2+3S_0^2)}{(N-1)(N-2)(N-3)S_0^2} - \frac{K(S_1(N^2-N)-2NS_2+6S_0^2)}{(N-1)(N-2)(N-3)S_0^2} - \left(\frac{1}{N-1}\right)^2 \tag{2.7}$$

where

$$S_1 = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (w_{ij} + w_{ji})^2, \tag{2.8}$$

$$S_2 = \sum_{i=1}^{N} \left(\sum_{i=1}^{N} w_{ij} + \sum_{j=1}^{N} w_{ji}\right)^2 \tag{2.9}$$

and

$$K = \frac{N \sum_{i=1}^{N} (z_j)^4}{\left(\sum_{i=1}^{N} (z_j)^2\right)^2}. \tag{2.10}$$

The sampling distribution of Moran's *I* statistic approaches a normal distribution asymptotically under the null hypothesis of no spatial autocorrelation as *N* increases to infinity. The z-score is calculated as:

$$Z = \frac{I - E[I]}{\sqrt{V[I]}} \ . \tag{2.11}$$

Apart from the measure of overall spatial autocorrelation within the region, often, the degree of local similarity of the values is of interest. The Local Moran's *I* is a LISA which measures the degree of the local similarity of observations around an individual site. The Local Moran's *I* (Anselin, 1995) is calculated as

$$I_i = \frac{z_i}{S_i^2} \sum_{j=1}^{N} w_{ij} z_j \tag{2.12}$$

where S_i is the aggregate of all the spatial weights:

$$S_i{}^2 = \sum_{j=1}^{N} \frac{z_j{}^2}{N-1}. \tag{2.13}$$

The mean and variance of Local Moran's I for a complete random spatial process are given by:

$$E[I_i] = \frac{-\sum_{j=1}^{N} w_{ij}}{N-1} \tag{2.14}$$

and

$$V[I_i] = \frac{(N-K)\sum_{j=1}^{N} w_{ij}^2}{N-1} - \frac{(2K-N)\sum_{k=1}^{N}\sum_{h=1}^{N} w_{ik}w_{ih}}{(N-1)(N-2)} - [\frac{\sum_{j=1}^{N} w_{ij}}{N-1}]^2 \tag{2.15}$$

where

$$K = \frac{N\sum_{i=1}^{N}(z_j)^4}{\left(\sum_{i=1}^{N}(z_j)^2\right)^2}. \tag{2.16}$$

The associated z-score of the test of significance is calculated as:

$$Z_i = \frac{I_i - E[I_i]}{\sqrt{V[I_i]}}. \tag{2.17}$$

Similar to the Global Moran's I, a significantly positive value of Local Moran's I indicates that the site has similar values clustering together. A significantly negative value indicates that the neighboring sites have dissimilar values.

2.3 THE SPATIAL WEIGHTS MATRIX

To model spatial autocorrelation explicitly in a spatial process model, the spatial weights matrix W must be defined first. Conceptually, the spatial weights matrix is a $N \times N$ matrix reflecting how much a site spatially interacts with all the rest of sites. Spatial weights are often row standardized, particularly with binary weighting strategies. Row standardization is used to create proportional weights, i.e. the weights are the value of each cell divided by the sum of the values of its row. In W, each off-diagonal element w_{ij} represents the magnitude of geographic interaction between a pair of values observed at sites i and j. Zero indicates a lack of spatial interaction between two observations. The diagonal element w_{ii} can be interpreted as "self-influence" and is defined as zero.

There are two fundamentally different ways to construct the weight matrix. One is to weight spatial interaction by inverse geographical distance. This inverse distance method assumes the spatial process is observed on a continuous (geostatistical) area. The other is to define spatial interaction with a binary variable to represent contiguity, meaning that two spatial sites share a common border of non-zero length.

The most common measure of distance is Euclidean distance:

$$d_{ij} = \sqrt{(i_{easting} - j_{easting})^2 + (i_{northing} - j_{northing})^2}, \qquad (2.15)$$

giving a weight of

$$w_{ij} = \frac{1}{d_{ij}}. \qquad (2.16)$$

There are other more complicated specifications of the weight matrix based on distance, such as inverse distance with higher order of power, $w_{ij} = (\frac{1}{d_{ij}})^{\alpha}$, and exponential distance, $w_{ij} = \exp(-\alpha d_{ij})$, where α is a positive constant for both weights.

In many settings, the spatial process is defined as a statistical summary over a geographical region, such as averages, ratios or counts over a state or county. This type of spatial process refers to areal data or regional summary data. Due to its convenience, the inverse distance weight matrix is often employed to model regional summary data, where the observations are assumed to be located at the center point of the area. However, modeling regional summary data with an inverse distance matrix gives rise to a conceptual inconsistency between "emptiness" among the data points versus the assumption of the continuous spatial indexing. Another commonly cited problem is the arbitrariness of assigning the summary value to the centroid (Wall, 2002).

One way to construct the weight matrix for an areal process is based on physical neighbors or contiguity relationship. Thus, instead of measuring geographic distance between data points, the spatial weight is defined by whether two areal units touch. There are two common ways to identify neighbors, both of which are based on a chessboard. One is called Rook's weight, in which neighbors share a common border:

$$w_{ij} = \begin{cases} 1, & if\ i\ and\ j\ share\ a\ border \\ 0, & otherwise \end{cases} \tag{2.17}$$

The other is called Queen's weight, in which neighbors share a common border or a vertex:

$$w_{ij} = \begin{cases} 1, & if\ either\ i\ and\ j\ share\ a\ border\ or\ a\ vertex \\ 0, & otherwise \end{cases} \tag{2.18}$$

There are "hybrid" methods of the two approaches often used in practice. The following weight matrices combine the two approaches.

<u>Radial Distance Weights</u>

Let d denote a threshold distance beyond which there is no spatial interaction between sites. The radial distance weight matrix, W, is defined as:

$$w_{ij} = \begin{cases} 1, & 0 \leq d_{ij} < d \\ 0, & d_{ij} < d \end{cases}. \tag{2.19}$$

<u>*k*-Nearest Neighbor Weights</u>

Let k denote the number of closest sites to i, Then for each given k, the k-nearest neighbor weight matrix, W, has spatial weights given by:

$$w_{ij} = \begin{cases} \frac{1}{d_{ij}}, & i, j \text{ are k-nearest points} \\ 0, & \text{otherwise} \end{cases}.$$

<u>Double-Power Distance Weights</u>

If d denotes the maximum radius of influence then the class of double-power distance weights is defined as:

$$w_{ij} = \begin{cases} \left[1 - (\frac{d_{ij}}{d})^{\alpha}\right]^{\alpha}, & 0 \leq d_{ij} < d \\ 0, & d_{ij} < d \end{cases}. \tag{2.21}$$

<u>Boundary Length Weights</u>

Let l_{ij} define the total boundary length shared between i and j and K be the total number of adjacent sites of i. Then the boundary length weights are:

$$w_{ij} = \begin{cases} \frac{l_{ij}}{\sum_{j=1}^{K} l_{ij}}, & \text{if } i \text{ and } j \text{ share a border} \\ 0, & \text{otherwise} \end{cases}. \tag{2.22}$$

<u>Combined Distance-Boundary Weights</u>

In the settings in which a spatial variable exhibits autocorrelation of both distance and boundary length, the combined distance-boundary weights are employed:

$$w_{ij} = \begin{cases} \frac{l_{ij} d_{ij}^{-\alpha}}{\sum_{j=1}^{K} l_{ij} d_{ij}^{-\alpha}}, & if\ i\ and\ j\ share\ a\ border \\ 0, & otherwise \end{cases}. \tag{2.23}$$

In this study, since crop yield of interest is a typical class of regional summary data, we present the Moran's *I* calculated by Queen's weight matrix, which is the specification most widely used in the setting of regional summary data.

2.4 LITURATURE REVIEW

The modeling of crop yield has garnered a great deal of attention in the agricultural economics literature. Many studies have attempted to investigate the best modeling approaches including parametric methods (Ramirez 1997; Ramirez et al. 2003; Sherrick et al. 2004) and nonparametrics methods (Goodwin and Ker, 1998). Most of these studies based on the assumption that crop yield are independently and identically distributed. In 1990s, with yield monitors being commercially available and global positioning systems (GPS) becoming increasingly operational for civilian use, spatial statistics methods were widely used in crop yield models. Lambert et al. (2004) compared OLS regression and four spatial regression methods. Their results suggest that all four spatial regression methods provided similar estimates and outperform OLS. Mainstream econometrics (Coley 1999; Auffhammer et al. 2013; Schlenker and Roberts, 2009) has used the variance-covariance (VC) matrix to represent the autocorrelation relationship in crop yield models. In the case of panel data, estimating the spatial covariance matrix does not require an explicit spatial process or a functional form for the distance decay when the time dimension is considerably greater than the cross-sectional dimension (T >> N). For instance, the seemingly uncorrelated regression (SUR) estimator is able to consider the full spatial covariance matrix. Recently, new methodologies such as copula-based methods (Vedenov, 2008, Woodard et al., 2011 and Goodwin et al., 2014) and Bayesian methods (Ker et al., 2015; Park et al., 2015) have been developed to model the dependence structures of crop yield in probabilistic settings. However, a key underlying assumption of these approaches is the stable spatial variability of crop yields over time, and few papers have discussed the variability of the autocorrelation structure.

CHAPTER 3

DATA AND EMPIRICAL RESULTS

3.1 BACKGROUND

China was the world's largest producer of cotton and largest consumer of cotton fiber, with a share of around 25% of global cotton production (USDA Agricultural Outlook Forum 2014). In 2018, China remains the world's second-largest producer, behind India. The major cotton producing areas in China are comprised of the northwest cotton region (i.e., Xinjiang Province), the Yellow River valley region and the Yangtze River valley region. Approximately 99.7% of cotton produced in China is produced in these three areas.

The diffusion of Bt cotton was rapid in China starting in 1996. Currently, over 90% of the cotton planted in the Yellow River valley region is Bt cotton; non-Bt cotton is concentrated in Xinjiang Province and the Yangtze River valley region where the bollworm pest is not a major peril to the crop. The Yellow River valley region is composed of three provinces: Henan, Hebei, Shandong (known as the Yellow Three). The cotton yield data and geographic information data such as longitude and latitude of the counties comes from the county-level crop yield dataset from the county-level data center of Ministry of Agriculture of China. The study period covers 1980 to 2015. Figure 1 presents a map of the three Provinces in Yellow River Valley and Xinjiang Province.

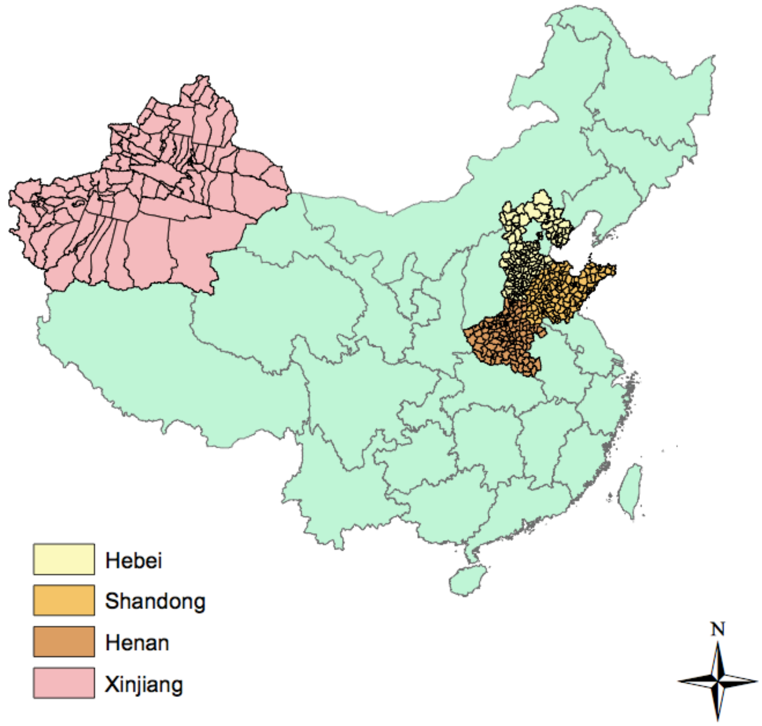


Figure 1. The Map of the Three Provinces in Yellow River Valley and Xinjiang Province

Bt cotton has been formally approved by the government for commercial use in Hebei and Shandong since 1997 and Henan since 1999. Table 1 reports the official Bt cotton adoption rate. Prior to the formal approval, some farmers had already distributed the seeds of Bt cotton varieties cultivated by the National Cotton Research Centre (located in Henan province) from the year 1996, which explains why there was Bt cotton adoption recorded in Henan before approval. However, according to Pemsl (2006), there was a black market in Shandong and Hebei selling Bt cottonseeds prior to approval as far back as 1993. Approximately 16% of farming households in these two provinces adopted Bt cotton prior to the official approval. This unofficial adoption was not reflected in the official data given in Table 1. For the cotton farmers in these three provinces, 1992 was a particularly bad year due to a severe bollworm infestation. This painful experience partially may result in the rapid diffusion of Bt cotton within this area (Liu, 2006).

Table 1. Adoption Rate (in %) of Bt Cotton in the Three Provinces

Year	Shandong	Henan	Hebei	China
1996	-	1.9	-	0.5
1997	0.0	1.1	22.8	2.9
1998	12.6	1.9	70.7	7.6
1999	74.4	19	90.3	21.0
2000	90.2	38.7	100	39.7
2001	100.0	70.3	94.7	45.7
2002	100.0	62.3	95.4	49.1
2003	98.1	78.0	98.3	64.0
2004	100.0	91.7	100.0	69.4
2005	100.0	83.1	100.0	70.9
2006	100.0	82.9	97.0	69.7

Data Source: Fok, Michel, and Naiyin Xu. "Variety market development: A Bt cotton cropping factor and constraint in China." (2011).

Figure 2 represents the county-average yield of the three provinces. Before the introduction of Bt cotton, the yield curves were stationary. As reported in Liu (2006), a huge shortfall can be seen in 1992. Following 1992, the yields recovered quickly and rose rapidly all the way to 2005, and continued to increase slowly to date. The histograms of county-average yield (Figure 3) show pronounced mixed distribution patterns. The normal distributions estimated via mixture model are overlaid on the histogram. Two periods of descriptive statistics of yields are present in Table 2, with 1996 as the cutting point. The average yields increased by 51.7%, 69.6% and 53.2% for Henan, Hebei and Shandong respectively. The results of t-test statistics show that the increases are all statistically significant at 1% confidence level. The data shows that the considerable increase in cotton yield occurs coincidently with the adoption of Bt cotton. However, the portion of the increase in yields that can be explained by the adoption of Bt cotton technology is another important research question.

Table 2. Summary Statistics of Cotton Yield (in ton/ha) before and after Bt Cotton

	Year	Mean	Std. Dev.	Min	Max
Henan	1980~1996	0.60	0.10	0.35	0.74
	1997~2015	0.91	0.09	0.68	1.02
Hebei	1980~1996	0.56	0.15	0.34	0.88
	1997~2015	0.95	0.14	0.56	1.10
Shandong	1980~1996	0.79	0.10	0.62	0.95
	1997~2015	1.21	0.15	0.88	1.44

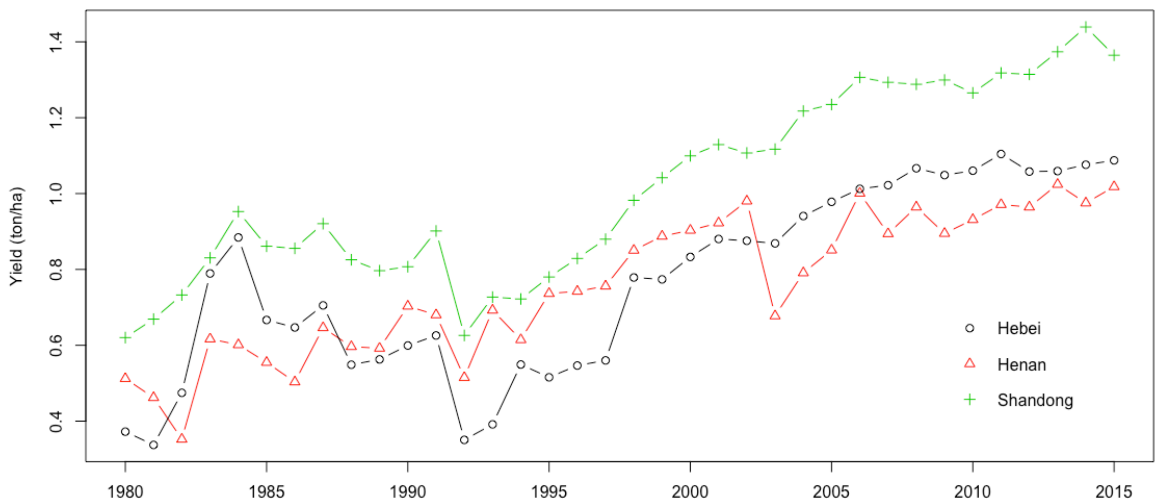


Figure 2. Historical County-Average Cotton Yield of the Three Provinces

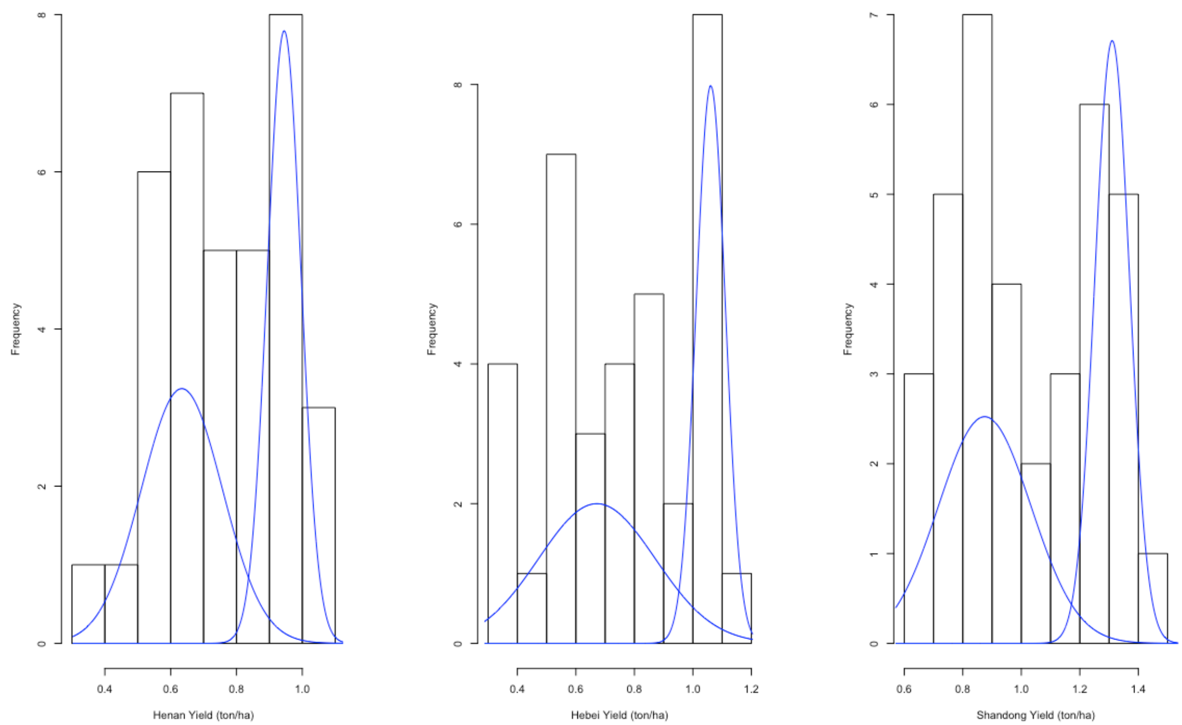


Figure 3. Distribution of County-Average Cotton Yields Fitted with a Mixture Normal Distribution

CHAPTER 4

EMPIRICAL RESULTS

This section presents the main results of the test of the research hypothesis. We use Moran's I to examine the change in spatial autocorrelation of cotton yield over time, and compare this change with the adoption of Bt cotton. To make the cotton yields comparable among different years, the yields were standardized by their temporal average for each year (Blackmore, 2000) by using Equation 4.1.

$$ST\_yield(s,t) = 100 \times \frac{yield(s,t)}{\frac{1}{N}\sum_{s=1}^{N} yield(s,t)} \tag{4.1}$$

where $yield(s,t)$ is the observed cotton yield for county s in year t, and N is the total number of counties. Moreover, the spatial data is not stationary if there is a significant spatial trend. The landscape of China in general is of high elevation in the west inland and descends to the east coast. Accordingly, cotton yield exhibits spatial trend due to the temperature, rainfall and soil conditions. Therefore, in order to remove trend effect in cotton yield, the data is detrend by regressing cotton yields on the easting-northing coordinates:

$$ST\_yield(s,t) = \mu(t) + \alpha * easting(s) + \beta * northing(s) + \varepsilon(s,t) \tag{4.2}$$

and

$$detrend\_yield(s) = \mu + \varepsilon(s) \tag{4.3}$$

where $ST\_yield(s,t)$ is the standardized cotton yield of county s in year t, μ is the portion of yield not spatially dependent in year t, $easting(s)$ and $northing(s)$ are the coordinates of the centroid of county s, $\alpha * easting(s) + \beta * northing(s)$ is a deterministic surface (the trend), and $\varepsilon(s,t)$ is a spatially autocorrelated error.

We examine Moran's *I* for three provinces separately instead of together as a whole. Global Moran's *I* for the detrended cotton yield of each province is shown in in Figure 3. The adoption rate is overlaid on the plots. A pronounced decline in global Moran's *I* can be seen to track with the adoption of Bt cotton in the three provinces, and the statistics show the yields are not spatially correlated at the 5% confidence level after 1996. Take Henan for example, from 1980 to 1999: the Moran's *I* of cotton yield is significantly above zero except the year 1981. With the diffusion of Bt cotton since 1996, the Moran's I continuously decreases to insignificant, meaning spatial autocorrelation gradually diminishes. Note that a nontrivial percentage of farmers in Shandong and Hebei had adopted Bt cotton back to 1994; however, the adoption rate in Figure 4 does not reflect the rate of early adoption before the official approval due to limited availability of black market data. To further examine the change of autocorrelation with the adoption of Bt Cotton, a simple OLS regression of Moran's *I* on the adoption rate was run for each province:

$$Moran's\ I(t) = s + adoption(t) + \epsilon(t), \tag{4.4}$$

where $Moran's\ I(t)$ is the Moran's I of cotton yield for the year t, $adoption(t)$ is the adoption rate, and $\epsilon(t)$ is a random error. In order to focus on any change occurring after the adoption of Bt cotton, we choose *t* from 1993 to 2010.

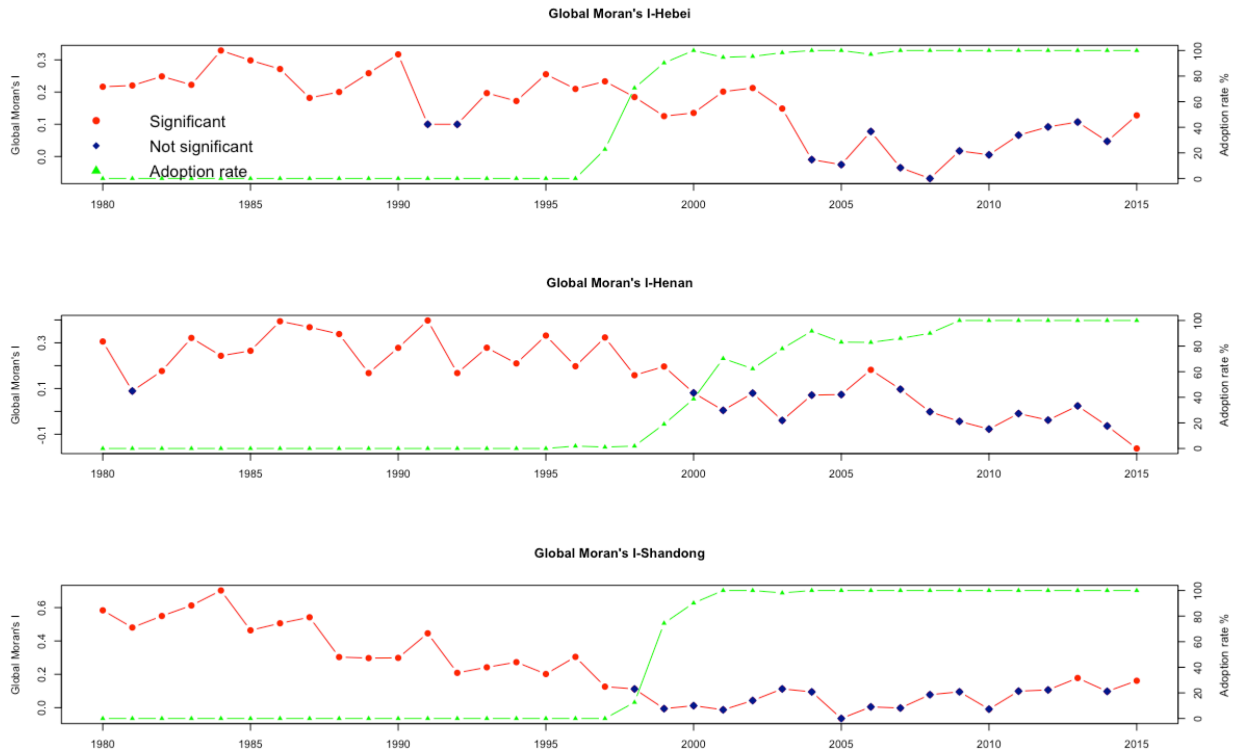


Figure 4. Global Moran’s *I* and Adoption Rate of Bt Cotton for the Three Provinces

Table 3 reports the results of the regressions. For three provinces, the coefficient of the adoption rates are negatively significant at 5% level, indicating that the adoption rate of Bt cotton is negatively associated with the degree of spatial autocorrelation of cotton yield, bolstering our research hypothesis. It is interesting to note that, for three provinces, the coefficients of adoption rates are -0.002 and the intercepts are about 0.2 This value can be interpreted that the Moran's *I* will be close to zero from significantly positive with the adoption increasing from zero to 100%.

Table 3. OLS Regression of Moran's *I* on Bt Cotton Adoption Rate

		Mean	Std. Dev.	t-value	Pr(>|t|)
Henan	Intercept	0.234	0.045	5.205	0.000***
	Adoption rate (in %)	-0.002	0.001	-3.024	0.009**
Hebei	Intercept	0.237	0.029	8.059	0.000***
	Adoption rate (in %)	-0.002	0.000	-5.013	0.000***
Shandong	Intercept	0.184	0.046	3.987	0.004**
	Adoption rate (in %)	-0.002	0.001	-2.862	0.021*

Significant at * $p < 0.05$, ** $p < 0.01$, *** $p < 0.000$

CHAPTER 5

ROBUSTNESS CHECK

This section presents further evidence to check the robustness of the conclusion drawn above. First, we discuss the effect of missing values of crop yield on the estimate of Moran's *I*, showing the number of missing values in our data does not cause any essential difference in the results presented in the above section. Moreover, two additional cases are given. One case is from another province in China where Bt cotton comprises only a small percentage of the cotton acreage: its Moran's *I* does not show a fundamental change during the same period of time. The other is an analogous case from three US states: the trend of increased Bt adoption rate and decreased spatial autocorrelation can be observed in these states. However, due to a large number of missing values in this data, we use these results as a robustness check instead of solid evidence to test the research hypothesis.

5.1 MISSING VALUES

The biasedness of Moran's *I* is jeopardized by the existence of missing values in the yield data. Possible solutions include replacing the missing values with a spatial mean, or spatially interpolated values; and omitting the sites of missing values from the dataset. However, none of these methods is able to fully overcome the potential bias without further information. Conceptually, the spatial interpolation method is likely to result in an upwardly biased estimate of the spatial autocorrelation, and simple omission may lead to cells without neighbors, which are referred to as "isolated islands" and lead to problematic results in spatial analysis. In the context of this study, there are multiple reasons for the missing values. For instance, the missing values may result from no cotton grown in a county for a certain year, or simply from an absent record. Without detailed information about the reason for the missing values, any method used to replace missing values would lead to considerable bias. However, for a data set with a large number of observations, any bias will be small if the percent of missing values is moderate. This study provides evidence that the missing values will lead to a conservative direction towards accepting the null hypothesis. In other words, if the percentage of missing values is moderate, say less than 5%, the bias of Moran's *I* will decrease the probability of accepting the null and result in a lower p-value.

Figure 5 presents the percentage of missing values of the three provinces. A rise in the percentage of missing values can be seen in recent years in Shandong and Hebei. In Henan, the missing value problem was severe prior to 1994. On average, the three provinces of Hebei, Henan, and Shandong have 1.98%, 2.36% and 4.71% missing values, respectively.

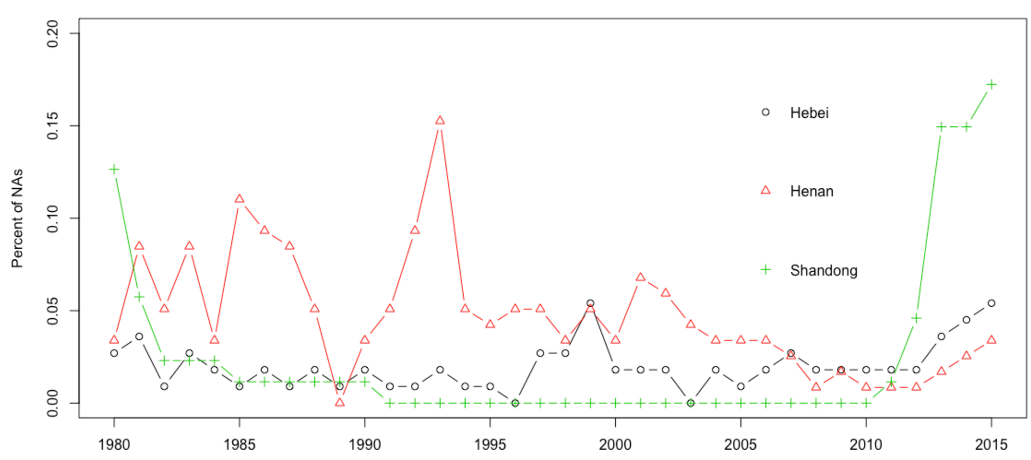
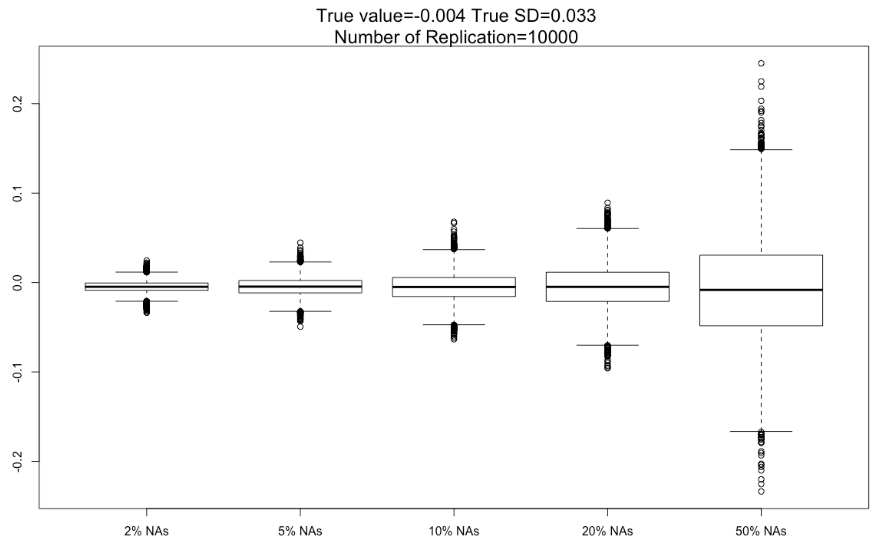
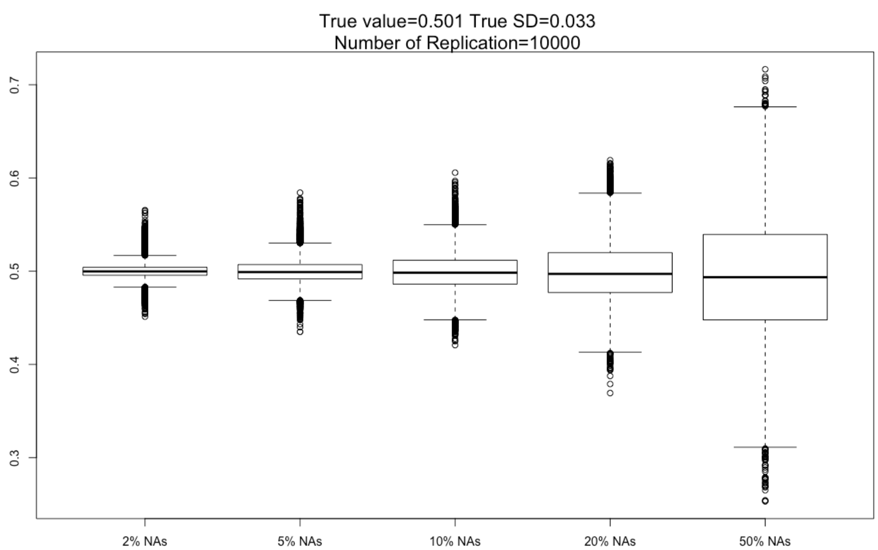


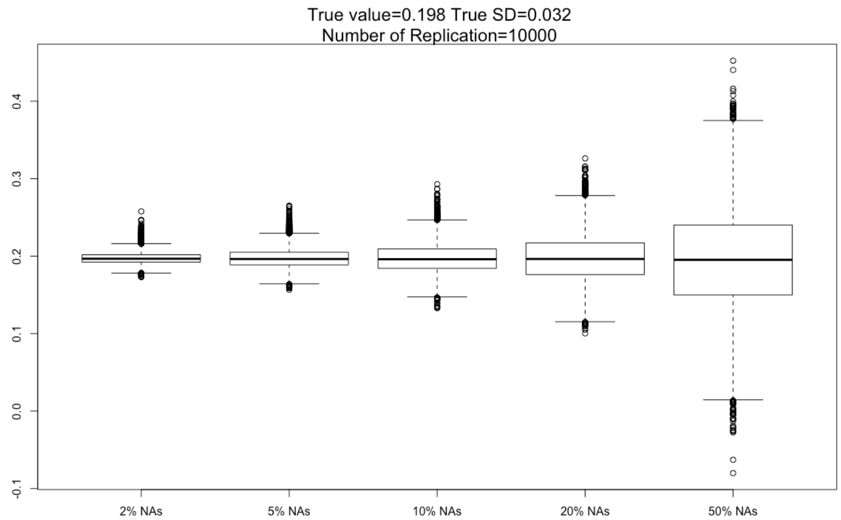
Figure 5. The Percentages of Missing Values of Cotton Yield in the Three Provinces

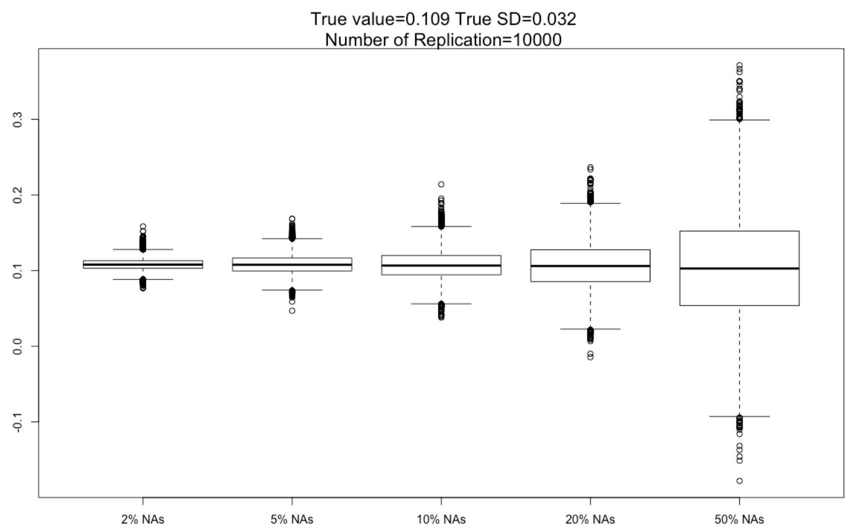
In order to examine the magnitude of the distortion on Moran's I caused by the missing values, we use a Monte Carlo simulation to present the deviation from a hypothesized true mean for varying percentages of missing data. We take the three provinces as a whole so that we have large number of sites (counties) to examine. The simulation was implemented as follows. First, we assume the cotton yield follows a normal distribution. The mean and variance were estimated with all the historical data. Second, we control values of Global Moran's I of cotton yield in these three provinces to be approximately 0.5, 0.2, 0.1, -0.2 and spatial randomness, respectively. We utilize the Global Moran's I calculated in the study period. For example, Global Moran's I was 0.195 in year 2004 and there were 10% counties with missing observations. We then filled the missing counties with spatial interpolation method to boost the Moran's I close to 0.2. Third, at each level, we randomly omit 2%, 5%, 10% 20% and 50% data, and compute the Global Moran's I each case. This step is repeated 10000 times. We also compare between the Global Moran's I with missing values and the "true" Moran's I. The frequencies of Type I error-rejection of a true null hypothesis and Type II error-failure to reject a false null hypothesis in the 10000 times simulation were also calculated. Figure 6 shows the boxplots of the simulation and Table 4 reports the corresponding probabilities of two types of error for each case. The probability of errors increases with the rise of percent of missing value. Global Moran's I is resistant to the potential bias caused by the missing values if there are less than 5% missing and the true value of Moran's I is high, like 0.5. However, if the amount of missing value takes over 50% of the whole dataset and the true value of Moran's I is low, like 0.1. The probability of making type I error will be over 50%. In another word, the result of Moran's I is not trustworthy.

In addition, global Moran's *I* seems easily distorted by missing values if the true value is close to zero.


True value=-0.004 True SD=0.033
Number of Replication=10000
0.2
0.1
0.0
-0.1
-0.2
2% NAs
5% NAs
10% NAs
20% NAs
50% NAs


True value=0.501 True SD=0.033
Number of Replication=10000
0.7
0.6
0.5
0.4
0.3
2% NAs
5% NAs
10% NAs
20% NAs
50% NAs


True value=0.198 True SD=0.032
Number of Replication=10000
0.4
0.3
0.2
0.1
0.0
-0.1
2% NAs
5% NAs
10% NAs
20% NAs
50% NAs


True value=0.109 True SD=0.032
Number of Replication=10000
0.3
0.2
0.1
0.0
-0.1
2% NAs
5% NAs
10% NAs
20% NAs
50% NAs

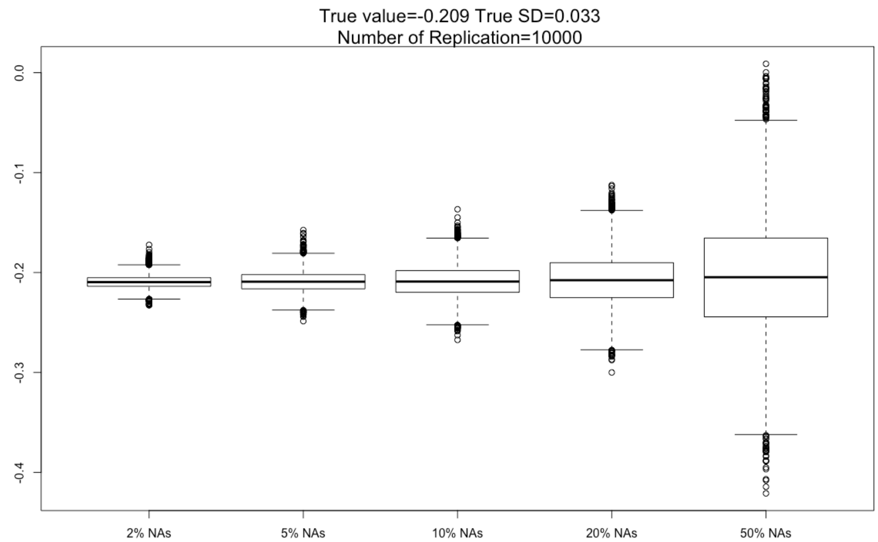


Figure 6. Summaries of the Simulation of Global Moran's I Versus the Percent of Missing Values

Table 4. Probability of errors caused by missing values

True value	1%	5%	10%	20%	50%
The probability of falsely accepting the Null of spatial randomness					
Global Moran's=0.5	0	0	0	0	0
Global Moran's=0.2	0	0	0	0	16.38%
Global Moran's=0.1	0	0.03%	1.95%	17.79%	64.07%
Global Moran's=-0.2	0	0	0	0	12.39%
The probability of falsely not accepting the Null of spatial randomness					
Global Moran's= -0.004	0	0	0	0.12%	2.35%

5.2 TWO COUNTERPARTS-XINJIANG PROVINCE AND THREE U.S. STATES

In order to test the generalization of the conclusion, we examine two counterpart regions. Xinjiang Province has been the largest cotton producing province of China in terms of the cultivation scale, the yield, and the quality (see Figure 7). Due to severe coldness in winter, the incidence of cotton bollworm, cotton aphid and other pests in Xinjiang is not one of the major threats for the cotton producers. In addition, the cotton-planting regions in Xinjiang are distributed across many oases segmented by deserts. These deserts serve as natural firewalls to prevent hazards and disasters on a large scale. The adoption of Bt cotton in Xinjiang is still very limited because of its later introduction and fewer available Bt cotton varieties compared to other regions. Bt cotton has been commercialized in Xinjiang in 1999; the adoption grew extremely slowly to 13% in 2008 (Wang et al., 2015, Pray, 2002). Therefore, we could expect that the spatial pattern is unlikely to alter with the emerging genetically modified technology. As present in Figure 7, the yield of Xinjiang Province grew continuously and smoothly and there was no sudden increase over the past three decades. The Global Moran's *I* of cotton yield in Xinjiang appears to be stationary (see Figure 8). There is no clear decrease in Global Moran's *I* after 1999 as expected. There are 19 years in which Moran's *I* is not significant. Eight of these years happened before the introduction of Bt cotton. The results of Moran's *I* echoes to the facts that cotton yields in Xinjiang Province are independent to each other due to the isolation formed by deserts.

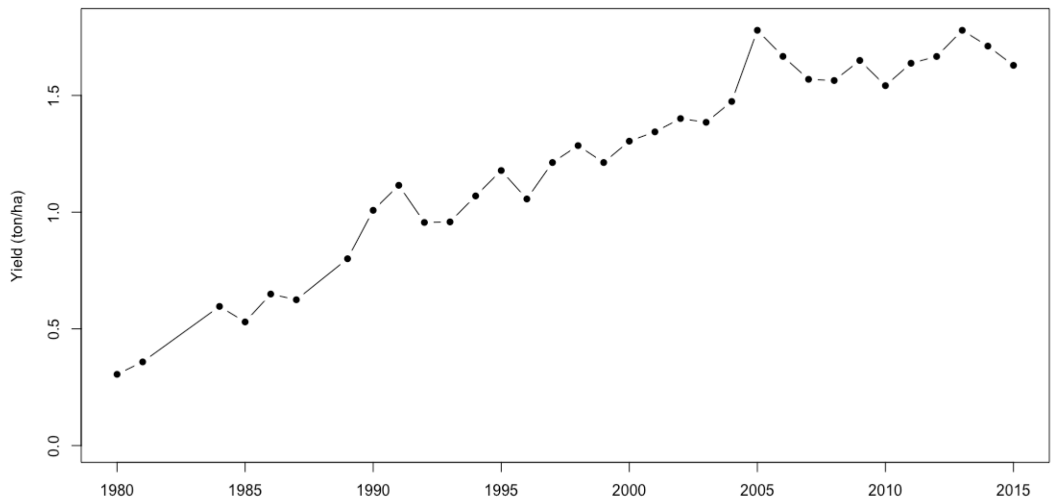


Figure 7. Historical County-Average Cotton Yield of Xinjiang Province

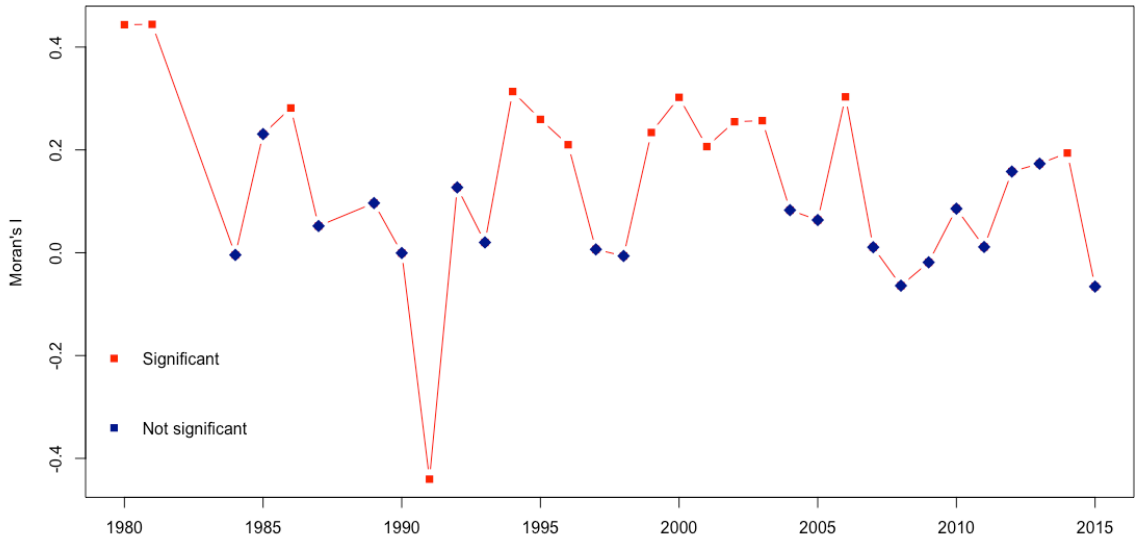


Figure 8. Global Moran's *I* for Xinjiang Province

Bt cotton was quickly diffused into the US cotton industry since it was first marketed in 1996. Producers in the southeast region adopted the new insecticidal cottons most rapidly to reduce the threat caused by the proliferation of a type of insecticide-resistant tobacco budworm. According to Layton et al. (1997), 42% of Mississippi's cotton crop was planted Bt varieties in 1996, the first year of commercialization. Nationwide, Bt cotton acreage expanded from 15% of U.S. cotton acreage in 1997 to 37% in 2001. Currently, 85% of U.S. cotton acres are planted with genetically engineered, insect-resistant seeds. Figure 9 presents the cotton yield of the three states. The average yield of the three states shows a similar pattern as those in Yellow River Valley. After the introduction of Bt cotton, the yield significantly rises. However, we must be very careful interpreting the results for U.S. cotton because approximately 60% of the overall sample is missing data (See Figure 10). The three states of Mississippi, Alabama, and Georgia have relatively fewer missing values than the neighboring states. A large number of the missing values are highly likely to be unreported records rather no cotton production. The simulation results in Chapter 4 shows that Global Moran's *I* is likely to be distorted if the percentage of missing values is over 20%. Due to this significant drawback, we only view the case of these three cotton production states of the U.S. as a supplemental reference, rather than as concrete empirical evidence. Figure 11 presents Global Moran's *I* of cotton yield for the three states as a whole area. Similarly to the story of three Chinese provinces, Global Moran's *I* significantly decreased with the increase in the adoption rate of Bt cotton. Before the year of 2000, the values of Moran's *I* were stable at around 0.5. After that, a downward trend can be seen.

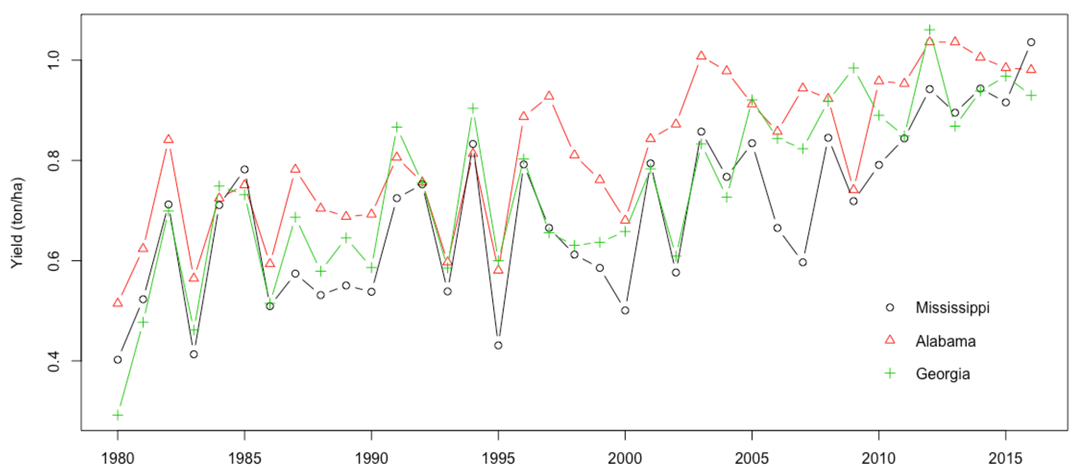


Figure 9. Historical County-Average Cotton Yield of Three States

Figure 10. The Percent of Missing Values in the Three Southeastern U.S. States

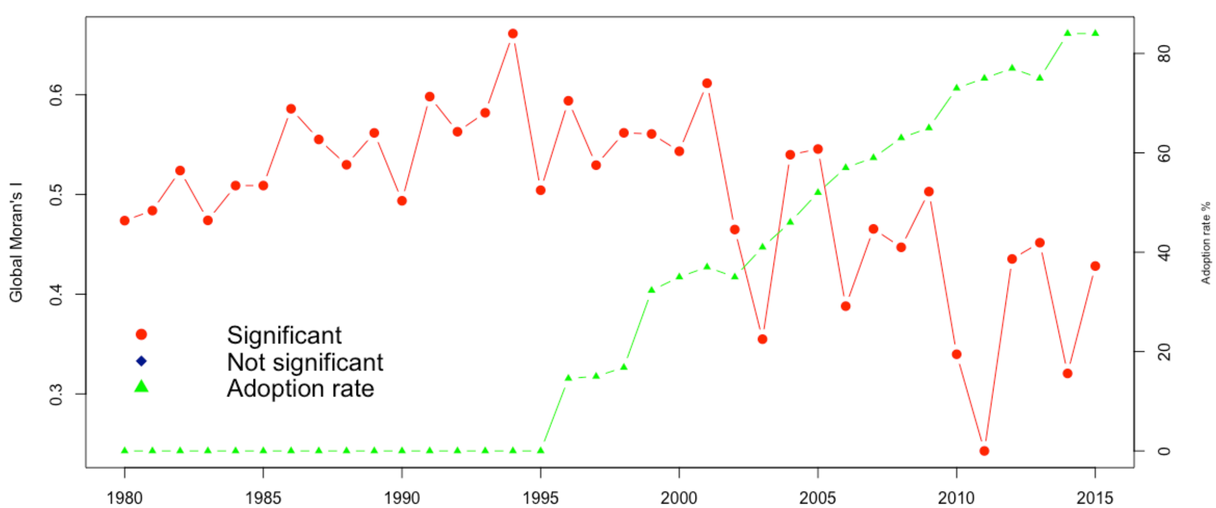


Figure 11. Global Moran's *I* of Cotton Yield for the Three Southeastern U.S. States

CHAPTER 6

APPLICATIONS

6.1 PRECISION AGRICULTURE

Traditional methods of crop production unavoidably lead to over- and under-applications of herbicides, pesticides, irrigation, and fertilizers. Precision agriculture (PA) is an approach to farm management that uses multiple technologies to realize optimal use of inputs with minimum inputs and achieve maximum productivity and healthy environment. It was born with the wide use of GPS in the early 1990s. To date, it has developed a system approach including using sensors to obtain real-time data of the crops, soil and ambient air, weather conditions, along with other relevant information such as the use of fertilizers and pesticides, labor costs and build predictive model.

Site-specific yield monitor data are expected to preform spatial structure, which violates the assumption of classic linear regression model. Therefore, successful application of precision agriculture depends on the understanding of spatial variability of crop yield and the other spatial variables (Pierce and Nowak, 1999). Anselin (2013) suggest two spatial process models: the spatial error model (Equation 2.2) and the spatial lag model (Equation 2.3). Crop yield response models assume crop yield follow the spatial error process.

$$Y = X\beta + \varepsilon, \varepsilon \sim N(0, \sigma^2 \mathrm{V}) \tag{6.1}$$

where Y represents crop yield, X can be a vector of inputs of interest and the error term is follow a spatial pattern characterized by $\sigma^2 V$.

When observations are available across space as well as over time, the additional dimension allows the estimation of the full covariance of one type of association, using the other dimension to provide the asymptotics (Anselin, 2013). However, this study suggests it is possible that V is a function of time as follows rather than simply invariant over time.

$$\varepsilon_t \sim N(0, \sigma^2 \mathrm{V}_t) \tag{6.2}$$

or

$$\varepsilon_t \sim N(0, {\sigma_t}^2 \mathrm{V}) \tag{6.3}$$

Before using temporal data to estimate a spatial covariance matrix, it is necessary to examine the spatial diagnostic statistics over time and make sure the spatial autocorrelation pattern is stable.

6.2 REDUCING CROP INSURANCE PRICE

Accurate modeling of crop yield is critical for the proper design of crop insurance contracts and the maintenance of profitable insurance programs. The spatial autocorrelation of crop yield or crop loss is considered as systemic risk, which is related to contract designing, premium setting and reinsurance strategies. As far as our best knowledge, there are few studies that discuss that technology change affects crop insurance premium through reducing spatial autocorrelation of yield.

Ideally, an insurable loss is independent among the insured units, and thus the insured units do not suffer loss at the same time. Auto insurance and home insurance typically covers this type of loss. As for crop insurance, if yield losses within a region, say a state or a province, are highly correlated with each other, it is highly likely that a large number of counties suffer losses in a year. As a result, insurance companies are not able to sufficiently diversify its risk by increasing the unit of insurance, i.e. selling more contracts in this region. This means insurance companies may face a huge amount of claims at once. In order to cope with catastrophic risk, insurance companies have to reserve a high portion of the collected premium or pay a high price to reinsure the risk. The associated cost is called a catastrophic risk loading, and is paid by farmers or the public in terms of premium subsidy or catastrophic plan. In a word, less spatial autocorrelation means lower premium.

The spatial autocorrelation effect on an insurance premium can be understood by examining insurer's risk exposure or actual payouts. Take as an example the three Yellow River provinces in this study. An insurance company insures the cotton yield of all the counties within these provinces at an actuarially fair price, meaning a price does not take into account administrative cost and profit. Assuming the insurance premium is set as a province-average yield, that is 100 for the standardized yield and 80% coverage level, we calculate its potential payout (known as indemnity) each year as:

$$Indemnity(t) = \sum_{s=1}^{N}(ST\_yield(s,t) - 100 * Coverage) * Ind(ST_{yield(s,t)} < 100 * Coverage) \qquad (6.2)$$

where $Indemnity(t)$ represents the indemnity occurring at year t, $ST\_yield(s,t)$ is the standardized yield value of county s in year t, N is the total number of counties, $Coverage$ represents coverage level, and $Ind(ST_{yield(s,t)} < 100 * coverage\ level)$ is the indicator function equal to 1 if standardized yield is less than the average standardized yield times coverage level. Here we set coverage level at 80%.

Figure 11 depicts the insurance payouts for the three provinces. To take into account the change in the number of counties planted in cotton, the payouts were divided by the number of non-missing observations. This value can be roughly interpreted as the actuarially fair premium under this simplified framework. A salient decrease can be observed after 1996, the year Bt cotton was commercially available. The decreased average payouts or premium implies that the yield variability at the county level is small and a fewer number of counties have cotton yields less than the insured coverage. The change in average insurance payouts before and after 1996, the year of approval Bt cotton, was assessed by analysis of variance (ANOVA). Table 5 reports the results of the ANOVA for the payouts on the adoption of Bt cotton. A significant difference ($P<0.001$) of average insurance payouts can be seen before and after the approval of Bt cotton. The result shows that the negative variation (less than the county-average) of crop yield among counties significantly decreased after the year of 1996. In another word, as expected, the resulting average insurance payouts decreased significantly.

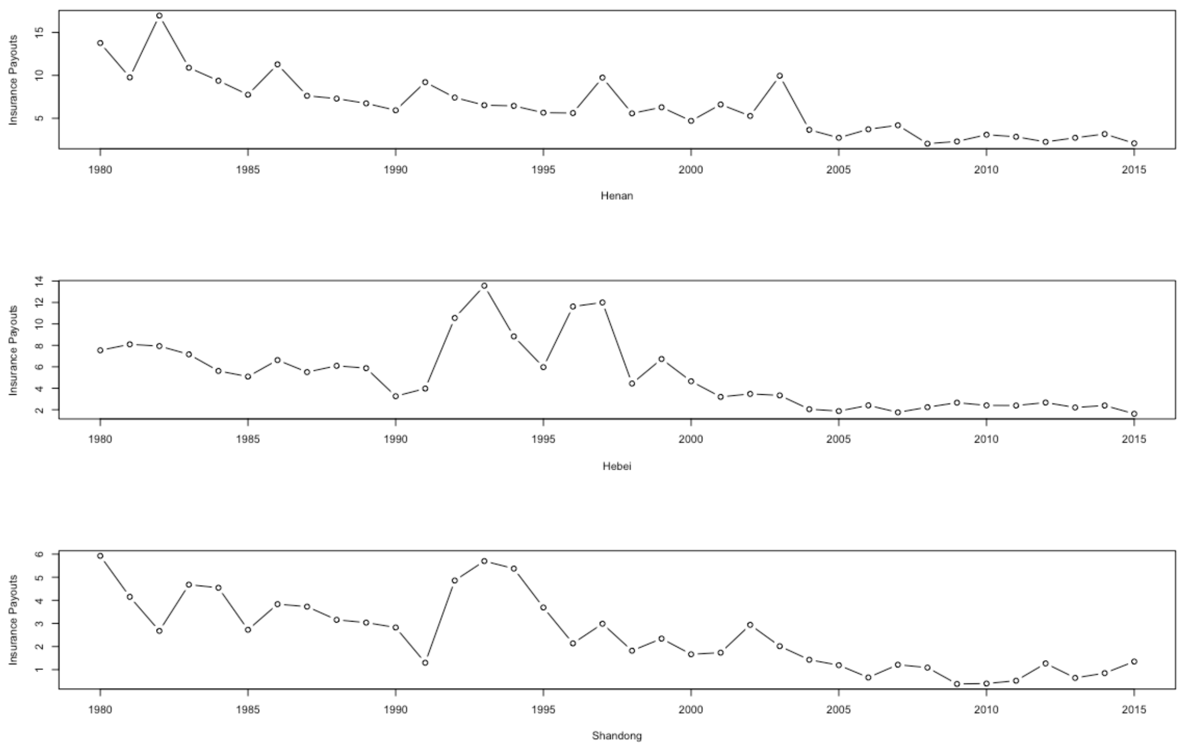


Figure 12. Average Insurance Payouts for the Three Yellow River Provinces

Table 5. ANOVA Results on the Adoption of Bt Cotton

		DF	Sum Square	Mean Square	F-value	Pr(>F)
Henan	Adopt Bt cotton	1	170.1	170.11	22.66	0.000***
	Residuals	34	255.2	7.51		
Hebei	Adopt Bt cotton	1	133.7	133.66	20.52	0.000***
	Residuals	34	221.5	6.52		
Shandong	Adopt Bt cotton	1	51.23	51.23	46.44	0.000***
	Residuals	34	37.51	1.1		

Significant at * $p < 0.05$, ** $p < 0.01$, *** $p < 0.000$

CHAPTER 7

CONCLUSION AND DISCUSSION

The implication of our analysis is twofold. First, this work provides a cautionary note on modeling crop yield based on the assumption of stable spatial autocorrelation. It is widely agreed that there is a spatial structure to crop yield distributions and considering the spatial structure may provide more accurate inference and prediction. To model the spatial structure of crop yield with spatial-temporal data, it is often assumed that the spatial autocorrelation of the crop yield is spatially invariant. However, the finding of this study suggests that spatial autocorrelation might not always be stable over a long period of time. In particular, the spatial autocorrelation may change with the introduction of significant advancements in agricultural technology such as genetically modified seeds. Therefore, it is necessary to be watchful of the varying spatial autocorrelation patterns within the sample period. This thesis proposes as the use of Moran's *I* to check whether there is salient change in the pattern of spatial autocorrelation of crop yield within the sample period.

Adding to the large body of literature on the impacts of the diffusion of Bt cotton, this thesis points out an unexplored potential impact, which is that the adoption of Bt cotton is likely to reduce the spatial autocorrelation of cotton yield within an area. In combination with other literature on the effects of Bt cotton on reducing the agricultural intensification and increasing biodiversity, we might expect the diffusion of Bt cotton to tend to increase resistance of cotton to adverse natural shocks and geographic conditions.

REFERENCES

Anselin, L. and Bera, A.K., 1998. Introduction to spatial econometrics. *Handbook of Applied Economic Statistics,* p.237.

Anselin, L., 1995. Local indicators of spatial association—LISA. *Geographical Analysis,* 27(2), pp.93-115.

Anselin, L., 2013. Spatial econometrics: methods and models *(Vol. 4). Springer Science and Business Media*.

Auffhammer, M., Hsiang, S.M., Schlenker, W. and Sobel, A., 2013. Using weather data and climate model output in economic analyses of climate change. *Review of Environmental Economics and Policy,* 7(2), pp.181-198.

Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society*, Series B 36, 192–236.

Blackmore, S., 2000. The interpretation of trends from multiple yield maps. *Computers and Electronics in Agriculture*, 26(1), pp.37-51.

Cattaneo, M.G., Yafuso, C., Schmidt, C., Huang, C.Y., Rahman, M., Olson, C., Ellers-Kirk, C., Orr, B.J., Marsh, S.E., Antilla, L. and Dutilleul, P., 2006. Farm-scale evaluation of the impacts of transgenic cotton on biodiversity, pesticide use, and yield. *Proceedings of the National Academy of Sciences,* 103(20), pp.7571-7576.

Chen, R., Huang, J. and Qiao, F., 2013. Farmers' knowledge on pest management and pesticide use in Bt cotton production in China. *China Economic Review*, 27, pp.15-24.

Cliff, A. and Ord, K., 1972. Testing for spatial autocorrelation among regression residuals. *Geographical Analysis,* 4(3), pp.267-284.

Conley, T.G., 1999. GMM estimation with cross sectional dependence. *Journal of Econometrics,* 92(1), pp.1-45.

Cressie, N. (1985). A geostatistical analysis of the mercer and hall wheat yield data. *Bulletin of the International Statistical Institute*, Contributed Papers 51, 277–78.

Cressie, N., 1992. Statistics for spatial data. *Terra Nova,* 4(5), pp.613-617.

Donegan, K.K., Palm, C.J., Fieland, V.J., Porteous, L.A., Ganio, L.M., Schaller, D.L., Bucao, L.Q. and Seidler, R.J., 1995. Changes in levels, species and DNA fingerprints of soil microorganisms associated with cotton expressing the Bacillus thuringiensis var. kurstaki endotoxin. *Applied Soil Ecology*, 2(2), pp.111-124.

Fok, M. and Xu, N., 2011. Variety market development: a Bt cotton cropping factor and constraint in China.

Geary, R.C., 1954. The contiguity ratio and statistical mapping. *The Incorporated Statistician,* 5(3), pp.115-146.

Getis, A. and Ord, J.K., 1992. The analysis of spatial association by use of distance statistics. *Geographical Analysis,* 24(3), pp.189-206.

Getis, A., 1991. Spatial interaction and spatial autocorrelation: a cross-product approach. *Environment and Planning* A, 23(9), pp.1269-1277.

Getis, A., 2009. Spatial weights matrices. *Geographical Analysis,* 41(4), pp.404-410.

Goodwin, B.K. and Hungerford, A., 2014. Copula-based models of systemic risk in US agriculture: implications for crop insurance and reinsurance contracts. *American Journal of Agricultural Economics*, 97(3), pp.879-896.

Goodwin, B.K. and Hungerford, A., 2014. Copula-based models of systemic risk in US agriculture: implications for crop insurance and reinsurance contracts. *American Journal of Agricultural Economics,* 97(3), pp.879-896.

Goodwin, B.K. and Ker, A.P., 1998. Nonparametric estimation of crop yield distributions: implications for rating group-risk crop insurance contracts. *American Journal of Agricultural Economics,* 80(1), pp.139-153.

Gouse, M., Pray, C.E., Kirsten, J. and Schimmelpfennig, D., 2005. A GM subsistence crop in Africa: the case of Bt white maize in South Africa. *International Journal of Biotechnology,* 7(1-3), pp.84-94.

Griffin, T.W., 2010. The spatial analysis of yield data. *In Geostatistical Applications for Precision Agriculture* (pp. 89-116). Springer, Dordrecht.

Huang, J., Hu, R., Rozelle, S., Qiao, F. and Pray, C.E., 2002. Transgenic varieties and productivity of smallholder cotton farmers in China. *Australian Journal of Agricultural and Resource Economics,* 46(3), pp.367-387.

Hubert, L.J., Golledge, R.G. and Costanzo, C.M., 1981. Generalized procedures for evaluating spatial autocorrelation. *Geographical Analysis*, 13(3), pp.224-233.

James, C., 2003. Global review of commercialized transgenic crops. *Current Science,* 84(3), pp.303-309.

Ker, A.P., Tolhurst, T.N. and Liu, Y., 2015. Bayesian estimation of possibly similar yield densities: implications for rating crop insurance contracts. *American Journal of Agricultural Economics,* 98(2), pp.360-382.

Kathage, J. and Qaim, M., 2012. Economic impacts and impact dynamics of Bt (Bacillus thuringiensis) cotton in India. *Proceedings of the National Academy of Sciences,* 109(29), pp.11652-11656.

Kouser, S. and Qaim, M., 2015. Bt cotton, pesticide use and environmental efficiency in Pakistan. *Journal of Agricultural Economics*, 66(1), pp.66-86.

Lambert, D.M., Lowenberg-Deboer, J. and Bongiovanni, R., 2004. A comparison of four spatial regression models for yield monitor data: A case study from Argentina. *Precision Agriculture,* 5(6), pp.579-600.

Layton, M.B., Williams, M.R. and Stewart, S., 1997. BT-cotton in Mississippi: the first year. *In 1997 Proceedings Beltwide Cotton Conferences,* New Orleans, LA, USA, January 6-10, 1997: Volume 2. (pp. 861-863). National Cotton Council.

Liu, E.M., 2013. Time to change what to sow: Risk preferences and technology adoption decisions of cotton farmers in China. *Review of Economics and Statistics,* 95(4), pp.1386-1403.

Maestrini, B. and Basso, B., 2018. Predicting spatial patterns of within-field crop yield variability. *Field Crops Research,* 219, pp.106-112.

Matheron, G., 1963. Principles of geostatistics. Economic Geology, 58(8), pp.1246-1266.

Moran, P.A., 1948. The interpretation of statistical maps. *Journal of the Royal Statistical Society*. Series B (Methodological), 10(2), pp.243-251.

Moran, P.A., 1950. Notes on continuous stochastic phenomena. *Biometrika,* 37(1/2), pp.17-23.

Myint, S.W., 2003. Fractal approaches in texture analysis and classification of remotely sensed data: comparisons with spatial autocorrelation techniques and simple descriptive statistics. *International Journal of Remote Sensing*, 24(9), pp.1925-1947.

Ord, J.K. and Getis, A., 2001. Testing for local spatial autocorrelation in the presence of global autocorrelation. *Journal of Regional Science,* 41(3), pp.411-432.

Park, E., Brorsen, B.W. and Harri, A., 2018. Using Bayesian Kriging for Spatial Smoothing in Crop Insurance Rating. *American Journal of Agricultural Economics.*

Pemsl, D.E., 2006. Economics of agricultural biotechnology in crop protection in developing countries: the case of Bt-cotton in Shandong province, China. *University of Hannover, Faculty of Economics and Management.*

Pierce, F.J. and Nowak, P., 1999. Aspects of precision agriculture. *In Advances in agronomy* (Vol. 67, pp. 1-85). Academic Press.

Ping, J.L., Green, C.J., Zartman, R.E. and Bronson, K.F., 2004. Exploring spatial dependence of cotton yield using global and local autocorrelation statistics. *Field Crops Research,* 89(2-3), pp.219-236.

Pray, C.E., Schimmelpfennig, D. and Gouse, M., 2004. The distribution of benefits from Bt cotton adoption in South Africa.

Qaim, M., Subramanian, A., Naik, G. and Zilberman, D., 2006. Adoption of Bt cotton and impact variability: Insights from India. *Review of Agricultural Economics*, 28(1), pp.48-58.

Ramirez, O.A., 1997. Estimation and use of a multivariate parametric model for simulating heteroskedastic, correlated, nonnormal random variables: the case of corn belt corn, soybean, and wheat yields. *American Journal of Agricultural Economics,* 79(1), pp.191-205.

Ramirez, O.A., Misra, S. and Field, J., 2003. Crop-yield distributions revisited. *American Journal of Agricultural Economics,* 85(1), pp.108-120.

Robinson, A.H., 1956. The necessity of weighting values in correlation analysis of areal data. *Annals of the Association of American Geographers,* 46(2), pp.233-336.

Sarkar, B., Patra, A.K., Purakayastha, T.J. and Megharaj, M., 2009. Assessment of biological and biochemical indicators in soil under transgenic Bt and non-Bt cotton crop in a sub-tropical environment. *Environmental Monitoring and Assessment,* 156(1-4), pp.595-604.

Schlenker, W. and Roberts, M.J., 2009. Nonlinear temperature effects indicate severe damages to US crop yields under climate change. *Proceedings of the National Academy of sciences*, 106(37), pp.15594-15598.

Sherrick, B.J., Zanini, F.C., Schnitkey, G.D. and Irwin, S.H., 2004. Crop insurance valuation under alternative yield distributions. *American Journal of Agricultural Economics,* 86(2), pp.406-419.

Thomas, E.N. and Anderson, D.L., 1965. Additional comments on weighting values in correlation analysis of areal data. *Annals of the Association of American Geographers,* 55(3), pp.492-504.

Vedenov, D.V., 2008. Application of copulas to estimation of joint crop yield distributions (No. 382-2016-22375).

Velmourougane, K. and Blaise, D., 2017. Impact of transgenic Bt cotton on soil health. *CAB Reviews.*

Wall, M.M., 2004. A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference,* 121(2), pp.311-324.

Wang, X., Xiang, C. and Huang, J., 2015. Adoption and uptake pathway of GM technology by Chinese smallholders: evidence from Bt cotton production.

Woodard, J.D., Paulson, N.D., Vedenov, D. and Power, G.J., 2011. Impact of copula choice on the modeling of crop yield basis risk. *Agricultural Economics,* 42, pp.101-112.