A Spatio-temporal Model for Agricultural Yield Prediction

Oleksiy Tokovenko[1]
Jeffrey H. Dorfman
Lewell F. Gunter

[1]Oleksiy Tokovenko (`leksa@uga.edu`) is a graduate research assistant and Jeffrey H. Dorfman (`jdorfman@uga.edu`) and Lewell F. Gunter (`lgunter@uga.edu`) are Professors in the Department of Agricultural and Applied Economics, College of Agricultural and Environmental Sciences, University of Georgia

A Spatio-temporal Model for Agricultural Yield Prediction

Oleksiy Tokovenko   Jeffrey H. Dorfman   Lewell F. Gunter

Abstract

The paper presents a spatio-temporal statistical model of agricultural yield prediction based on spatial mixtures of distributions. The proposed method combines several hierarchical and sequential Bayesian estimation procedures that allow the general problem to be addressed with a series of simpler tasks, providing the required flexibility of the model while decreasing the complexity associated with the large dimensionality of the spatial data sets. The data used for the study are 1970 - 2009 annual Iowa state county level corn yield data. The spatial correlation hypothesis is studied by comparing the alternative models using the posterior predictive criterion under squared loss function.

Research in progress. Do not quote without authors' permission.

Introduction

Despite the extensive knowledge accumulated over time in the field of modeling crop yield distributions it remains an important area of research due to its role in modern agricultural economics. Accurate information about the behavior of crop yields is a key component of successful policy applications in many areas of agribusiness and finance such as farm decision planning, designing agricultural insurance and government supported policy making. A long-standing result concluded from the empirical studies in the field is the rejection of the assumption about the normality of crop yield distribution in favor of the various nonsymmetric alternatives such as the beta (e.g., Nelson and Preckel 1989, Hennesy, Babcock

and Hayes 1997), the gamma (e.g., Gallagher 1987), the log-normal (e.g. Goodwin, Roberts and Coble 2000) and SU family distributions (e.g. Moss and Shonkwiler (1994), Ramirez 1997). In addition a variety of nonparametric and semi-parametric solutions to the problem were also offered in the literature (e.g. Goodwin and Ker 1998, Ker and Coble 2003 and Racine and Ker 2004). In particular, Norwood, Roberts and Lusk (2004) found the method of Goodwin and Ker (1998) to out-perform other models in out-of-sample prediction power. However there is still no consensus as to what yield model is superior for empirical work, since the results of normality and performance tests depend significantly on the variety of assumptions and the specifications as well as the data used for each study (see e.g. Ramirez and McDonald 2006 for a comment on the Norwood, Roberts and Lusk 2004 result and Just and Weninger 1999 for a discussion of methodologi-cal problems that occur in typical crop yield distribution analyses that can make the validity of results questionable). An interest in the spatio-temporal compo-nent of the yield models emerged significantly in recent years (e.g. Wang and Zhang 2003, Ozaki, Ghosh, Goodwin and Shirota 2008, Harri, Erdem, Coble and Knight 2009 and Ozaki and Silva 2009). The spatio-temporal approach to crop yield modeling allows increasing the scale of the studied problems and carrying on the analysis in its full efficiency by avoiding the errors of aggregation through the proper use of spatial information. However the computational complexity of the spatio-temporal methods remains a problem since it often imposes restrictions on one of the components of the analysis – spatial, temporal or distributional.

The objective of this study is to develop a method for modeling crop yield distributions that will allow one to characterize their dynamic behavior by in-corporating spatial information to increase the efficiency of analysis and make it available for disaggregate levels of data while retaining the flexibility of the

shapes of the crop yield distribution for each spatial unit. The conditional nature of the underlying estimation algorithm assumes that only a small problem will be addressed at a time, providing the advantage of maintaining the efficiency and feasibility of the analysis in large scale models with computational complexity growing linearly in the number of spatial units included.

Model

The proposed approach to agricultural yield prediction models the yield distribution of interest as a spatial mixture of unobserved dynamic processes distributed normally at each period of time such that

$$p(y_{it}) = \sum_{j \in A_i} w_{ij} \phi(\mu_{jt}, \sigma_j^2) \tag{1}$$

where $\phi(\mu, \sigma^2)$ denotes the normal density function with mean μ and variance σ^2. A set of spatial unit indices A_i defines the spatial neighborhood for unit i, with $i \in A_i$, describing the primal spatial relation between y_i and μ_j's as shown in Fig 1(a). The neighbors' contribution structure is completely described by the set of spatial weights w such that $\sum_{j \in A_i} w_{ij} = 1$ and is assumed to be constant over time. Let B_i denote a set of spatial unit indices j for which $i \in A_j$, that defines the spatial neighborhood for unit i latent process and describes the dual relation between μ_i and y_j's as shown in the Fig 1(b). Let us denote the number of members of A_i (and, correspondingly, the number of mixture components of $p(y_i)$) as k_i and the number of members of B_i as m_i. The nature of the spatial mixture definition of yield distribution (1) assumes that only one of k_i latent processes μ_j, $j \in A_i$, will actually contribute to the realization of y_i at any given moment of time $t =$

3

$1, \ldots, T$. Therefore, at any given moment of time t only a subset $B_{it} \subseteq B_i$ of the spatial units $j \in B_i$ will be active recipients of the latent signal μ_i. The number of members in B_{it} is thus assumed to vary over time and can be denoted as $m_{it} \leq m_i$, with $m_{it} = 0$ meaning that the spatial unit i is not providing information actively to any distribution, while $m_{it} = m_i$ implying its complete spatial contribution. The
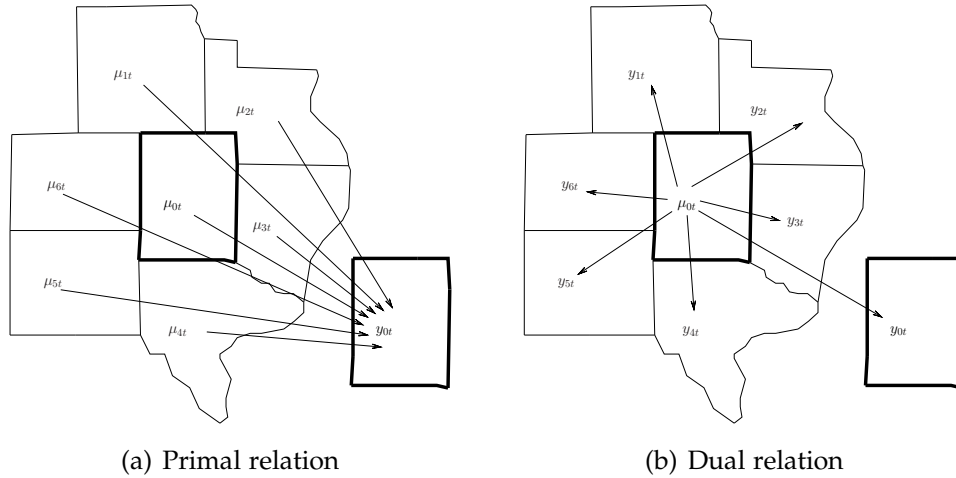

(a) Primal relation   (b) Dual relation

Figure 1: Spatial relations: a) primal, latent information to observed and b) dual, observed information to latent.

stochastic trend specification of the latent spatial processes estimated in this study is that of a local linear type which is quite general and shown to be well suited for a variety of applications including agricultural yield prediction (see, e.g., Moss and Shonkwiler 1993). It consists of a set of m_{it} measurement equations

$$y_{jt} = \mu_{it} + \epsilon_{jt} \tag{2}$$

where $\epsilon_{jt} \sim N(0, \sigma_i^2)$, for all $j \in B_{it}$, and two transition equations that govern the dynamics of the unobserved spatial unit mean μ_{it}

$$\mu_{it} = \mu_{it-1} + \eta_{it-1} + \nu_{it} \tag{3}$$

$$\eta_{it} = \eta_{it-1} + \xi_{it} \tag{4}$$

where $\nu_{it} \sim N(0, \delta_i^2)$ and $\xi_{it} \sim N(0, \gamma_i^2)$, such that $E(\nu_{it}, \xi_{is}) = 0$ for all t and s. For compactness of notation, let us denote $\alpha_{it} = \{\mu_{it}, \eta_{it}\}$ and $w_{it} = \{\nu_{it}, \xi_{it}\}$ to be stacked 2×1 vectors of state variables and state errors, respectively. Similarly, let $u_{it} = \{y_{jt}\}_{j \in B_{it}}$ and $v_{it} = \{\epsilon_{jt}\}_{j \in B_{it}}$ be $m_{it} \times 1$ stacked vectors of observations in (2) and measurement errors associated with them. Then using the vector notation introduced above, equations (2) - (4) form the following state-space model:

$$u_{it} = H_{it}\alpha_{it} + v_{it} \tag{5}$$

$$\alpha_{it} = F\alpha_{it-1} + w_{it} \tag{6}$$

where $v_{it} \sim N(0, R)$ and $w_{it} \sim N(0, Q)$, such that $E(v_{it}, w_{is}) = 0$ for all t and s. Matrices H_t and R in (5) are defined as follows

$$H_{it} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \end{bmatrix} = [i_{m_{it}} 0_{m_{it}}] \quad \text{and} \quad R = \begin{bmatrix} \sigma_i^2 & 0 & \dots & 0 \\ 0 & \sigma_i^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_i^2 \end{bmatrix} = \sigma_i^2 I_{m_{it}} \tag{7}$$

and matrices F and Q in (6) are defined as follows

$$F = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad Q = \begin{bmatrix} \delta_j^2 & 0 \\ 0 & \gamma_j^2 \end{bmatrix} \tag{8}$$

Estimation

The Bayesian treatment of the finite mixture models suggests augmenting the data likelihood (1) with the set of mixture component labels, $\{z_{ijt}\}$, where $\{z_{ijt}\} = 1$ indicates that the observation y_{it} is generated from the jth labeled component of the mixture distribution (see, e.g. Koop, Porier and Tobias (2007) for details). In the context of this study, $\{z_{ijt}\} = 1$ implies that the latent process j contributed to the observed realization of the ith spatial unit yield y_i at time t. The resulting expression for the model parameters likelihood conditional on the values of the latent mixture component indicators is

$$L(\Gamma, \{\mu_i\}_i^N; \{y_i\}_i^N) = \prod_{t=1}^{T} \prod_{j \in A_i} \phi(y_{it}|\mu_{jt}, \sigma_j^2)^{z_{ijt}} \tag{9}$$

where Γ denotes a set of all variance parameters of the model. For computational purposes (9) has a more convenient form than the original unconditional likelihood. The model is completed by choosing the set of prior distribution specifications, where two issues has to be considered when working with the mixtures of Normal densities, as discussed in Koop (2006, Sec. 10.3.3). First, the likelihood function for this class of models is unbounded and therefore informative priors are required. Second, the likelihood function is also invariant under relabeling of the mixture components. As the result, any of $k!$ combinations of possible labeling of the k mixture components will yield the same likelihood function. This

second property of the mixture of Normals models, called "label switching", is essentially an identification problem and can be irrelevant in cases where the researcher is only interested in analyzing the quantities and functions based explicitly on the likelihood value. However, if we do not have enough prior information to distinguish between the mixture components the invariance of the likelihood to all possible permutations of parameter vectors will lead to a posterior distribution which is also the same for all possible combinations of component labels. One of the conventional ways of dealing with the label switching is to choose the prior distribution that will impose labeling restriction through the identifiability constraints on the model parameter space, such as $\sigma_1^2 < \sigma_2^2 < \cdots < \sigma_N^2$, $\mu_1^2 < \mu_2^2 < \cdots < \mu_N^2$ or $w_1^2 < w_2^2 < \cdots < w_N^2$, where only one such constraint is required. In many cases the choice of constraint is naturally suggested by either the underlying economic theory or the type of the mixture used. However, there is no obvious strict ordering of the parameter space, such as the ones discussed above, for the problem studied here. Indeed, there is not enough information to believe that the variation in yield realization is always higher for one spatial unit than the other and that such a relation exists for all counties and defines the strict ranking of σ_i^2, $i = 1, \ldots, N$. Similarly, identifying the mixture components by restricting the latent process space requires imposing very strong assumptions that lack the formal statistical or economic logic to support them. We suggest that the solution to the identification problem can be found by examining the structure of spatial weights. It still remains difficult to establish the strict ordering of weights within any given spatial neighborhood. At the same time the initial definition of the problem clearly implies that the own effect of the latent spatial process μ_i must dominate the effects of contributions of the rest of spatial neighbors. Formally this condition can be stated as the following inequality restriction, $w_{ii} > w_{ij}$

7

for all i and $j \in A_i$. Note, that the established inequality will only allow one to differentiate the own mixture component from the rest of the contributors which is generally not enough for identification purposes. However, applying this constraint to each of $i = 1, \ldots, N$ sets of spatial weights w_i provides the required N identifying conditions for each of corresponding N elements of parameter space.

We fit the model using a Gibbs sampler with data augmentation where the posterior simulations are being conducted by iteratively drawing according to Steps $1 - 5$ below.

Step 1: $\{z_i\}_i^N | \Gamma, \{\mu_i\}_i^N, y$

The natural choice of prior distribution for the mixture component labels vectors $z_{it} = \{z_{ijt}\}_{j \in A_i}$ is the multinomial distribution $M(1, w_i)$. Combining this information with the augmented likelihood yields the following posterior densities of z_{it}

$$z_{it} \sim M \left(1, \left\{ \frac{w_{ij}\phi(y_{it}|\mu_{it},\sigma_j^2)}{\sum_{j \in N_i} w_{ij}\phi(y_{it}|\mu_{jt},\sigma_j^2)} \right\}_{j \in A_i} \right) \tag{10}$$

Step 2: $\{\mu_i, \nu_i\}_i^N | \Gamma, \{z_i\}_i^N, y$

Let $p(\alpha_i) = p(\alpha_{i1}, \ldots, \alpha_{it}, \ldots, \alpha_{iT})$ denote the prior for each of the state vectors α_i. Assuming $p(\alpha_{i1}, \ldots, \alpha_{it}, \ldots, \alpha_{iT})$ to have the form of kT-dimensional multivariate normal prior density (where k is the number of states) the posterior inference about α_i can be carried out using the conventional Bayesian methods for linear models. Note, that the problem of estimating the components of the latent state

vectors is essentially the problem of estimating kT time-varying linear regression coefficients, k for each time period, leading to a kT-dimensional multivariate normal posterior. Despite the relative ease of this approach, obtaining the draws from the posterior distribution of α_i can become difficult in practice due to a large T and possible high correlation between the coefficients (Koop (2006), Sec. 8.3.1) that result into low numerical stability and inefficiency of the algorithm. Alternatively, rewriting the expression for $p(\alpha_i)$ as the product of appropriate conditional densities and applying the Markov property of the state space models will yield the following result

$$p(\alpha_{i1}, \ldots, \alpha_{it}, \ldots, \alpha_{iT}) = p(\alpha_{i1}) \ldots p(\alpha_{it}|\alpha_{i1}, \ldots, \alpha_{it-1}) \ldots p(\alpha_{iT}|\alpha_{i1}, \ldots, \alpha_{iT-1}) \quad (11)$$

$$= p(\alpha_{i1}) \ldots p(\alpha_{it}|\alpha_{it-1}) \ldots p(\alpha_{iT}|\alpha_{iT-1}) \quad (12)$$

which establishes a hierarchical type of construct in both prior and posterior distributions of α_i. The hierarchical structure of the problem allows us to use the Bayesian sequential approach to state vectors estimation where the posterior inference about each component of α_i is obtained conditionally on the posterior value of the previous component in the time sequence. The particular results based on the specification of the measurement (5) and the state equation (6) can be derived according to the Theorem 15.1 in West and Harrison (1989) for general multivariate dynamic linear models as follows. For $t = 2, T$, the prior distribution for α_{it} is implied by the state transition rule as the bivariate normal density

$$(\alpha_{it}|\mathcal{I}_{t-1}) \sim N(d, D) \quad (13)$$

with

$$d = Fa_{t-1} \quad \text{and} \quad D = H_{it}P_{t-1}H_{it}^T + Q$$

where a_{t-1} and P_{t-1} are the posterior mean and variance of α_{it-1}, respectively, and $\mathcal{I}_{t-1}$ denotes the past information available. Updating the prior information with the observed values of y and corresponding allocation variables z at time t gives the following bivariate normal posterior distribution for α_{it}

$$(\alpha_{it}|\mathcal{I}_t) \sim N(a_t, P_t) \tag{14}$$

with $K_{it} = DH_{it}(H_{it}DH_{it}^T + R)^{-1}$ such that

$$a_t = d + K_{it}(u_{it} - H_{it}d) \quad \text{and} \quad P_t = D - K_{it}H_{it}^T D$$

where a_t and P_t are the posterior mean and variance of α_{it}, respectively, and $\mathcal{I}_t$ denotes the current information available.

Step 3: $\sigma_j^2|\Gamma_{-\sigma_j^2}, \{z_i\}_i^N, \{\mu_i\}_i^N, y$

Given the $IG(a_j, b_j)$ prior distribution the posterior density of σ_j^2 is defined as

$$\sigma_j^2 \sim IG\left(\frac{1}{2}\sum_i n_{ij} + a_j, \left[b_j^{-1} + \frac{1}{2}\sum_i\sum_t z_{ijt}(y_{it} - \mu_{jt})^2\right]^{-1}\right) \tag{15}$$

where $n_{ij} = \sum_{t=1}^T z_{ijt}$ denotes the number of time periods the latent process j contributed to the observed realization of the ith spatial unit yield y_i. Note that exactly the same expression for posterior of σ_j^2 in (15) can be obtained from both finite mixture of normals and Gaussian state space estimation procedures since

10

both models are just two different representations of one spatio-temporal process.

Step 4: $\delta_j^2, \gamma_j^2 | \Gamma_{-\sigma_j^2}, \{z_i\}_i^N, \{\mu_i\}_i^N, y$

Assuming the conventional $IG\left(c1_j, c2_j\right)$ and $IG\left(h1_j, h2_j\right)$ priors for δ_j^2 and γ_j^2 parameters, respectively, the corresponding posterior densities are derived as

$$\delta_j^2 \sim IG\left(\frac{T}{2} + c1_j, \left[c2_j^{-1} + \frac{1}{2}\sum_t (\mu_{jt} - \mu_{jt-1} - \eta_{jt-1})^2\right]^{-1}\right) \tag{16}$$

$$\gamma_j^2 \sim IG\left(\frac{T}{2} + h1_j, \left[h2_j^{-1} + \frac{1}{2}\sum_t (\eta_{jt} - \eta_{jt-1})^2\right]^{-1}\right) \tag{17}$$

Note, that the form of prior distributions for δ_j^2 and γ_j^2 defines the degree of smoothness of state variables series and has to be specified by researcher. The choice to favor the higher variation in the stochastic trend will improve the in sample fit of the model, decreasing however the forecast power.

Step 5: $w_i | \Gamma_{-w_{ij}}, \{z_i\}_i^N, \{\mu_i\}_i^N, y$

Given the Dirichlet prior for component probabilities $w_i \sim D(\{\alpha_{ij}\}_{j \in A_i})\mathbb{1}(w_{ii} > w_{ij})$ the posterior draws of w_i for $i = 1, \dots, N$ can be obtained from the following conditional densities

$$w_i \sim D(\{n_{ij} + \alpha_{ij}\}_{j \in A_i})\mathbb{1}(w_{ii} > w_{ij}) \tag{18}$$

Prediction

The general one-step ahead prediction can be computed using the one-step fore-cast result from Theorem 15.1 in West and Harrison (1989) by drawing from the following predictive density of u_{it}

$$(u_{ti}|\mathcal{I}_{t-1}) \sim N(H_{it}d, H_{it}DH_{it}^T + R) \tag{19}$$

Data

The data used for the study are obtained from the National Agricultural Statistical Service (NASS) and are 1970 - 2009 annual Iowa state county level corn yield data in bushels. The $N \times N$ connectivity matrix C was computed using OpenGeoDa software by applying the first order Queen contiguity criterion. The value of the matrix element $C(i,j) = 1$ means that the spatial unit i is the neighbor to the spatial unit j while $C(i,j) = 0$ implies no connectivity between i and j according to the chosen contiguity criterion. Note, that unlike the conventional spatial analysis methods the algorithm proposed in this study requires treating spatial unit i as the neighbor to itself based on the idea of the own latent process contribution. As the result the connectivity matrix we use has the values of its diagonal elements all equal to 1 (see Fig. 2). The Iowa state cartographic boundary files was obtained from U.S. Census Bureau Census 2000 County and County Equivalent Areas Cartographic Boundary Files Database at http://www.census.gov/geo/www/cob/co2000.html.
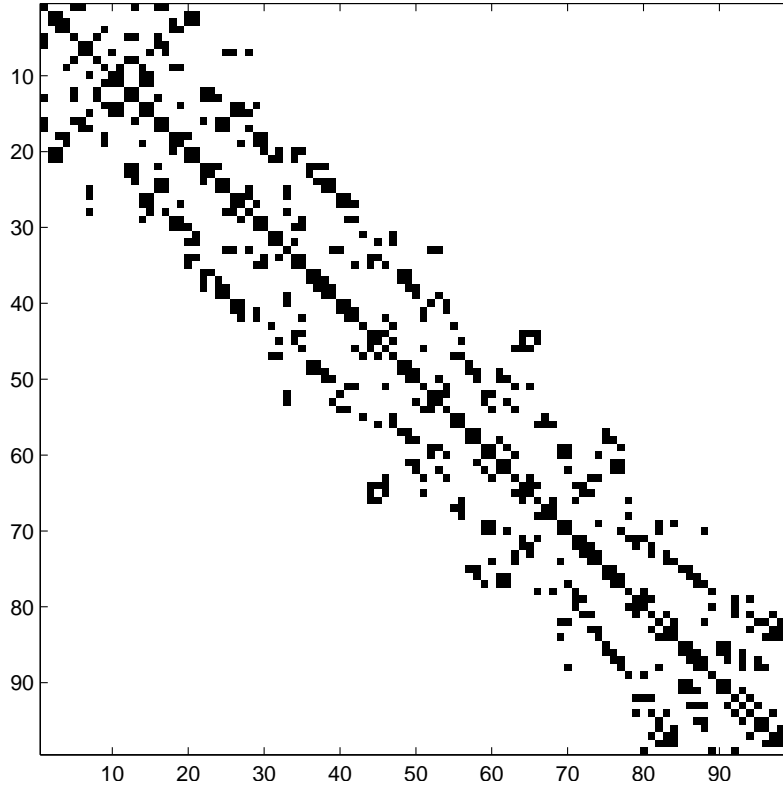
Figure 2: Connectivity matrix of Iowa state counties.

Expected results

Based on the general dynamic linear model equations (2) - (4) we fit two competing models, that are different in the set of spatial weights they use. Model 1 assumes no restrictions on the space of w_{ij} besides the natural nonnegativity and adding up to 1 constraints, thus allowing for spatial correlation between the agricultural yields of any spatial units i and j. In context of the spatial mixture methods discussed in Section 3.2, Model 1 explicitly implies non-normality of the underlying yield distribution y_i if more than one of the spatial weights w_{ij} for $j \in A_i$ is greater than zero. An alternative, Model 2, is a special case of Model 1 that requires $w_{ii} = 1$ for all $i = 1, \ldots, N$. Such a restriction essentially prohibits any spatial correlation by assuming that only own latent process μ_i is going to

13

be an active information contributor to the distribution of yield y_i at any time t thus reducing the Model 1 to a normality case. We have used the first 35 observations covering the period from 1970 to 2004 fit both models using the algorithm described in Section 3.3 while the last 5 years of data for the period from 2005 to 2009 served as the basis for the model comparison based on the out-of-sample predictive power. To assess the predictive abilities of both models we use the squared loss function form of predictive criterion developed by Gelfand and Ghosh (1998) that incorporates both the goodness-of-fit and the penalty for higher predictive variance measures.

References

[1] Gallagher, P. 1987. "U.S. Soybean Yields: Estimation and Forecasting With Nonsymmetric Disturbances." *American Journal of Agricultural Economics*, 69: 796 - 803.

[2] Goodwin, B. and A. Ker. 1998. "Nonparametric Estimation of Crop Yield Distributions: Implications for Rating Group-Risk Crop Insurance Contracts". *American Journal of Agricultural Economics*, 80: 139 - 153.

[3] Goodwin, B.K., M.C. Roberts and K.H. Coble. 2000. "Measurement of Price Risk in Revenue Insurance: Implications of Distributional Assumptions." *Journal of Agricultural and Resource Economics*, 25(1): 195 - 214.

[4] Harri, A.,C. Erdem, K.H. Coble and T.O. Knight. 2009. "Crop Yield Distributions: A Reconciliation of Previous Research and Statistical Tests for Normality". *Review of Agricultural Economics*, 31(1): 163 - 182.

[5] Hennessy, D.A., B.A. Babcock and D.J. Hayes. 1997. "Budgetary and Producer Welfare Effects of Revenue Insurance." *American Journal of Agricultural Economics*, 79: 1024 - 1034.

[6] Just, R. and Q. Weninger. 1999. "Are Crop Yields Normally Distributed?" *American Journal of Agricultural Economics*, 81: 287 - 304.

[7] Ker, A. and K. Coble. 2003. "Modeling Conditional Yield Densities". *American Journal of Agricultural Economics*, 85(2): 291 - 304.

[8] Moss, C.B. and J.S. Shonkwiler. 1993. "Estimating Yield Distributions Using a Stochastic Trend Model and Nonnormal Errors." *American Journal of Agricultural Economics* 75 (5) : 1056 - 62.

[9] Nelson, C.H. and P.V. Preckel. 1989. "The Conditional Beta Distribution as a Stochastic Production Function." *American Journal of Agricultural Economics*, 71: 370 - 378.

[10] Norwood, B., M.C. Roberts and J.L. Lusk. 2004. "Ranking Crop Yield Models Using Out-of-Sample Likelihood Functions". *American Journal of Agricultural Economics*, 86(4) : 1032 - 1043.

[11] Ozaki, V.A., S.K. Ghosh, B.K. Goodwin and Ricardo Shirota. 2008. "Spatio-Temporal Modeling of Agricultural Yield Data with an Application to Pricing Crop Insurance Contracts." *American Journal of Agricultural Economics* 90 (4) : 951 - 961.

[12] Ozaki, V.A., B.K. Goodwin and Ricardo Shirota. 2008. "Parametric and Nonparametric Statistical Modelling of Crop Yield: Implications for Pricing Crop Insurance Contracts." *Applied Economics*, 40 : 1151 - 1164.

[13] Ozaki, V.A. and R.S. Silva. 2009. "Bayesian Ratemaking Procedure of Crop Insurance Contracts with Skewed Distribution." *Journal of Applied Statistics*, 36(4) : 443 - 452.

[14] Racine, J. and A. Ker. 2006. "Rating Crop Insurance Policies with Efficient Nonparametric Estimators that Admit Mixed Data Types". *Journal of Agricultural and Resource Economics*, 31(1): 27 - 39.

[15] Ramirez, O. 1997. "Estimation and Use of a Multivariate Parametric Model for Simulating Heteroscedastic, Correlated, Nonnormal Random Variables: The Case of Corn Belt Corn, Soybean, and Wheat Yields." *American Journal of Agricultural Economics*, 79: 191 - 205.

[16] Ramirez, O. and T. McDonald. 2006. "Ranking Crop Yield Models: A Comment". *American Journal of Agricultural Economics*. 88(4): 1105 - 1110.

[17] Wang, H.H. and H. Zhang. 2003. "On the Possibility of a Private Crop Insurance Market: A Spatial Statistics Approach". *The Journal of Risk and Insurance*, 70(1) : 111 - 124.

[18] West, M. and J. Harrison. 1989. *Bayesian Forecasting and Dynamic Models*. Springer-Verlag, New York Inc, NY.