SPATIO-TEMPORAL MODELING OF AGRICULTURAL YIELD DATA WITH AN APPLICATION TO PRICING CROP INSURANCE CONTRACTS

Vitor A. Ozaki,

Sujit K. Ghosh,

Barry K. Goodwin,

and

Ricardo Shirota*

Abstract:

This article presents the statistical modeling of agricultural yield data based on a set of hierarchical Bayesian models. A hierarchical Bayesian framework allows the joint modeling of the temporal and spatial autocorrelation observed in the yield data. One of the major advantages of this framework is that an estimate of the premium rate is obtained directly from the posterior predictive distribution. This method captures all possible inference uncertainties involved in predicting the insurance premium rates as opposed the more traditional *ad hoc* two-stage methods based on estimation and prediction. A county-average yield data set was analyzed for the State of Paraná, Brazil for the period of 1990 through 2002. The choice of the best model from among the several non-nested models considered was based on a posterior predictive criterion. The methodology used in this article proposes substantial improvements in the statistical methods often applied to the calculation of insurance premium rates. These improvements are especially relevant to situations when the availability of data is limited.

*Ozaki and Shirota are with the University of São Paulo, Brazil. Ghosh and Goodwin are with North Carolina State University. Research support from the North Carolina Agricultural Research Service and CAPES (Brazil) are gratefully acknowledged.

SPATIO-TEMPORAL MODELING OF AGRICULTURAL YIELD DATA WITH AN APPLICATION TO PRICING CROP INSURANCE CONTRACTS

Introduction

Historically, crop insurance in Brazil has been offered by the government at both the federal and state levels. In spite of the government's efforts, the experience with crop insurance in Brazil has generally not been satisfactory. The absence of relatively long data sets and a suitable actuarial method to price crop insurance contracts is one of the main reasons for the poor performance and ultimate failure of this agricultural risk management program. High premium rates inhibited the demand for the insurance by producers and, at the same time, selected only those with higher probability of receiving the indemnity. This is the classic problem of adverse selection.

In recent years, efforts have been made to improve the performance of the programs and to make crop insurance more popular among producers. In December 2003, the federal government of Brazil approved Law No. 10,823, which authorizes the government to subsidize the crop insurance premium, according to the sort of insurance, type of crop and animal species, categories of producers, and production regions. These scheme benefits those engaged in activities considered risk-reducing or technology-enhancing.

Beyond the federal government's efforts, state governments have attempted to stimulate producers' demand for crop insurance. The State of São Paulo (Southeast region), through a pilot project initiated in 2003, subsidized the premium paid by the producers. In the State of Rio Grande do Sul (South region), the state government began regulating the crop insurance state system through the Law No. 11,352 in 1999. The program is subsidized and operates with three types of insurance which vary according to the producer and the total amount of subsidy.

This article concentrates on statistical methods of pricing an alternative crop insurance contract based on county yields. This type of insurance is widely available in the United States (named group risk plan), India, Sweden and Canada (Miranda, Skees

and Hazel, 1999) and, currently, is offered in Brazil in the South Region. It is important to point out that the methodology proposed in this paper can also be applied to pricing others forms of insurance contracts, such as those based on individual yields, as long as there are enough data.

Statistical Modeling Framework

A wide variety of statistical methods are often adopted in the estimation of crop insurance rates and a number of issues relating to the modeling of crop yields are pertinent to these methods. For example, one often must address issues related to the fact that yields tend to have substantial trends over time and tend to be significantly correlated over space due to the systemic nature of weather. One subtlety often overlooked in crop insurance pricing models pertain to the fact that a degree of uncertainty also applies to the estimated parameters of any model. In this analysis, we adopt a Bayesian inferential framework that accounts for such sources of uncertainty while estimating the appropriate premium rate.

Over many years, the statistical issues underlying agricultural yields have been a controversial point in the crop insurance literature. Several statistical approaches have been considered, including parametric yield models, semiparametric methods (Ker and Coble, 2003), nonparametric models (Goodwin and Ker, 1998; Turvey and Zhao, 1999) and empirical Bayes nonparametric approaches (Ker and Goodwin, 2000).

Within the parametric modeling approach, some researchers have concluded that crop yields tend to follow a Normal distribution (Just and Weninger, 1999). However, a large number of other researchers including Day (1965), Taylor (1990), Ramirez (1997), and Ramirez et al. (2003) have found evidence against Normality.

Other suggestions included the use of a Beta distribution (Nelson and Preckel, 1989), Inverse Hyperbolic Sine Transformations (Moss and Shonkwiler, 1993), and Gamma distributions (Gallagher, 1987). Sherrick et al. (2004) used several parametric distributions including the Normal, Lognormal, Beta, Weibull and Logistic distributions to model individual yield data.

Of course, the characteristics of crop yields may be idiosyncratic and may vary by location, crop, and production practice. In most empirical work, the only information known at time *t* is the time index and previously realized yields. Thus, in these analyses, the conditional density is based only on the temporal generating process of the data.

In this study the temporal aspect of the data generating process is addressed, but we also give attention to the spatial dimension of the data generating process. In particular, we explicitly recognize the fact that the events that underlie yield realizations (e.g., weather, disease, and pest damages) tend to affect large areas at any single time. Thus, adjacent regions may experience substantial spatial correlations of yields over time. Taking this fact into account, space and time were combined in order to construct a spatio-temporal model.

In other words, we simultaneously model the time trend, the temporal and spatial autocorrelation in contrast to two-stage methods[1] making the premium rate calculation less *ad hoc*, in the sense that rates are derived directly in the model after the simulation through Markov Chain Monte Carlo algorithm (MCMC). Moreover, when calculating rates we are able to capture its estimation uncertainty through the standard errors.

The fact that our data set is not large (in time dimension) creates additional difficulties regarding the forecast or prediction[2] of crop yields in future years. In the construction of crop insurance contracts, it is typically the case that the terms and parameters of the contract must be available one to two years prior to the insurance cycle[3].

In our case, the last observation recorded was for the year 2002. Assuming that there is a two year lag between the receipt of historical yield data and the deadline required for filing new contract terms[4]. In light of this objective, we model the structure of the yield

[1] A two-stage method first detrend the time series and them treat the detrended data (known as "normalized yields") as "observed" data to estimate the premium rate. Thus, this method fails to capture the uncertainty of the premium rate estimate.
[2] In this article, forecast and prediction and density and distribution will be used interchangeably.
[3] Administrative issues relating to the operation of any program require substantial lead time in providing the parameters of the contract offering.
[4] Such a two year lag is inherent in all U.S. crop insurance programs.

mean and assume that the precision of our models is conditionally constant throughout the analysis[5].

Under this approach, the conditional mean μ_{it} was considered as identical to E(y_{it}), where i represents the space variable index and t the temporal index. Thus, y_{it} is the agricultural yield in county i and in time t, where i = 1, 2, ... , S and t = 1, 2, ... , T. The objective is to model the stochastic mean component, so that μ_{it} reflects the temporal effects, spatial variation and the spatio-temporal relationships.

Modeling the structure underlying the mean yield realization by adopting hierarchical models is intuitive and facilitates the visualization of each component in the analysis instead of modeling such structure directly through the y_{it}[6]. However, one limitation of the building correlation structure by hierarchical models is that all of the pairwise correlations would be positive.

In situations where relatively little is known about the hyper-parameters, diffuse prior distributions can be adopted. Nevertheless, we must be careful to recognize that improper priors may yield improper posterior distributions[7]. In a practical sense, this problem can be prevented by considering proper prior distributions that assure that the Gibbs sampling process will be well-behaved, where ignorance can be represented as values for the precision parameter close to zero[8] (Gelfand and Smith, 1990).

Bayesian Mixture of Gaussian distributions

[5] Modeling the mean component rather than the precision in forecasting problems results in more effective results (Gelfand et al., 1998).

[6] For this alternative version, Anselin (1988) shows several spatial and spatial-temporal models, such as, SUR (seemingly unrelated regression), where the Beta coefficients are allowed to vary in one of the two dimensions and the error term is correlated in the other dimension. In those models the dependence structure is modeled through the error term ε_{it}, where $y_{it} = x_{it}\beta_{it} + \varepsilon_{it}$.

[7] In this context, Hobert and Casella (1996), estimated the parameters of a hierarchical linear mixed model using the Gibbs sampler and warned about using a non-informative prior distribution that can lead us to an improper posterior distribution.

[8] However, even in this case Gelman (2004) raises some computational and numerical issues.

Extending the work of Ker and Goodwin (2000), we modeled μ_{it} as coming from two subpopulations or groups, a catastrophic and a non-catastrophic group[9]. Under this approach, we fit a hierarchical Bayesian model based on a mixture of two Gaussian distributions, where the density of the first (catastrophic) group lives in the inferior tail of the second group.

Because catastrophic events are, by definition, much less frequent and the observed yield in such years is inferior relative to yields in regular years, one can expect a smaller mass in the first group and that such concentration lies in the left tail of the non-catastrophic distribution. If we had information about such catastrophic events for each region and each year, we could use it as an indicator variable within a regression model. However, in most cases, such information is not observable and thus must be considered to be represented by latent variables.

The general mixture model can be written as:

$$f(y \mid \theta_1,...,\theta_j,\gamma_1,...,\gamma_j) = \sum_{j=1}^{J} \gamma_j f(y \mid \theta_j), \quad (1)$$

where θ_j is the parameter vector, J is the number of components, and for $j = 1, 2, \ldots, J$ $\gamma_j \geq 0$ is a weighting parameter representing the ratio of the population attributed to the component j, such that $\sum_j \gamma_j = 1$. If the distribution $f(y \mid \theta_j, \gamma_j)$ is represented by a Gaussian distribution, then we have $\theta_j = (\mu_j, \sigma_j^2)$. Thus, eq. (1) can be written as

$$f(y \mid \theta_1,...,\theta_J,\gamma_1,...,\gamma_J) = \sum_{j=1}^{J} \gamma_j N(y \mid \theta_j) \quad (2)$$

The previous model can be specified in an alternative manner by introducing an unobserved (latent) indicator variable that identifies the component from which the observation is drawn. This indicator variable I receives values equal j when y is drawn from the jth component. Equivalently, thus the mixture model in (1) can be represented as:

[9] A catastrophic event can be defined by an adverse climatic event that occurs in a determined period of time (such as drought, hail, etc.). Consequently, if such an adverse event occurs, the agricultural yield will be drawn from the catastrophic group. Alternatively, yields are considered to be drawn from the non-catastrophic group when normal weather events are realized. In this manner, one can think of yield realizations as being drawn from a finite mixture of two distributions.

$$y \mid I,\ \theta \sim \mathrm{f}(y \mid \theta_I)$$
$$I \mid \underline{\gamma} \sim DCat\,(\gamma), \tag{3}$$

where *DCat ( )* is the Categorical distribution such that P[$I = j$] = γ_j , $j = 1,\ldots,J$. We assume that we do not know from which component each observation is drawn. In this case, if we consider that the parameters θ and γ are independent, then the prior distribution can be considered as the product of the two distributions. As we assign a Categorical prior distribution for *I*, the conjugate prior for γ will be the Dirichlet distribution[10] with hyper-parameter α :

$$f(\gamma) = \frac{\Gamma(\sum_j \alpha_j)}{\Pi_j \Gamma(\alpha_j)} \Pi_j \gamma_j^{\alpha_j - 1}, \tag{4}$$

where $0 < \gamma_j < 1$ and $\sum_j \gamma_j = 1$, $\alpha_j > 0$, $j = 1, \ldots J$.

Gelman et al. (2003) suggest that the ratio between the two variances should be considered as fixed or, alternatively, one should assign a proper prior distribution. In this analysis, we assign an Inverse Gamma distribution (*a*, *b*) to assure that the posterior distribution is proper (assuming $J = 2$), and adopt Normal priors for the μ_j terms and a Dirichlet distribution for the γ_j terms.

Temporal modeling

Considering the temporal component as an integral part of μ_{it}, we will model it initially by assuming that $\Psi_t = \beta + u_t$, where Ψ_t is a constant mean for all regions plus an error term, where $u_t \overset{\text{iid}}{\sim} \mathrm{N}(0,\ \sigma^2)$. This model will be expanded incorporating time as a covariate in the analysis. In this case, time may be represented by a polynomial in *t* according to $\Psi_t = \sum_{l=1}^{p} \beta_l t^l + u_t$. For this type of deterministic trend model, the variable *t*

[10] For more details see Spiegehalter et al. (2003)

was centered in order to improve the MCMC speed of convergence, such that $t^* = (t - (N+1)*0.5)$.

As an initial data exploration technique, we use empirical plots to evaluate the type of trend that might be present in the data. This evaluation indicated that a quadratic trend was sufficient to capture deterministic trend effects. Beyond the deterministic trend models, stochastic trend models and its interactions were analyzed.

The stochastic trend component was modeled as a first-order autoregressive model AR (1), where, $\Psi_t = \rho\ \Psi_{t-1} + u_t$, where $-1 \leq \rho \leq 1$[11]. Note that this specification includes the random walk model as a special case.

Assumptions regarding the specification of the model must be made. First, the correlation parameter ρ in the stochastic trend models varies according to the region. Second, an exchangeable[12] Normal prior was assigned to the parameter ρ and Normal and inverse Gamma hyper-distributions for the mean and variance parameters, respectively[13].

Initially, a first-order polynomial function in t added to the stochastic component was considered. A subsequent model emerges naturally summing the second order term, resulting in $\Psi_t = \beta_0 + \beta_1 t^* + \beta_2 t^{*2} + \eta_t$, and $\eta_t = \rho\eta_{t-1} + u_t$. Similarly, the correlation coefficient was reparameterized as in the previous case and Normal prior distributions were assigned for β_0 and β_1 and $\beta_,$, with a prior precision parameter $\tau \rightarrow 0$.

---

[11] In light of the small sample size, a more sophisticated temporal model was not possible. For example, Ker and Goodwin (2000 p. 465) proposed an IMA(1,1) process, represented by $y_t = y_{t-1} + \theta_0 + \theta e_{t-1} + e_t$. The number of observations used in their article was small as well, though larger than in our case. Thus modeling an IMA(1,1) process can become a troublesome with regard to the stability and convergence of the parameters. In this manner, because we can express an MA(1) process as an AR(∞) process, they modeled the temporal process as a AR(4), such that, $y_t = y_{t-1} + \beta_0 + \beta_1(y_{t-1} - y_{t-2}) + \beta_2(y_{t-2} - y_{t-3}) + \beta_3(y_{t-3} - y_{t-4}) + \beta_4(y_{t-4} - y_{t-5}) + e_t$.

[12] The random variables $X_1, \ldots, X_n$ are exchangeable if any permutation of any subset of them of size k ($k \leq n$) has the same distribution. Note that a generalization of iid random variables is exchangeable random variables (Casella and Berger, 2002). This idea was first introduced by de Finetti (1972). One immediate consequence of exchangeability is that all marginal distributions must be the same (Migon and Gamerman, 1999).

[13] We can also reparameterize the parameter ρ so that a prior distribution could be assign to ρ, such that $\rho = 2\eta - 1, 0 \leq \eta \leq 1$. Naturally, the Beta ($c$, d) distribution emerges as a prior for the parameter η where $c = \xi\psi$ and $d = (1 - \xi)\psi$, $0 < \xi < 1$, $\psi > 0$ and hyper-prior distributions for ξ and ψ.

In light of random effects models, β's will be exchangeable. Such a result is convenient and it is reasonable to assume that the parameters may be different from one another, although they arise from the same population distribution. Thus, the preceding model (named "exchangeable model") takes the form $\beta \sim N_3(b, \Sigma)$, where the hyper-prior distributions for the vector b and the matrix Σ will be, respectively, $b \sim N_3(\mu_b, \Sigma_b)$, where $\mu_b = 0$ and Σ_b is the diagonal covariance matrix with diagonal elements that approach ∞ and $\Sigma \sim W(R, k)$, where Σ is a p x p symmetric positive definite matrix, with a density proportional to:

$$|R|^{k/2} |\Sigma|^{(k-p-1)/2} \exp [-1/2(\mathrm{Tr}(R \Sigma))], \quad (5)$$

where $k \geq p$ yields the Wishart distribution (Anderson, 1982).

Spatial modeling

In the traditional literature of Bayesian spatial models, a variable Φ_i denoting the spatial aspects can be represented initially in terms of a set of covariates placed in a vector Φ_i representing a given characteristic of a certain area, contributing a component $g(\Phi_i)$, where g would be a specific parametric function. In the absence of such covariates, random effects are introduced in order to capture unobserved features among different regions (Gelfand et al., 1998).

This approach is particularly appealing because of the nature of agriculture yield data. Random effects by region need to be separated. Using the hierarchical structure of our Bayesian model one can separate these effects. Under the exchangeability assumption heterogeneity variables and clustering variables can then be estimated. The former through its prior distribution and the latter assuming a special form of spatial prior distribution.

This treatment addressed the problem of spatial dependence between counties. Identification of the parameters in the likelihood function in this case is verified in the

hierarchical model by assuming a conditional autoregressive[14] (CAR) prior distribution for ξ_i and exchangeable Normal priors for v_i. Moreover, when using the temporal term as a covariate, the autocorrelation and the time trend are considered, without any data transformation.

In previous studies (Besag, 1974; Clayton and Kaldor, 1987; and, Cressie and Chan, 1989), the non-structured variable is assumed to follows a Normal distribution, such that $v_i \sim N(\mu_v, \sigma_v^2)$. In addition, we assume that the spatially structure variable ξ_i conditional on ξ_j ($j \neq i$), is proportional to:

$$\xi_i \mid \xi_j \sim \exp\{-1/2\sigma_\xi^2(\varphi_i\xi_i - \sum_{j\neq i}\omega_{ij}\xi_j)^2\}$$

where $\varphi_i \geq 0$ is a "sample size" associated with region i and $\omega_{ij} \geq 0$ is the weight reflecting the influence of ξ_j on the conditional mean of ξ_i. We let $\omega_{ij} = 1$ if j is neighbor of i and 0 otherwise and set φ_i equal the number of neighbors of i. Thus, the conditional distribution $\xi_i \mid \xi_j$ simplifies to $\xi_i \sim N(\bar{\xi}_i, \sigma_\xi^2 / n_i)$, where $\bar{\xi}_i$ is the average of the ξ_j's, in which j indexes the neighboring sites of i. The variance parameters σ_v^2 and σ_ξ^2 are assigned an inverse Gamma prior distribution.

Bernardinelli et al. (1995a) pointed out that the choice of the dispersion parameter must be made with caution. Carrying out a simulation of a Poisson model applied to disease mapping they verified that the heterogeneity parameter has standard deviation approximately equal to 0.7 times the standard deviation of the clustering parameter, var $(v_i) \approx 0.7$ var (ξ_i).

Thomas et al. (2002) suggested that a restriction must be imposed on the random effects parameters such that those effects sum to zero. In other words, an intercept parameter must be included in the model receiving an improper (uniform) prior distribution.

[14] The reader must not confuse the term "autoregressive" commonly used in the time series analysis. In spatial statistics or econometrics, autoregressive refers to the mean of the variable in the neighbor regions.

Gelfand et al. (1998) noted that, if both parameters were placed in the model, then one must allow E(v_i) = 0. In the same fashion, if both parameters v_i and ξ_i were included in the model and one attributed a non-informative prior for v_i, then either $v_i = 0$ or $\sum v_i = 0$. Moreover, as they pointed out, if ξ_i and v_i are included in the model, the prior distribution will have greater weight in the posterior density. If one allows v_i to be centered around zero with a small variance, then the component ξ_i will have greater weight in the term Φ_i.

Due to convergence issues in the MCMC algorithm, Gefland et al. (1998) suggested that one should choose to include either the spatially non-structured variable or the structured variable, but not both. Because the objective of their article was to obtain predicted values, they concluded that the model including ξ_i yielded better results.

Spatio-temporal modeling

One of first articles related to the spatial-temporal analysis using a log-linear Poisson model in disease mapping was provided by Bernardinelli et al. (1995b). In general, the model can be represented by: (intercept + area) + (time + area*time). To capture the dependence between v_i and ξ_i, or in other words, the intercept and trend, they assumed that v_i follows a univariate Normal distribution and that ξ_i came from a conditional Normal distribution.

Based on this research, Dreassi (2003) modeled the relative risk for each period and city in Italy, incorporating an ordinal covariate that allows one to determine in which time lag the disease, in this particular case, lung cancer is affected by socio-economic factors.

Another approach to modeling spatio-temporal effects was proposed by Waller et al. (1997). In this model, instead of capturing the spatio-temporal variation in a multiplicative form, they considered a nested model, where the spatial effect and the heterogeneity effect were allowed to vary in time. The general model considered was:

$$\mu_{ist} = x_{is}^T \beta + z_i^T \omega + \xi_{it} + v_{it}, \qquad (6)$$

where $x_{is}^T\beta$ is the covariate representing the effect for each sub-group *s*, $z_i^T\omega$ represents the regional covariate, ξ_{it} is the spatial effect for the *i*th region in year t and v_{it} is the random effect for the *i*th region in year t.

Because of the conditional interchangeability associated with time, the resulting prior distribution assigned to the heterogeneity can be represented by $v_i^{(t)} \overset{\text{iid}}{\sim} N(\mu_v^{(t)}, \sigma_v^{2(t)})$. For the spatial effect $\xi_i^{(t)}$ in the *i*th region in year *t,* an intrinsic CAR prior distribution was adopted (Waller et al., 1997). Thus, $\xi_i^{(t)} \sim N(\overline{\xi}_i^{(t)}, \sigma_\xi^{2(t)} / n_i)$, where $\overline{\xi}_i^{(t)}$ is the average of the *j*th contiguous areas of *i*. The precision parameters ($\sigma_v^{2(t)}$ and $\sigma_\xi^{2(t)}$) follow an inverse Gamma.

Some restrictions also must be imposed in spatio-temporal models in order to ensure identification. The inclusion of the former effect makes unnecessary the addition of v_i and ξ_i. Moreover, the model is incapable of identifying $\xi_i^{(t)}$ and Ψ_t if both are included in the model and a non-informative prior is assigned to Ψ_t given the time t. If both $v_i^{(t)}$ and $\xi_i^{(t)}$ are included in the model, then $\mu_v^{(t)} = 0$. In this study, the spatial effects were nested within the temporal process, such that the parameters of the deterministic trend (β's) are modeled using the CAR prior[15]. Thus we have the following general expression for the mean component: $\mu_{it} = \beta_0^{(i)} + \beta_1^{(i)} t^* + \beta_2^{(i)} t^{*2} + u_{it}$.

Model selection criteria

As we have demonstrated in the preceding review, several models emerge as potential candidates for our particular problem. A basic question is thus how to select the best model, taking into account one of the objectives of this work - prediction of agricultural yields. Traditional criteria of model selection, such as the Bayes factor, are not applicable in cases like ours where non-informative or conditional autoregressive (CAR) prior distributions are used. Carlin and Louis (2000, pg. 220), have shown that the use of

[15] Intuitively, one can think of the trend parameters as being correlated across space, given time.

improper priors results in improper conditional predictive distributions, limiting the use of Bayes factor as a model selection criterion in these cases.

The application of the classical approach to model selection is also difficult in these cases. Penalized likelihood criteria based on asymptotic efficiency requires the determination of the dimension of the model or the number of the parameters. In hierarchical models with random effects (such as the ones used in this paper), the dimension is difficult to characterize. Criteria based on cross validation are also difficult to implement when more sophisticated models are considered, due to the inclusion of heterogeneity and clustering variables defined only by the prior (Waller, 1998).

In this article, a criteria based on predictive densities was considered (Laud and Ibrahim, 1995). In this context, Gelfand and Ghosh (1998) formalized a predictive criteria using a general form of loss function. The objective is to minimize the posterior predictive loss. The posterior predictive distribution is given by:

$$f(y_{new} \mid y_{obs}) = \int f(y_{new} \mid M)\, p(M \mid y_{obs})\, dM \tag{7}$$

where M represents the set of all parameters in a given model and y_{new} is the replicate of the vector of observed data y_{obs}.

The criteria of model selection is based on a discrepancy function $d(y_{new}, y_{obs})$, and the objective is to choose the model that minimizes the expectation of the discrepancy function, conditional on y_{obs} and M_m, where M_m represents all the parameters in the model m. If we consider Gaussian models, the discrepancy function is given by $d(y_{new}, y_{obs}) = (y_{new} - y_{obs})^T (y_{new} - y_{obs})$:

$$D_{M_m} = E[(y_{new} - y_{obs})^T (y_{new} - y_{obs}) \mid y_{obs}, M_m]$$

$$D_{M_m} = \sum_n E[(y_{n,obs} - y_{n,new})^2 \mid y_{obs}, M_m]. \tag{8}$$

Equation 8 can be factored into two additive terms G_{M_m} and P_{M_m}, where the first term $G_{M_m} = \sum_n [y_{n,obs} - E(y_{n,new} \mid y_{obs})]^2$ represents the sum of squared errors, which is a measure of goodness-of-fit, and the second term $P_{M_m} = \sum_n \mathrm{var}(y_{n,new} \mid y_{obs})$ is a penalty term. In models that are over- or under-fit, the predicted variance tends to be large and

thus P_{M_m} is large. The penalty is considered in the analysis without regard to the dimension of the model.

In this work, a slightly different version of the model selection criterion will be utilized. Instead of using the quadratic predicted error, the mean squared predictive error will be considered relative to the number of regions used in the analysis. Note that the inclusion of a common denominator to all models does not affect the criterion.

Empirical Analysis

Data Description

The agricultural yield data used in this study were provided by the IBGE (Statistical and Geography Brazilian Institute) and correspond to the period of 1990 trough 2002 for corn in the state of the Paraná, located in the southern region of Brazil. The state is made up of 399 counties. Annual yield observations for all 13 years are only available in 290 counties. Consequently, we carry out the analysis with only those counties with the largest number of observations. The five largest counties in terms of average yields are Castro (6142 kg/ha), Ponta Grossa (5629 kg/ha), Marilândia do Sul (5488 kg/ha), Tibagi (5346 kg/ha) and Catanduvas (4923 kg/ha).

Empirical Application

We begin our analysis by choosing the model that minimizes the posterior predictive loss. Among the several models that were considered as candidates (25 in all), we only present results for the 10 best models (that resulted in minimum D_m, according to the criteria described above). Results for the model selection criteria are presented in Table 1.

Note that all of the models chosen by the ten best values of the predictive error criterion include the temporal component and the stochastic trend. This clearly demonstrates the importance of the stochastic trend in the analysis. The optimal model, or in other words, the model that minimizes the quadratic predictive error, includes both the stochastic and deterministic components. In addition, allows the intercept to vary from

one county to another. Further, this model was expanded to include spatial dependence in the slope parameters.

The difference between models 1 and 2 lies in the prior distributions assigned to the β parameters. The superscript C indicates that a conditional autoregressive prior was assigned to the parameter. Otherwise, β receives a Normal prior. Comparing models 4 and 9, one can note that the presence of heterogeneity variable results in smaller D_m as compared to the inclusion of the clustering effect. Comparing models 6 and 7, adding the spatially structured latent variable (clustering) indexed by t results in larger value of D_m as compared to the model that holds the clustering variable fixed in time. If we include the deterministic term, the model with a clustering effect that varies in time becomes slightly superior to the model considering the same effect constant in time ($D_8 < D_9$). The results in Table 1 also demonstrate that the quadratic deterministic trend model and mixture of Normal models were not included in the top ten best model specifications due to unsatisfactory values of D_m.

Basically, model 1 can be expressed according to the hierarchical structure:

$$y_{i,t} \sim N(\mu_{i,t}, \tau)$$
$$\mu_{i,t} = \rho_i y_{i,t-1} + \beta_{1_i} + \beta_{2_i}^C t^* + u_{i,t}$$

Prior distributions:

$\rho_i \sim N(\mu^\rho, \tau^\rho)$ - exchangeable prior distributions

$\beta_{2_i}^C = \xi_i + c$

$\xi_i \sim CAR(\bar{\xi}_i, \sigma_\xi^2 / n_i)$

β_{1_i} and c Normal priors with mean zero and low precision.

$\sigma = 1/(\tau)^{0.5}$ an Inverse Gamma distribution

Hyperprior distributions:

μ^ρ Normal hyperprior and $\sigma^\rho = 1/(\tau^\rho)^{0.5}$ an Inverse Gamma distribution

We run three chains to check the mixing of the Markov sequence and also check for all the parameters the graphical diagnostics of convergence. Results showed that all parameters achieved good convergence and mixing.

One of the main advantages of Bayesian analysis is that one can incorporate uncertainty when estimating the parameter value. Table 2 shows the expected value of the parameter, its standard deviation and the percentiles 5%, median and 95%. For these counties, the average standard deviation is for β_1, β_2 and ρ equal to 582, 3.9 and 0.11.

Because of the limited space, we will show only descriptive statistics of the 290 counties. Thus, the maximum predicted values of β_1, β_2 and ρ are respectively 2410, 46.85 and 0.83. The minimum values are 550, 46.73 and 0.30 and the average, 1174, 46.79 and 0.61. The average standard deviation is 430, 3.95 and 0.13. Due to the small number of observations, we do not correct for conditional heteroskedasticity. Instead we assume that series are conditionally homoskedastic[16].

In Table 3 we show the predicted values of yields and its respective standard deviation and percentiles 5, 50 and 95% for the chosen counties. The variance of the predicted value tends to increase as the time lag increases.

Rating Crop Insurance Contracts

The insurance premium rate (PR) represents a proportion (or percentage) of total liability. In the simple case where a proportion λ ($0 \leq \lambda \leq 1$) of the expected crop yield y^e is used to form the basis of insurance, the premium rate is given by (Goodwin and Ker, 1998):

$$\text{Premium Rate } (PR) = \frac{F_Y(\lambda y^e) E_Y[\lambda y^e - (Y \mid y < \lambda y^e)]}{\lambda y^e}, \tag{9}$$

where E is the expectation operator and F is the cumulative distribution function of yields. At this point we show how rates can be derived directly from our Bayesian hierarchical model. A slightly different derivation of the premium rate is convenient for our purposes. If we reparameterize y, such that, $y^* = y / \lambda y^e$, then equation (9) becomes:

$$PR = P(y^* < 1) E_{y^*}[1 - (y^* \mid y^* < 1)] \tag{10}$$

[16] If series were relatively longer, a procedure that could be used to verify heteroskedasticity would be assign to the precision parameter tau indexes i and t, or in other words, make the parameter vary in time and space and, later on, monitor such parameter to verify the variation in the precision and correct it, when necessary.

Note that the support of the random variable Y remains the same in this transformation. If we consider $w = 1 - y^*$, then equation (10) can be rewritten such that:

$$PR = P(w > 0)[1 - E_w(1 - w|w > 0)]$$

$$PR = P(w > 0)E_w[w|w > 0)] \quad (11)$$

After some simplification, the premium rate equation reduces to:

$$PR = \int_0^1 wf(w)dw \quad (12)$$

Equation 12 can be written as $PR = E[w\mathrm{I}(0 < w < 1)]$. Because of the change of variable, the support also changed such that w lies now in between 0 and 1. In our model, we can easily implement computationally equation (11) using the predicted yields. This expression represents the mean of w, or more specifically in the Bayesian jargon the "posterior mean" of w, which is the PR calculated for each county and for each level of coverage. Moreover, the Bayesian approach allows one to derive standard error estimates, but in our context these estimates are called Monte Carlo standard error of the mean[17]. In table 4, we show some rates and their MC errors. Antonina county is an illustrative example of the variability in rates. The standard deviation and consequently the MC error are much higher in this county comparing to the others. In this case, the uncertainty on rates will be much higher.

A natural advantage of having a viable measure of the uncertainty associated with an individual premium rate estimate can be found in the common insurance practice known as "loading"[18]. These adjustments are typically *ad hoc* and are based upon the actuary's confidence in the estimate. The standard errors of the premium rate estimates provide a natural metric to guide such loading practices. In particular, higher load adjustments can be applied to those rates which reflect a greater degree of uncertainty. The standard errors account for all of the uncertainty that encompasses the model, including the estimation of yield trend effects and spatial correlation factors.

[17] For further details see Spiegehalter et al. (2003).

[18] Loading refers to markups or additive factors that are often applied to premium rates to account for uncertainty and is also commonly used to build reserves, to cover administrative and operating costs and to ensure a positive profit for the insurer.

In figure 1 below, we illustrate aggregate premium rates by coverage levels and by regions in the state of Paraná. The state was divided into 10 great regions[19]: 1 - Occidental Centre (21), 2 - Oriental Centre (10), 3 - Centre-South (11), 4 - Metropolitan of Curitiba (28), 5 - Northwest (33), 6 - Central North (64), 7 - Pioneer North (43), 8 - West (36), 9 - Southeast (18), 10 - Southwest (26).

Conclusions

We have discussed a statistical method of pricing a crop insurance contract based upon hierarchical Bayesian models. We point out that this methodology can also be applied to contracts based on individual yields, as long as there are enough data. Conventional methods of pricing this type of individual contract using aggregate yield data, such as, county averages, are not recommended, because they do not reflect accurately the risk structure of an individual producer, thus increasing the problem of the adverse selection.

The methodology developed in this article was used to forecast corn yields for selected counties in the State of Paraná using data covering 1990 through 2002. Using the posterior predictive criteria of Gelfand and Ghosh (1998), we chose from among several models appropriate for this forecasting and insurance pricing problem. The optimal model was used in the calculation of premium rates for insurance coverage based on regional yield indexes.

Our analysis considers not only the temporal aspect of yield movements but also the spatial correlation that exists between counties. The resulting spatial-temporal model is thus more flexible and less ad hoc compared to other potential specifications that have been considered in the literature.

In other words, temporal and spatial effects can be incorporated into the model trough prior distributions. Moreover, premium rates can be derived directly considered as another parameter to be estimated by the model. One advantage is that standard errors of premium rates will also be calculated and used to load insurance premiums.

[19] Number of counties analized in brackets.

Future research will evaluate methods of pricing insurance contracts for individual yields using the methods developed in this analysis.

References

AKAIKE, H. Information theory and an extension of the maximum likelihood principle, in Proceedings of the International Symposium on Information Theory, eds PETROV, B. N.; CZAKI, F. Budapest: Academia Kiado, p. 267-281.

AKERLOF, G. A. The market for "lemons": quality uncertainty and the market mechanism. Quarterly Journal of Economics, v. 84, n. 3, p. 488-500, ago 1970.

ANSELIN, L. Spatial Econometrics: Methods and Models. Kluwer Academic Publishers, 1988, 284 p.

BERNARDINELLI, L.; CLAYTON, D.; MONTOMOLI, C. Bayesian estimates of disease maps: how important are priors? Statistics in Medicine, v. 14, p. 2411-2431, 1995a.

BERNARDINELLI, L.; CLAYTON, D.; PASCUTTO, C.; MONTOMOLI, C.; GHISLAND, M.; SONGINI, M. Bayesian analysis of space-time variation in disease risk. Statistics in Medicine, v. 14, p. 2433-2443, 1995b.

BESAG, J. Spatial Interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society, Ser. B, v. 36, 192-236, 1974.

BESAG, J.; YORK, J. MOLLIÉ, A. Bayesian image restoration, with applications in spatial statistics. Annals of the Institute of Statistical Mathematics, v. 43, p. 1-59, 1991.

BOTTS, R. R.; BOLES, J. N. Use of Normal-curve theory in crop insurance ratemaking. Journal of Farm Economics, v. 40, n. 2, p. 733-740, 1958.

CARLIN, B. P.; LOUIS, T. A. Bayes and empirical bayes methods for data analysis. Chapman & Hall/CRC, 2000, 440 p.

CASELLA, G.; BERGER, R.L. **Statistical inference**. Belmont: Duxbury Press, 2002. 660p.

CLAYTON, D. G.; KALDOR, J. Empirical bayes estimates of age-standardized relative risks for use in disease mapping. Biometrics, v. 43, p. 671-681, 1987.

CRESSIE, N. A. C.; CHAN, N. H. Spatial modeling of regional variables. Journal of the American Statistical Association, v. 84, p.393-401, 1989.

DAY, R. H. Probability distributions of field crop yields. Journal of Farm Economics, v. 47, n. 3, p. 713-741, 1965.

DE FINETTI, B. Probability, induction and statistics. New York: Wiley. 1972

DREASSI, E. Space-time analysis of the relationship between material deprivation and mortality for lung cancer. Envirometrics, v. 14, p. 511-521, 2003.

GALLAGHER, P. U.S. Soybean yields: estimation and forecasting with nonsymmetric disturbances. American Journal of Agricultural Economics, v. 69, n. 4, p. 796-803, 1987.

GELFAND, A. E.; GHOSH, S. K. Model choice: a minimum posterior predictive loss approach. Biometrika, v. 85, p. 1-11, 1998.

GELFAND, A. E.; GHOSH, S. K.; KNIGHT, J. R.; SIRMANS, C. F. Spatio-temporal modeling of residential sales data. Journal of Business & Economic Statistics, v. 16, p. 312-321, 1998.

GELFAND, A. E.; SMITH, A. F. M. Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association, v. 85, p.398-409, 1990.

GELMAN, A.; CARLIN, J. B.; STERN, H. S.; RUBIN, D. B. Bayesian data analysis. Chapman & Hall/CRC, 2003, 668 p.

GOODWIN, B. K.; KER, A. P. Nonparametric estimation of crop yield distributions: Implications for rating Group-Risk crop insurance contracts. American Journal of Agricultural Economics, v. 80, p. 139-153, 1998.

GOODWIN, B. K.; MAHUL, O. Risk modeling concepts relating to the design and rating of agricultural insurance contracts. World Bank, 2004.

HOBERT, J. P.; CASELLA, G. The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. Journal of the American Statistical Association, v. 91, p.1461-1473, 1996.

JUST , R. E.; WENINGER, Q. Are crop yields Normally distributed? American Journal of Agricultural Economics, v. 81, p. 287-304, 1999.

KER, A. P.; GOODWIN, B. K. Nonparametric estimation of crop insurance rates revisited. American Journal of Agricultural Economics, v. 83, p. 463-478, 2000.

KER, A. P; COBLE, K. Modeling conditional yield densities. American Journal of Agricultural Economics, v. 85, p. 291-304, 2003.

LAUD, P. W.; IBRAHIM, J. G. Predictive model seletion. Journal of the Royal Statistical Society, Ser. B, v. 57, p. 247-262, 1995.

MIGON, H.S.; GAMERMAN, D. Statistical inference: an integrated approach. New York: Oxford University Press, 1999. 262p.

MIRANDA, M; SKEES, J.; HAZELL, P. Innovations in agricultural and natural disaster insurance for developing countries. Working paper, Dep. of Agr., Envir. and Development. Econ., The Ohio State University, 1999.

MOSS, C. B.; SHONKWILER, J. S. Estimating yield distributions with a stochastic trend and nonNormal errors. American Journal of Agricultural Economics, v. 75, n. 4, p. 1056-1062, 1993.

NELSON, C. H.; PRECKEL, P. V. The conditional Beta distribution as a stochastic production function. American Journal of Agricultural Economics, v. 71, n. 2, p. 370-378, 1989.

RAMÍREZ, O. A. 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, v. 79, p. 191-205, 1997.

RAMÍREZ, O. A.; MISRA, S.; FIELD, J. Crop-yield distributions revisited. American Journal of Agricultural Economics, v. 85, p. 108-120, 2003.

ROTHSCHILD, M.; STIGLITZ, J. Equilibrium in competitive insurance markets: an essay on the economics of imperfect information. Quarterly Journal of Economics, v. 90, n. 4, p. 629-649, nov. 1976.

SKEES, J. R.; REED, R. R. Rate making for farm-level crop insurance: implications for adverse selection. American Journal of Agricultural Economics, v. 68, p. 653-659, 986.

TAYLOR, C. R. Two practical procedures for estimating multivariate nonNormal probability density functions. American Journal of Agricultural Economics, v. 72, p. 210-217, 1990.

SHERRICK, B. J.; ZANINI, F. C.; SCHNITKEY, G. D.; IRWIN, S. H. Crop insurance valuation under alternative yield distributions. American Journal of Agricultural Economics, v. 86, p. 406-419, 2004.

SPIEGEHALTER, D.; THOMAS, A.; BEST, N.; LUNN, D. Winbugs user manual. Medical Research Council Biostatistics Unit: Cambridge, 2003.

SCHWARZ, G. Estimating the dimension of a model. The Annals of Statistics, v. 6, 461-464, 1978.

THOMAS, A.; BEST, N.; ARNOLD, R. SPIEGEHALTER, D. GeoBugs user manual. Medical Research Council Biostatistics Unit: Cambridge, 2002.

TURVEY, C.;ZHAO, C. Parametric and nonparametric crop yield distributions and their effects on all-risk crop insurance premiums. Working paper, Dep. of Agricultural Economics and Business, University of Guelf, Jan, 1999.

WALLER, L. A.; CARLIN, B. P.; XIA, H.; GELFAND, A. E. Hierarchical spatio-temporal mapping of disease rates. Journal of the American Statistical Association, v. 92, p.607-617, 1997.

Table 1. Model Selection Criteria

M	D_m	Model for u_{it}

1	667800	$\rho_i y_{i,t-1} + \beta_{1_i} + \beta_{2_i}^C t*$
2	673200	$\rho_i y_{i,t-1} + \beta_{1_i}^C + \beta_{2_i}^C t*$
3	700100	R-W
4	728500	$\rho_i y_{i,t-1} + \beta_{1_i} + \beta_{2_i} t* + v_i$
5	736800	AR(1)
6	737900	$\rho_i y_{i,t-1} + \xi_i$
7	739900	$\rho_i y_{i,t-1} + \xi_i^t$
8	751400	$\rho_i y_{i,t-1} + \beta_{1_i} + \beta_{2_i} t* + \xi_i^t$
9	751700	$\rho_i y_{i,t-1} + \beta_{1_i} + \beta_{2_i} t* + \xi_i$
10	761300	Exchangeable model

Table 2. Predicted parameter values, standard deviation and percentiles 5, 50 and 95%, of selected counties.

County	parameter	predicted value	standard deviation	0.05	0.95
Castro	β_1	1366	683.6	201.3	2475
	β_2	46.83	3.938	40.33	53.29
	ρ	0.8073	0.1143	0.6236	1.002
Catanduvas	β_1	1545	515.7	687.5	2397
	β_2	46.78	3.95	40.29	53.34
	ρ	0.7147	0.1032	0.5447	0.8851
Marilândia do Sul	β_1	1446	592.6	461	2426
	β_2	46.78	3.937	40.28	53.28
	ρ	0.7703	0.1092	0.5904	0.9502
Ponta Grossa	β_1	1511	612.3	490.1	2523
	β_2	46.82	3.94	40.34	53.3
	ρ	0.7553	0.11	0.5749	0.9413
Rolândia	β_1	2109	526.4	1260	2993
	β_2	46.79	3.937	40.31	53.28
	ρ	0.5579	0.1082	0.374	0.733
Tibagi	β_1	1380	563	450.5	2306
	β_2	46.82	3.941	40.32	53.32
	ρ	0.7751	0.1062	0.6021	0.9526

Table 3. Predicted yield values, standard deviation and percentiles 5, 50 and 95%, of selected counties, in 2003 and 2004.

county	year	predicted yield	standard deviation	0.05	median	0.95
Catanduvas	2003	5968	758	4716	5972	7195
	2004	5833	903	4350	5813	7316
Castro	2003	8301	791	6990	8303	9591
	2004	8455	1114	6647	8443	10280
Marilândia do Sul	2003	7499	786	6208	7492	8784
	2004	7624	1074	5883	7614	9405
Ponta Grossa	2003	6553	760	5296	6550	7793
	2004	6638	1008	5021	6628	8338
Rolândia	2003	7336	777	6068	7342	8615
	2004	7461	1079	5745	7433	9280
Tibagi	2003	7730	793	6419	7733	9019
	2004	7779	1094	6035	7753	9613

Table 4. Premium rates (%) for selected counties by level of coverage, standard deviation and MC error estimates

County	Level of Coverage (%)	Premium Rate	Stand. Dev.	MC error	County	Level of Coverage (%)	Premium Rate	Stand. Dev.	MC error
Antonina	70	7.295	0.139	0.00154	Ponta Grossa	70	0.006	0.002	1.77E-05
	75	9.478	0.153	0.00171		75	0.039	0.005	5.64E-05
	80	11.810	0.165	0.00186		80	0.178	0.012	1.45E-04
	85	14.230	0.177	0.00200		85	0.553	0.022	3.00E-04
	90	16.690	0.186	0.00212		90	1.326	0.035	5.33E-04
Castro	70	0.014	0.003	2.72E-05	Rolândia	70	0.001	0.001	6.34E-06
	75	0.084	0.008	9.51E-05		75	0.013	0.003	2.68E-05
	80	0.318	0.016	2.63E-04		80	0.063	0.007	6.64E-05
	85	0.897	0.029	5.17E-04		85	0.220	0.014	1.37E-04
	90	2.041	0.044	8.34E-04		90	0.593	0.023	2.41E-04
Catanduvas	70	0.017	0.003	3.25E-05	Tibagi	70	0.016	0.003	2.89E-05
	75	0.096	0.008	9.26E-05		75	0.096	0.008	1.00E-04
	80	0.342	0.017	2.07E-04		80	0.356	0.017	2.41E-04
	85	0.905	0.029	3.74E-04		85	0.980	0.030	4.51E-04
	90	1.920	0.044	5.76E-04		90	2.120	0.045	7.08E-04
Marilândia do Sul	70	0.013	0.003	2.73E-05					
	75	0.076	0.007	7.44E-05					
	80	0.314	0.016	1.78E-04					
	85	0.899	0.028	3.78E-04					
	90	2.007	0.044	6.83E-04					

FIGURE 1. Premium rates (%) by level of coverage and aggregated by regions in the state of Paraná