



AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

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

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

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

aesearch@umn.edu

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

**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 focuses on the modeling of agricultural yield data using hierarchical Bayesian models. In recovering the generating process of these data, we consider the temporal, spatial and spatio-temporal relationships pertinent to the prediction and pricing of insurance contracts based on regional crop yields. 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 the posterior predictive criterion. The methodology used in this article proposes improvements in the statistical and actuarial methods often applied to the calculation of insurance premium rates. These improvements are especially relevant to situations of limited data. These conditions are frequently encountered, especially at the individual level.

Selected Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Providence, Rhode Island, July 24-27, 2005. Copyright 2005 by the authors. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on such copies.

*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 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, which characterized historical efforts at developing crop insurance in Brazil.

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. These premium subsidies will be specified according to the sort of insurance, type of crop and animal species, categories of producers, and production regions. A priority of the new insurance is to provide coverage to those engaged in activities considered to be 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, through a pilot project initiated in 2003, subsidized the premium paid by the producers by 50% for five crops: banana, orange, grape, corn and beans in 219 cities of the State. In 2004, the subsidy program was expanded to cover 14 more crops in 534 of the 645 cities of the State. These crops included cotton, peanuts, irrigated rice, cassava, soybeans, sorghum, wheat, pineapples, plums, kaki, guava, passion fruit, peaches, and cabbage. In the State of Rio Grande do Sul, 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 and actuarial methods with the objective of pricing an alternative crop insurance contract based on an index of regional yields. This type of insurance is widely available in the United States, India, Sweden and Canada (Miranda, Skees and Hazel, 1999) and, currently, is offered in Brazil in the southern state of Rio Grande Do Sul. It's important to point out that such methods addressed 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 to do the analysis.

In the analysis that follows, a number of alternative parametric, statistical models were applied to the data set with the objective of modeling the stochastic generating process of yield data and, in particular, properly recognizing the temporal, spatial and spatio-temporal dynamics underlying crop yields. To select among a large number of potential candidate models, a minimum mean square prediction error criteria was used.

This article is arranged according to the following outline. In the next section, the history of agricultural insurance in Brazil will be described briefly. Particular attention will be given to the institutional, legal and operational aspects of crop insurance at both the federal and state levels. In the next section, we review the literature that has addressed statistical aspects of agricultural yield data. To be precise, statistical modeling of yield data will be considered. We also describe temporal, spatial, spatio-temporal Bayesian modeling methods. As we discuss in detail, alternative distributional assumptions and techniques, including mixtures of normal distributions and the normal distribution, are considered. We also consider methods appropriate for choosing among alternative candidate models. The next section of the paper builds upon this discussion by applying the empirical methods to an analysis of Brazilian yield data for maize. In particular, we describe our optimal model specification and pursue an analysis of the problem of pricing a crop insurance contract using corn yield data from the State of Paraná, Brazil between 1990 through 2002. It should be noted that, although our application is to a case of regional yield insurance, our methods are entirely applicable to other types of contracts. We conclude the paper with a brief summary and concluding comments.

A BRIEF HISTORY OF CROP INSURANCE IN BRAZIL

Agricultural insurance was introduced in Brazil in 1938. This early form of federal crop insurance was specific peril—being directed toward hail damages to cotton. Hail insurance was later expanded to cover additional crops, including grapes and other horticulture crops. The early performance of the program was poor, with loss ratios (i.e., the ratio of indemnities paid out to premiums collected) above 3.8 being observed in the early years.

In January 1954, the federal government of Brazil established the norms and regulations for crop insurance in Brazil through Law No. 2,168. This legislative action established the Agrarian Insurance Stability Fund to guarantee insurance market stability, allow the gradual adjustment of premium rates, cover catastrophic risks, and to provide a number of other initiatives to improve crop insurance. The Institute of Reinsurance of Brazil (IRB), through its Technical Board, was responsible for the administration of the Fund's resources. The law created the National Company of Crop Insurance (CNSA) with the objective of gradually developing the operational aspects of agricultural insurance. All responsibilities not assumed by the insurance companies or the CNSA were reinsured in the IRB. If these companies did not find reinsurance within the country they could, through the IRB, seek reinsurance in international markets.

The company functioned between 1954 and 1966, but was forced to interrupt its activities in November 1966 because of a legal decree (Law No. 73) which closed down the CNSA. One of the main causes of the failure of the company was the extreme centralization of the program's administration in the city of Rio de Janeiro. Even with branch offices spread in many regions in the central and south of the country, the central administration was thought to monopolize the decision-making process. Thus, important programmatic aspects that should have been taken into account were ignored. For example, the construction of regional plans that considered the features and peculiarities of local environments should have been considered. Likewise, the area to be insured and selection of the producers should have been subject to tighter constraints. These are

crucial program design issues that one must consider when dealing with crop insurance in a large country like Brazil with so many different climates and types of soil.

In contrast to such a reasoned approach, the federal company applied generically-designed plans throughout the entire country. Moreover, due to the fact that CNSA was not allowed to operate in other types of more profitable insurance lines, their exposure to risk was concentrated on agricultural risks in a small volume of business. Further, the operational resources available to the company were modest and the company ran large deficits.

After the dissolution of the CNSA (through Law No. 73/66), the government did not provide any risk management or protection mechanism to the agricultural sector until 1973. In December 1973, the Farming Activity Guarantee Program, called PROAGRO was created by Law No. 5,969. The federal government, because of a need to offer risk protection to producers due to the increasing supply of rural credit, implemented the PROAGRO program to protect the capacity of the financial system in case of large-scale defaults on loan obligations by producers. From its beginnings through 1993, the program accumulated large deficits (\$1.6 billion). Many problems were raised during this period; including delays and non-payment of indemnities, technical and operational deficiencies and concerns regarding the presence of fraud and abuse in the program.

The results in the first five years of the program were disastrous. The loss ratios for the years between 1975 and 1979 were 84.8, 8.2, 10.1, 9.3 and 16.5, respectively.¹ The period corresponding to the beginning of the PROAGRO until August of 1991, was characterized by unsatisfactory results. The average loss ratio in the period of 1980 through 1991 was 2.4. In 1991, a number of important modifications to the program were announced, though these modifications did not halt the growing deficit of the fund, which reached \$264.6 million by May of 1994. Consequently, in August of 1994, there were additional operational modifications to the program, although some deficiencies still persisted. Program shortcomings included an absence of monitoring, multiple risks covered by the program, and a substantial delay in the payment of indemnities.

In October of 1995, the Ministry of Agriculture and Supply, in partnership with the Foundation of Technological and Scientific Enterprises (FINATEC) and the University of

¹ The loss ratio is the total amount of indemnity paid out divided by the total amount of premium collected.

Brasilia, implemented the “Project to Reduce Climatic Risk in Agriculture.” The objective of this project was to develop a methodology to reduce losses in agricultural production and to induce the adoption of techniques that the producer could use to manage climatic risk, specifically rainfall risks. The project was concluded in August of 1996 and the National Monetary Council in accepting the recommendations, redefined the objectives of the PROAGRO program.

New legislation in 1996 stipulated that producers that adhered to agricultural production recommendations would be charged differentiated insurance premium rates. Moreover, the project determined the proper seeding period for each city and for the following crops: rice, cotton, soybeans, corn and dried beans for the entire country and for wheat in the central and southern regions. As result, there was a significant improvement in the performance of the program. The loss ratio over the period of August of 1991 to 1997 was 0.94. At the state level, some initiatives were introduced though these were initially restricted to only a few crops.

Several private insurance companies currently offer crop insurance in Brazil. Although the amount of business is still small, some pilot projects have been implemented throughout the country. For example, a private company implemented an insurance plan based on county yield (that is, an area-wide insurance plan) in the State of Rio Grande do Sul. In this type of insurance, the producer will be indemnified only on the basis of a large area’s yield experience. In particular, the level of protection and indemnities are based upon the difference between the area-wide guaranteed yield and the observed county yield. Indemnities are paid only if the observed yield is lower than the county-level guarantee, regardless of an individual farmer’s experience. Other private insurance companies are offering contracts covering yields at the individual farm level.

Other types of crop insurance can also be found in the country. A mutual, cooperative form of insurance is the oldest type of insurance in Brazil. Through a formal contract, individuals get together and agree to divide damages or losses that individual producers might experience due to certain unexpected events. Instead of paying premiums, the insured growers contribute according to a quota necessary to cover administrative costs.

The responsibility for risk is shared by everyone. Some examples of mutual insurance can be found in the southern regions of Brazil.²

STATISTICAL MODELING FRAMEWORK

A fundamental parameter of any insurance contract is the premium rate. An actuarially fair premium rate is a rate that is set such that premiums collected are equal to expected indemnities. An inaccurate premium rate results in distortions to the insurance pool and thus may result in losses as agents adversely select against the insurance provider. In particular, low risk agents may be overcharged and high risk agents may be undercharged. This will distort participation in favor of the higher risks and thus premiums will not be sufficient to cover indemnity payments. This condition of adverse selection has been well documented for a number of insurance plans. The eventual failure of an insurance program as a result of such selection is often called the “death spiral of adverse selection.” Optimally, an insurance provider would prefer to calculate individual premium rates for each farmer on the basis of that farmer’s risks and expected yields. However, individual data are rare at best and thus crop insurance plans are often based upon more aggregate data—such as data at the county level. Such index-based crop insurance plans were developed to overcome the problem of short or nonexistent individual crop yield series.

Another important aspect of insurance contract design pertains to the actuarial procedures used in the calculation of insurance premium rates. In particular, the derivation of such rates generally requires a statistical analysis of crop yields. 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 yield models pertains to the fact that a degree of uncertainty also applies to the parameters of

² This includes the Cooperativa Agropecuária Batavo, the Cooperativa Agrária Mista Entre Rios, and Associação dos Fumicultores do Brasil (Afubra).

any model used to describe the uncertainty of yields. For example, it is common to detrend yields using standard regression models and then to use the detrended yields to measure yield uncertainty. However, a certain degree of uncertainty is also inherent in the models used to detrend yields. In this analysis, we adopt a Bayesian inferential framework that accounts for all 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. Thus, it is unlikely that any single parametric approach will be universally supported across different applications.

As we have pointed out, a related problem pertains to the limited number of yield observations typically available for empirical models. This is true even when aggregated data are considered. This limitation typically precludes the use of individual farm-level data for the purposes of modeling yields and rating insurance contracts. The choice of a statistical model that adequately reflects the conditional density of yields is an important consideration in the actuarial calculation of an accurate premium rate. In doing this, one must try to recover the probability generating process of the yield data. Agricultural yields follow a spatio-temporal process, in the sense that if we take the average in a region conditional on the underlying temporal process, one can recover the conditional

density yield $f(y | \Omega_t)$ at a certain moment in time and point in space, where Ω_t is the minimum σ -algebra generated by the information known at moment t (Ker and Goodwin, 2000).

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. Our work addresses this temporal aspect of the data generating process, 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 correlations of yields over time. Thus, our models combine the two aspects of space and time in order to construct a spatio-temporal model of crop yields.

The fact that our data set is not large in the time dimension creates additional difficulties regarding the forecast or prediction of crop yields in future years.³ 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. This reflects the fundamental fact that an insurance provider will not want to offer coverage after the insurance buyers already have information about their yields. In addition, administrative issues relating to the operation of any program require substantial lead time in providing the parameters of the contract offering. In our case, the last observation recorded was for the year 2002. We will assume that there is a two year lag between the receipt of historical yield data and the deadline required for filing new contract terms. Such a two year lag is inherent in all U.S. crop insurance programs. In this context, we must attempt to choose the best statistical model to predict yields for the following 2 years. In light of this objective, we model the structure of the yield mean and assume that the precision of our models is conditionally constant throughout the analysis. Gelfand et al. (1998), point out that modeling the mean component rather than the precision in forecasting models results in more effective results.

³ In this article, forecast and prediction and density and distribution will be used interchangeably.

Under this approach, we consider the mean μ_{it} as being 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 $i = 1, 2, \dots, S$ and $t = 1, 2, \dots, T$. The objective of this portion of the analysis is to model the stochastic mean component, so that μ_{it} reflects the covariates, the temporal effects, spatial variation and the spatio-temporal relationships relevant to agricultural yields.

In some applications, statistical models may be comprised of a large number of parameters. This is especially true in analyses of data that have been pooled over time and cross-sections. In such cases, a natural way of modeling the parameters is through hierarchical models. Under such an approach, the dependence structure between the parameters can be represented by the joint probability distribution. Consequently, we can define a prior distribution for these parameters assuming that they can be considered as a sample from a common population distribution.

Hierarchical models are usually specified in several stages, thus suggesting a conventional notation. If the model has k stages, the joint distribution of the observed variable y and parameters θ 's, can be written in a multiplicative form, such as:

$$f(y | \theta_1) f(\theta_1 | \theta_2) f(\theta_2 | \theta_3) \dots f(\theta_{k-1} | \theta_k) f(\theta_k).$$

We consider hierarchical models to be more natural for incorporating the correlation structure. Thus, in our model, the first hierarchical stage assumes that y_{it} is conditionally independent, given μ_{it} . In other words, any parameters added to our representation of μ_{it} will be random. If μ_{it} includes a random effect indexed by t then, marginally, y_{it} will reflect the temporal dependence in a given year. Similarly if we include an effect indexed by i , marginally, y_{it} will reflect the spatial dependence within certain region. In such a case, a spatial effect can be introduced into the model, thus allowing for spatial dependence among the observed variables y_{it} . 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} .⁴

⁴ 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

In many applications, the observed variable is modeled conditional on a given number of parameters that receive a prior probability distribution, which in turn receive other parameters, called hyper-parameters. One can assign probability distributions to these parameters (hyper-prior distributions). Priors can then be chosen to reflect prior knowledge of these hyper-parameters. In situations where relatively little is known about the hyper-parameters, diffuse prior distributions can be adopted. However, we must be careful to recognize that improper priors may yield improper posterior distributions.⁵

Consider, for example, the following prior distribution for the parameter $\theta \sim N(\mu, \sigma^2)$. If we assign a hyper-prior distribution for σ^2 , such as, $f(\sigma) = I/(\sigma^2)^a$ where $a = 0, 0.5, 1$, it may happen that the joint posterior distribution is improper, although the final results based on numerical output seem reasonable and the analyst may not realize the problem. In such a case, an analyst will be making inferences about a non-existent posterior distribution. In a practical sense, as shown in Gelfand and Smith (1990), 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.⁶

Initially, extending the work of Ker and Goodwin (2000), we modeled μ_{it} as coming from two subpopulations or groups, a catastrophic group and a non-catastrophic group. 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.

Under this approach, we fit 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

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 $y_{it} = x_{it}\beta_{it} + \varepsilon_{it}$.

⁵ 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.

⁶ However, even in this case Gelman (2004) raises some computational and numerical issues.

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:

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

where θ_j is the parameter vector, J is the number of components, such that $j = 1, 2, \dots, J$ and $\gamma_j \geq 0$ is a weighting parameter representing the ratio of the population attributed to the component j , and $\sum_j \gamma_j = 1$. If the distribution $f(y | \theta_j, \gamma_j)$ is represented by a

Gaussian distribution, then we have $\theta_j = (\mu_j, \sigma_j^2)$. Thus, eq. (1) can be written as

$$f(y | \theta_1, \dots, \theta_j, \gamma_1, \dots, \gamma_j) = \sum_{j=1}^J \gamma_j N(y | \theta_j) \quad (2)$$

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 j th component. Equivalently, thus the mixture model in (1) can be represented as:

$$\begin{aligned} y | I, \theta &\sim f(y | \theta_I) \\ I | \underline{\gamma} &\sim DCat(\underline{\gamma}), \end{aligned} \quad (3)$$

Where $DCat(\cdot)$ is the categorical distribution and $P[I = j] = \gamma_j, j = 1, \dots, J$. 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 with hyperparameter α :

$$f(\gamma) = \frac{\Gamma(\sum_j \alpha_j)}{\prod_j \Gamma(\alpha_j)} \prod_j q_j^{\alpha_j - 1}, \quad (3)$$

where $0 < q_j < 1$ and $\sum_j q_j = 1$, $\alpha_j > 0$, $j = 1, \dots, J$.

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_t , 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 $u_t \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$. This model, though initially lacking in realism, provides a convenient benchmark which will be expanded in a fashion that allows subsequent models to incorporate time as a covariate in the analysis. In this deterministic trend model, 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, we center the variable t in order to improve the speed of convergence of our Markov Chain Monte Carlo (MCMC) algorithm. Thus, we have $t^* = (t - (N+1)*0.5)$. We consider $p = 1, 2$ in the model estimation and use the normal distribution to form prior distributions for the intercept and trend parameters of the deterministic trend models.

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 in the yield data. Beyond the deterministic trend models, one can also analyze in a complementary fashion stochastic trend models and the interactions between stochastic and deterministic models. In this paper, we modeled the stochastic trend component as a first-order auto-regressive model

AR(1), where, $\Psi_t = \rho u_{t-1} + u_t$, where $-1 \leq \rho \leq 1$.⁷ Note that this specification includes a standard random walk model as a special case. We also adopt two assumptions regarding the exact specification of the model. First, the correlation parameter ρ in the stochastic trend models is allowed to vary according to the region. Second, an exchangeable normal prior was assigned to the parameter ρ with normal and inverse gamma hyper-distributions for the mean and variance parameters, respectively.⁸

The interaction between the deterministic and stochastic trend was analyzed initially by considering a first-order polynomial function in t added to the stochastic component. This implies a subsequent model which emerges if we sum the second order term, which yields $\Psi_t = \rho y_{t-1} + \beta_0 + \beta_1 t^* + \beta_2 t^{*2} + u_t$. In a similar way, the correlation coefficient was reparameterized as in the previous case and normal prior distributions were assigned for β_0 and β_1 and β_2 , with a prior precision parameter $\tau \rightarrow 0$.

If we consider a random effects model, then all of the β parameters 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.

One can then consider the preceding model as an exchangeable model that takes the form $\beta \sim N_3(\mathbf{b}, \Sigma)$, where the hyper-prior distributions for the vector \mathbf{b} and the matrix Σ will be, respectively, $\mathbf{b} \sim N_3(\mu_b, \Sigma_b)$, where $\mu_b = 0$ and Σ_b is the diagonal covariance matrix with diagonal elements that approach ∞ and $\Sigma \sim W(\mathbf{R}, k)$, where Σ is a $p \times p$ symmetric positive definite matrix, with a density proportional to:

⁷ 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 $y_t = y_{t-1} + \theta_0 + \theta_1 e_{t-1} + e_t$. 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, $y_t = y_{t-1} + \beta_0 + \beta_1(y_{t-1} - y_{t-2}) + \beta_2(y_{t-2} - y_{t-3}) + \beta_3(y_{t-3} - y_{t-4}) + \beta_4(y_{t-4} - y_{t-5}) + e_t$.

⁸ 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 .

$$|\mathbf{R}|^{k/2} |\Sigma|^{(k-p-1)/2} \exp [-1/2(\text{Tr}(\mathbf{R} \Sigma))], \quad (4)$$

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

SPATIAL MODELING

In the traditional literature of spatial models, a variable Φ_i denoting the spatial aspects of the data 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 covariates, random effects are introduced in order to capture heterogeneity among different regions. Gelfand et al. (1998), in the absence of covariates, used random effects as surrogate to the covariates in order to capture the effects of heterogeneity in a context where a hedonic price model was used to predict the future selling price of houses.

In their article, they identified a variable v_i that reflected such characteristics as, for example, quality of the construction, income in the subdivision and socio-economic variables, such as, race and education. In addition to unstructured heterogeneity, a latent variable was introduced to catch the spatial effect ξ_i that represented the geographic nature of each subdivision and the importance of each area in relation to the selling price of the houses. Thus, the spatial variable can be represented as $\Phi_i = \xi_i + v_i$, where v_i is a spatially non-structured latent variable (representing heterogeneity) and ξ_i is a spatially structured latent variable (representing clustering). Identification of the parameters in the likelihood function in this case is verified in the hierarchical model by assuming a conditional auto-regressive prior distribution (CAR) for ξ and exchangeable priors for v_i .

Based in the work of Besag (1974), Clayton and Kaldor (1987) used the concept of spatial dependence applied to the problem of disease mapping. Their application was to modeling cancer rates in Scotland. The spatial correlation was modeled on the basis of the geographic proximity of a particular region in relation to other adjacent regions. Cressie and Chan (1989) studied the sudden infant death syndrome with a data set collected during the period of 1974 through 1984 in counties of North Carolina using

space models. They noted that, in a manner analogous to the time series case where one may try to show how the actual observations are influenced by its past values, in spatial processes one may try to verify how a particular value is influenced by its "neighboring" values.

In these prior studies, the non-structured variable is assumed to follow 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 , where $j \neq i$, can be modeled such that $\xi_i \sim N(\bar{\xi}_i, \sigma_\xi^2 / n_i)$, where $\bar{\xi}_i$ is the average of the ξ_j 's and j indexes the neighboring sites of i . The variance parameters σ_v^2 and σ_ξ^2 are assigned an inverse gamma prior distribution. One can note that these terms determine a spatial process in accordance with the terms defined by Besag et al. (1991). Bernardinelli et. al. (1995a) pointed out that the choice of the dispersion parameter must be made with caution. They carried out a simulation study of a Poisson model applied to a model of disease mapping and verified that the heterogeneity parameter has standard deviation approximately equal to 0.7 times the standard deviation of the clustering parameter, $\text{var}(v_i) \approx 0.7 \text{var}(\xi_i)$. 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.

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, Gelfand 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 the pioneering articles related to the spatial-temporal analysis using a log-linear Poisson model in the disease mapping was provided by Bernardinelli et al. (1995b). In this article, they represented the spatial effect, which can be interpreted as the rate of variation of a certain disease in a given area by a random effect variable v_i . The temporal term is captured through a trend coefficient and the interaction between the space and temporal effect reflected by the spatially correlated covariate ξ_i . 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 arose from 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:

$$\mu_{ist} = x_{is}^T \beta + z_i^T \omega + \xi_{it} + v_{it}, \quad (5)$$

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 .

Using the principle of parsimony, simpler models were chosen among the various models considered in the article. Because of the conditional interchangeability associated with time, the resulting prior distribution assigned to the heterogeneity can be represented by $v_i^{(t)} \stackrel{\text{iid}}{\sim} N(\mu_v^{(t)}, \sigma_v^{2(t)})$. For the spatial effect $\xi_i^{(t)}$ in the i th region in year t , Waller et al. (1997) adopted an intrinsic CAR prior distribution. Thus, $\xi_i^{(t)} \sim N(\bar{\xi}_i^{(t)}, \sigma_\xi^{2(t)} / n_i)$, where

$\bar{\xi}_i^{(t)}$ is the average of the j th contiguous areas of i . An inverse Gamma was used to form hyper-priors for $\sigma_v^{2(t)}$ and $\sigma_\xi^{2(t)}$. 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_i if both are included in the model and a non-informative prior is assigned to Ψ_i given the time t . If both $v_i^{(t)}$ and $\xi_i^{(t)}$ are included in the model, then one must let $\mu_v^{(t)} = 0$.

We also allow the spatial effects to be nested within the temporal process, such that the parameters of the deterministic trend (β 's) are modeled using the CAR prior. Intuitively, one can think of the trend parameters as being correlated across space, given time. Thus we have the following general expression for the mean component $\mu_{it} = \beta_0^{(i)} + \beta_1^{(i)}t + \beta_2^{(i)}t^2 + u_{it}$. As was described in the previous subsection, we can incorporate the stochastic term in the general expression and reparameterize the correlation term. We also reparameterized the trend parameters by recentering the trend.

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 auto-regressive (CAR) prior distributions are used. Carlin and Louis (2000, pg. 220), have shown that the use of improper priors results in improper conditional predictive distributions, limiting the use of Bayes' factor as a model selection criterion in these cases.

In the simplest case, when both models have the same parameterization and the hypotheses are simple, one can see that Bayes' factor is equivalent to the likelihood ratio between the two models. The application of the classical approach to model selection is also difficult in these cases. Penalized likelihood criteria based on asymptotic efficiency

require 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. Further, in more sophisticated models the dimension of the model increases with the sample size, thus invalidating the use of popular model selection criteria. Examples of such conventional criteria include the Akaike Information Criteria (AIC) (Akaike, 1973), which in terms of change from model 1 (M_1) to model 2 (M_2) is given by $-2\log(\sup_{M_1}f(x/\theta))(\sup_{M_2}f(x/\theta))^{-1} - 2(p_2 - p_1)$, where p_1 and p_2 are the number of parameters, the Bayesian Information Criteria (BIC) (Schwarz, 1978), which is equal to $-2\log(\sup_{M_1}f(x/\theta))(\sup_{M_2}f(x/\theta))^{-1} - (p_2 - p_1)(\log n)$ and the Deviance Information Criteria (DIC), given by $(E_{\theta|x}(D) - D(E_{\theta|x}(\theta)))$, where the first term is the expectation of the deviance and the second is the deviance estimated at the expectation of the posterior distribution. 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, we select our model specification by adopting a criteria based on predictive densities. As Laud and Ibrahim (1995) pointed out, these criteria are easy to interpret since they are not based on asymptotic analysis and they allow for the incorporation of prior distributions. Working in the predictive space, the penalty appears without the necessity of asymptotic definitions. Intuitively, one can think that good models must result in predictions close to what is observed in identical experiments.

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:

$$f(x_{new} | x_{obs}) = \int f(x_{new} | M) p(M | x_{obs}) dM \quad (6)$$

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

The criteria of model selection is based on a discrepancy function $d(x_{new}, x_{obs})$, and the objective is to choose the model that minimizes the expectation of the discrepancy function, conditional on x_{obs} and M_i , where M_i represents all the parameters in the model

i. If we consider Gaussian models, the discrepancy function is given by $d(x_{new}, x_{obs}) = (x_{new} - x_{obs})^T(x_{new} - x_{obs})$:

$$D_{M_i} = E[(x_{new} - x_{obs})^T(x_{new} - x_{obs}) | x_{obs}, M_i]$$

$$D_{M_i} = \sum_n E[(x_{n,obs} - x_{n,new})^2 | x_{obs}, M_i]. \quad (7)$$

Gelfand and Ghosh (1998) demonstrate that D_{M_i} can be factored into two additive terms G_{M_i} and P_{M_i} , where the first term $G_{M_i} = \sum_n [x_{n,obs} - E(x_{n,new} | x_{obs})]^2$ represents the sum of squared errors, which is a measure of goodness-of-fit, and the second term $P_{M_i} = \sum_n \text{var}(x_{n,new} | x_{obs})$ is a penalty term. In models that are over- or under-fit, the predicted variance tends to be large and thus P_{M_i} 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 through 2002 for corn in the state of the Paraná, located in the southern region of Brazil. The state of Paraná is the largest producer of corn in the country, with a total amount produced in 2002 equal to 9,797,816 tons, a little bit more than 27% of all Brazilian production. Corn yields in Paraná are generally the fourth largest in Brazil (3,987 Kilograms per hectare – kg/ha in 2002).

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). The evolution of the corn yields in the state of the Paraná between the years of 1990 and 2002 can be seen in Figures 1 and 2.

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, the optimal model allows the intercept to vary from one county to another. Further, this model includes 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 results in a smaller D_m as compared to the inclusion of the clustering effect. In a comparison of models 6 and 7, the addition of the spatially structured latent variable (clustering) which varies in time results in a 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. This is because they resulted in unsatisfactory values of D_m .

According to Table 1, given the fact that the best model was the one including the temporal and stochastic component, we will present a detailed description of this model in the discussion that follows. Initially, to give a better visualization and understanding, model 1 will be written graphically, in Figure 3.

In Figure 3, nodes of the Directed Acyclic Graph (DAG) are stochastic variables, rectangles are constants, arrows with simple and parallel straight lines represent, respectively, parameters of the distributions and logical links between the variables. The index i and t in the rectangles denote, respectively, the variable of space and time. The parameter β_2 has logical links between c and ξ (*zeta*), so that we can redefine it, in order to recentralize it. The first variable receives a Normal prior distribution and the second one a conditional auto-regressive prior.

The parameters of the variable *zeta*, named *neigh*, *wei*, and *num* are, respectively, the j -th adjacent counties in relation to a central county i , the weights assigned to each neighboring county, such that, adjacent counties receive weight equal 1 and 0, otherwise, and the sum of the adjacent counties to a central county. The variance parameter of the *zeta* variable receives an Inverse Gamma prior 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. Figure 4 shows the convergence of β_1 , β_2 and ρ in selected counties.

Next, in Figure 5, we show the decomposition of μ_{it} , according to model 1, in its temporal component, deterministic and stochastic, for the counties of Castro, Ponta Grossa, Marilândia do Sul, Tibagi, Catanduvas and Rolândia. One can see that in the five first counties, the stochastic term has larger weight in the composition of u_{it} and the residual are around zero for the six counties.

One of the main advantages of the Bayesian analysis is that one can incorporate the uncertainty when estimating the parameter value. Taking this fact into account, 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.

Figure 6 shows that the number zero is in the tail part of the posterior distributions of the parameters β_1 , β_2 and ρ . Thus, we can confirm that, in fact, the slope term of the deterministic term is different from zero and the correlation parameters were, in average, equal to 0.73 for these six counties.

Because of the series being relatively short, we do not correct for conditional heteroskedasticity. Instead we assume that the series are conditionally homoskedastic. If series were relatively longer, a procedure that could be used to verify heteroskedasticity would be assign to the precision parameter τ (Figure 3) 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.

In Table 3 we show the predicted values of yields and its respective standard deviation and percentiles 5, 50 and 95% for the counties of Castro, Ponta Grossa, Marilândia do Sul, Tibagi, Catanduvas and Rolândia. The variance tends to increase as the time lag increases.

RATING THE CROP INSURANCE CONTRACT

Pricing an insurance contract accurately is essential for the viability and existence of an agricultural insurance market. Premium rates that are too high result in an insurance pool made up of only high risk individuals. Likewise, rates that are universally too low will result in insurance losses since premiums are not adequate to cover indemnity outlays. The selection problem that is brought about by inaccurate rates is known as adverse selection. In the literature of insurance economics, this is often also referred to as the hidden information problem since agents tend to know more about their risks than does the insurance provider.⁹

⁹ In the context of the principal-agent theory, the problem of hidden information or adverse selection occurs when the characteristics of the agent are imperfectly observed by the principal. In a classical article,

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:

$$\text{Premium Rate (PR)} = \frac{F_Y(\lambda y^e) E_Y[\lambda y^e - (Y | y < \lambda y^e)]}{\lambda y^e}, \quad (8)$$

where E is the expectation operator and F is the cumulative distribution function of yields. Note that the premium rate is completely transparent to the price at which yields are valued since the price term would appear in both the numerator and denominator of the premium rate expression.

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 (8) becomes:

$$PR = P(y^* < 1) E_{y^*}[1 - (y^* | y^* < 1)] \quad (9)$$

Note that the support of the random variable Y remains the same in this transformation. If we consider $w = 1 - y^*$, then equation (9) can be rewritten such that:

$$\begin{aligned} PR &= P(w > 0)[1 - E_w(1 - w | w > 0)] \\ PR &= P(w > 0) E_w[w | w > 0] \end{aligned} \quad (10)$$

After some simplification, the premium rate equation reduces to:

$$PR = \int_0^1 w f(w) dw \quad (11)$$

Akerlof (1970) analyzes the market of used cars (“lemons”), arguing that in a market where sellers have more information about the quality of the used car than buyers, then only the bad quality cars will remain in this market. One can note that, on average, the quality will be inferior compared to the price paid and in the limit the market for used cars will not exist due to the problem of asymmetric information. Analogous to the “lemons market”, Akerlof also pointed out that the health insurance market is also affected by the adverse selection. The higher the premium, the riskier is the insurance market, or in other words, only those individuals who really need to buy the insurance contract will do so, thus selecting only those more likely to receive the indemnity. Rothschild and Stiglitz (1976) analyzed a general model of a competitive insurance market. In this market the nature of the imperfect information lies in the fact that insurance companies are not capable to differentiate risks among buyers. Consequently, they argue that in a competitive insurance market the equilibrium cannot be reached. Possible “second best” solutions would be desirable for the viability of this market, such as, the implementation of the “auto-selective” insurance contracts, in which the insured would reveal to the insurers his risk structure constraint to the zero profit condition by the firm. Thus, contracts that could voluntarily be chosen would be offered to individuals, such that, the low risk individuals would be offered contracts with partial coverage at a lower premium. Likewise, low incentives (e.g., higher premiums) are given to the high risk group to buy the contract. The auto-selection mechanism was demonstrated to be a Pareto improvement if an individual can be categorized according to signals correlated with their risk. Another “second best” solution would be to formulate the contract according to the available information. If the insurer can monitor the insured, even imperfectly, then a Pareto improvement situation can be reached.

We can similarly write (11) as $PR = E[wI(0 < w < 1)]$. Because of the change of variable, the support also changed such that w lies now in between 0 and 1.

Premium rates were evaluated and compared using the mean posterior of w . In figure 6 below, we illustrate aggregate premium rates for regions in the state of Paraná. The state was divided into 10 large regions: Occidental Centre of Paraná (1), Oriental Centre of Paraná (2), Centre-South of Paraná (3), Metropolitan of Curitiba (4), Northwest of Paraná (5), Central North of Paraná (6), Pioneer North of Paraná (7), West of Paraná (8), Southeast of Paraná (9), Southwest of Paraná (10).

Figure 6 represents the average premium rate for the coverage levels (i.e., the percentage of the expected yield that is insured) of 70% through 90%, in multiples of 5, for each region. One can note that the largest rates are located in regions in the north and northwest of the state. We should note that for our model of yields that underlies rates is strongly influenced by the last observed yields (for 2002). Consequently, premium rates are strongly influenced by this value. To understand how yields in 2002 affect the premium rate, we must go back to the reparameterization of y in equation (9). One can see that, the smaller the value of y in relation to λy^e , the smaller will be y^* , thus, the larger the value of w . Because of the fact that rates are directly proportional to w , an increase in this variable will result in higher rates. This is, of course, a natural consequence of having such a short time-series of data. As more experience is accumulated, the effect of any single observation will be muted. Going back to Figure 2 one can see that counties situated in the northwest region, regions 1, 5, part of regions 6 and 8, in Figure 6, had relatively lower yields in 2002. Consequently, this situation results in higher premium rates in these regions. On the other hand, regions 2, 3 and 9 had relatively strong yield performance in 2002 and thus have relatively lower premium rates. Considering counties 1, 2, 3 and 4, in Figures 7 and 8, (below) one can clearly note that counties 1, 2 and 3, whose yields were relatively low in 2002, resulted in higher premium rates in relation to county 4. Thus, we again see that the premium rate is highly influenced by yields in 2002.

CONCLUSIONS

We have discussed a statistical and actuarial method of pricing a crop insurance contract that is based upon hierarchical Bayesian models. Our models of the probability generating process of yield data consider temporal and spatial effects as well as the interaction between these two effects, resulting in spatial-temporal models. The contracts are based upon a regional crop yield index. Such crop insurance plans have been adopted in many areas, including in the United States. Area-wide plans of this sort are now being implemented as an alternative risk management tool in the South of Brazil. We point out that this methodology can also be applied to contracts based on individual yields, as long as there are enough data to conduct the statistical analysis. 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 use of these new risk management tools, together with the approval of the Law n° 10,823 in December 2003, provide support for the development of a crop insurance market in Brazil. Likewise, these developments improve incentives for the entrance of new private insurance companies in this market. Finally, the new legislation includes improved incentives for agricultural producers to buy crop insurance contracts in the form of premium subsidies.

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 compared to other potential specifications that have been considered in the literature. In light of the rather small sample of data available, we demonstrate the sensitivity of premium rates to the yield observed in 2002. In particular, higher rates were found in the regions where yields were lower in this year.

We discuss the potential application of our methods to the general problem of pricing insurance contracts for individual coverage. We note that, to the extent that sufficient data are available, these methods may be applicable to the problem of pricing crop insurance contracts with individual coverage. 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. MOLLÍE, 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.
- 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.
- 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 selection. *Journal of the Royal Statistical Society, Ser. B*, v. 57, p. 247-262, 1995.
- 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, 1986.
- 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 Guelph, 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	$\rho_i y_{t-1} + \beta_{1_i} + \beta_{2_i}^C t^*$
2	673200	$\rho_i y_{t-1} + \beta_{1_i}^C + \beta_{2_i}^C t^*$
3	700100	R-W
4	728500	$\rho_i y_{t-1} + \beta_{1_i} + \beta_{2_i} t^* + v_i$
5	736800	AR(1)
6	737900	$\rho_i y_{t-1} + \zeta_i$
7	739900	$\rho_i y_{t-1} + \zeta_i^t$
8	751400	$\rho_i y_{t-1} + \beta_{1_i} + \beta_{2_i} t^* + \zeta_i^t$
9	751700	$\rho_i y_{t-1} + \beta_{1_i} + \beta_{2_i} t^* + \zeta_i$
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
Castro	2003	8301	791	6990	8303	9591
	2004	8455	1114	6647	8443	10280
Ponta Grossa	2003	6553	760	5296	6550	7793
	2004	6638	1008	5021	6628	8338
Marilândia do Sul	2003	7499	786	6208	7492	8784
	2004	7624	1074	5883	7614	9405
Tibagi	2003	7730	793	6419	7733	9019
	2004	7779	1094	6035	7753	9613
Catanduvas	2003	5968	758	4716	5972	7195
	2004	5833	903	4350	5813	7316
Rolândia	2003	7336	777	6068	7342	8615
	2004	7461	1079	5745	7433	9280

Figure 1: Corn yields in the state of Paraná, kilograms per hectare in 1990.

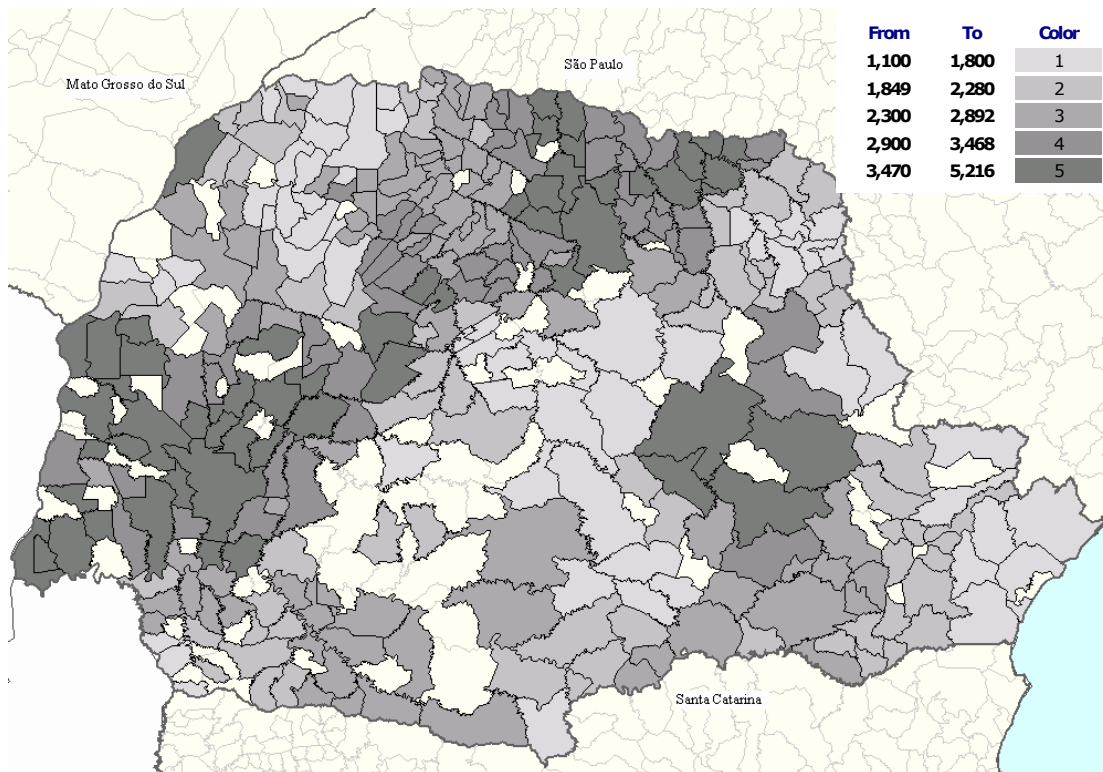


Figure 2: Corn yields in the state of Paraná, kilograms per hectare in 2002.

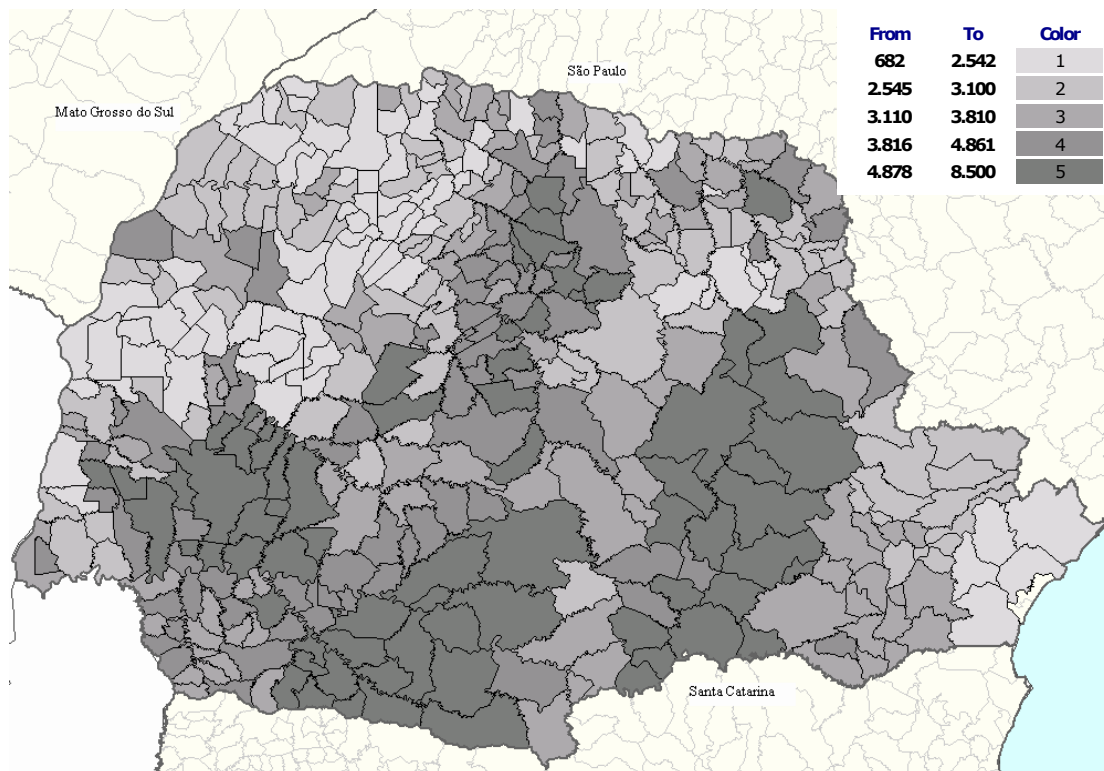


FIGURE 3. Graphical Description of Model 1

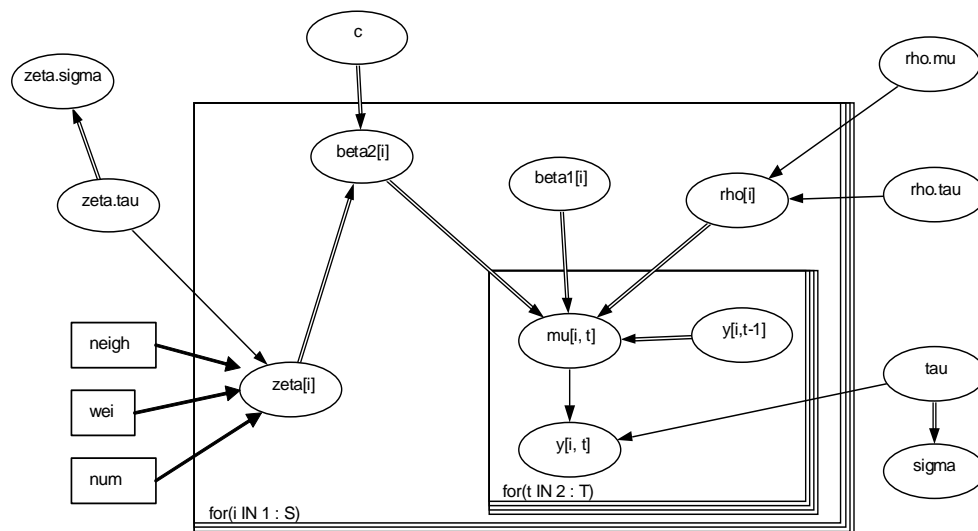
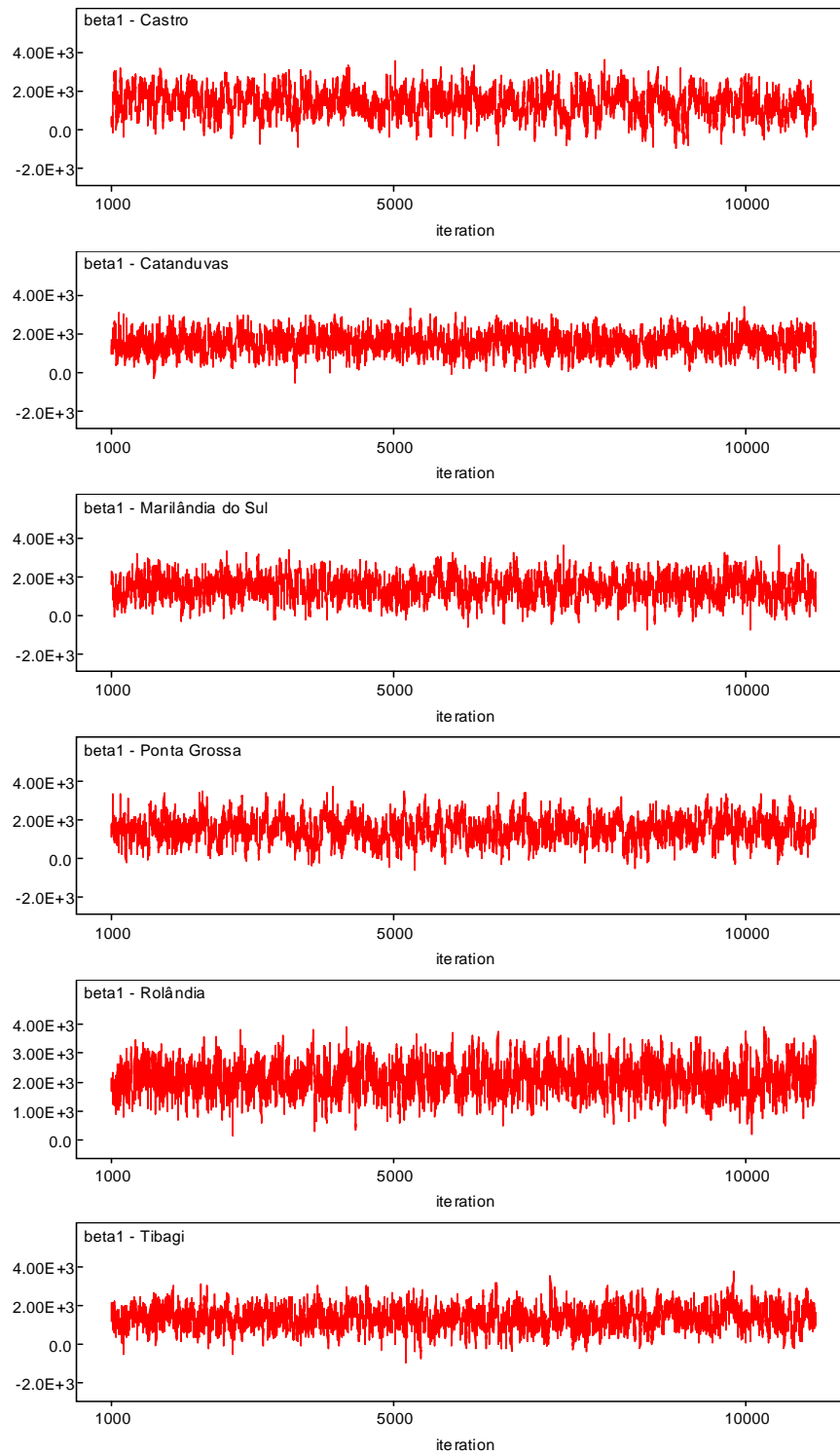
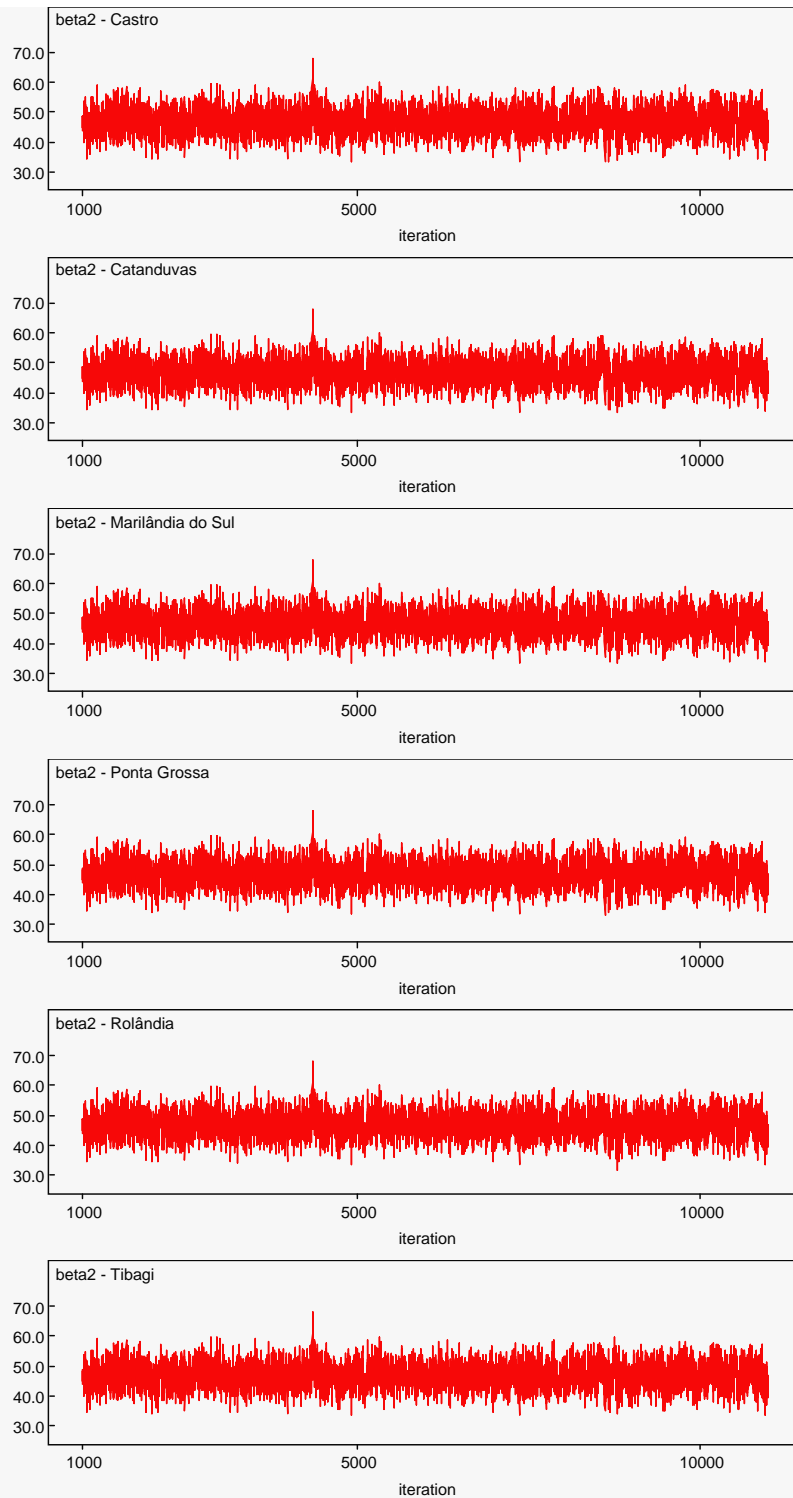


FIGURE 4. Convergence checking of β_1 , β_2 and ρ , respectively, for Castro, Ponta Grossa, Marilândia do Sul, Tibagi, Catanduvas and Rolândia.





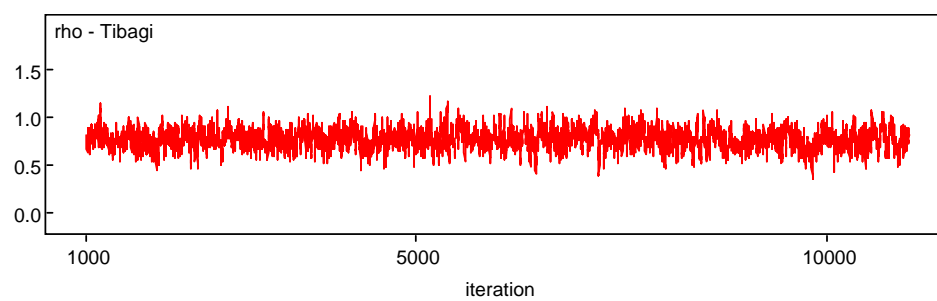
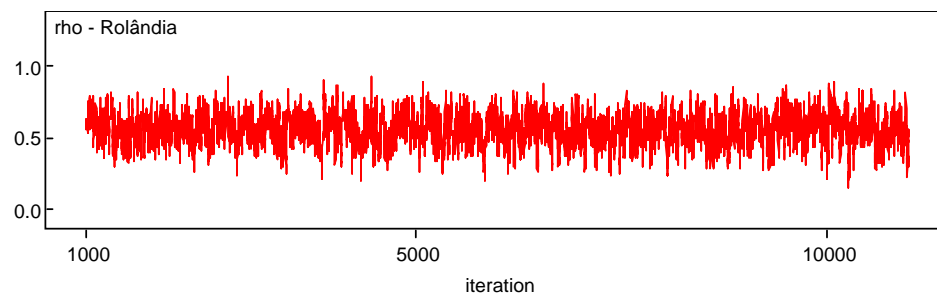
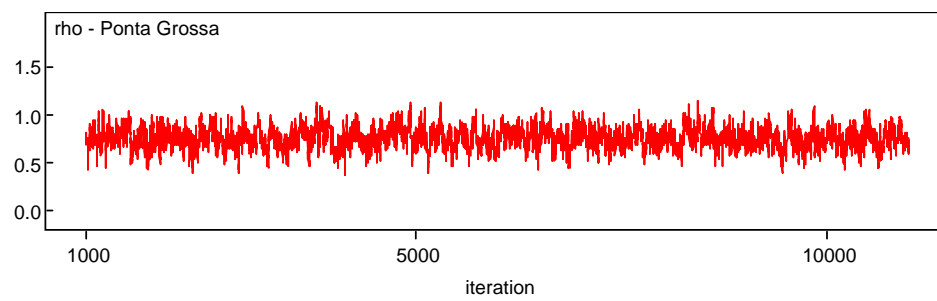
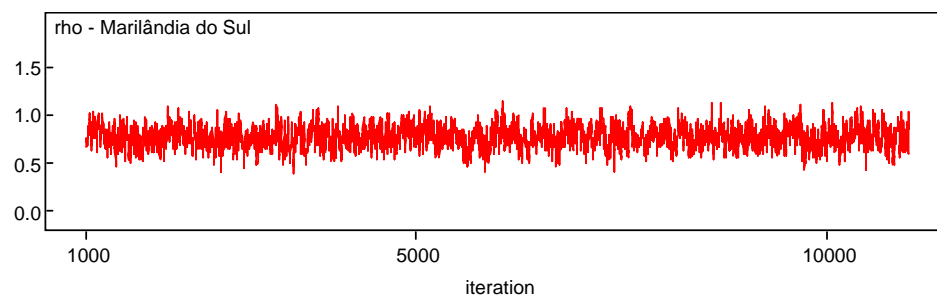
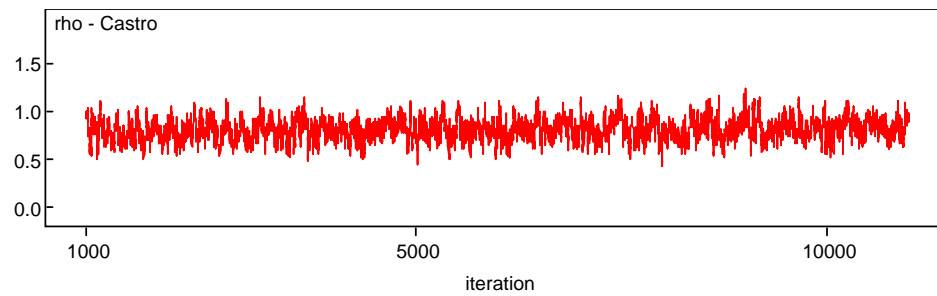


FIGURE 5. Decomposition of u_{it} in its deterministic and stochastic components.

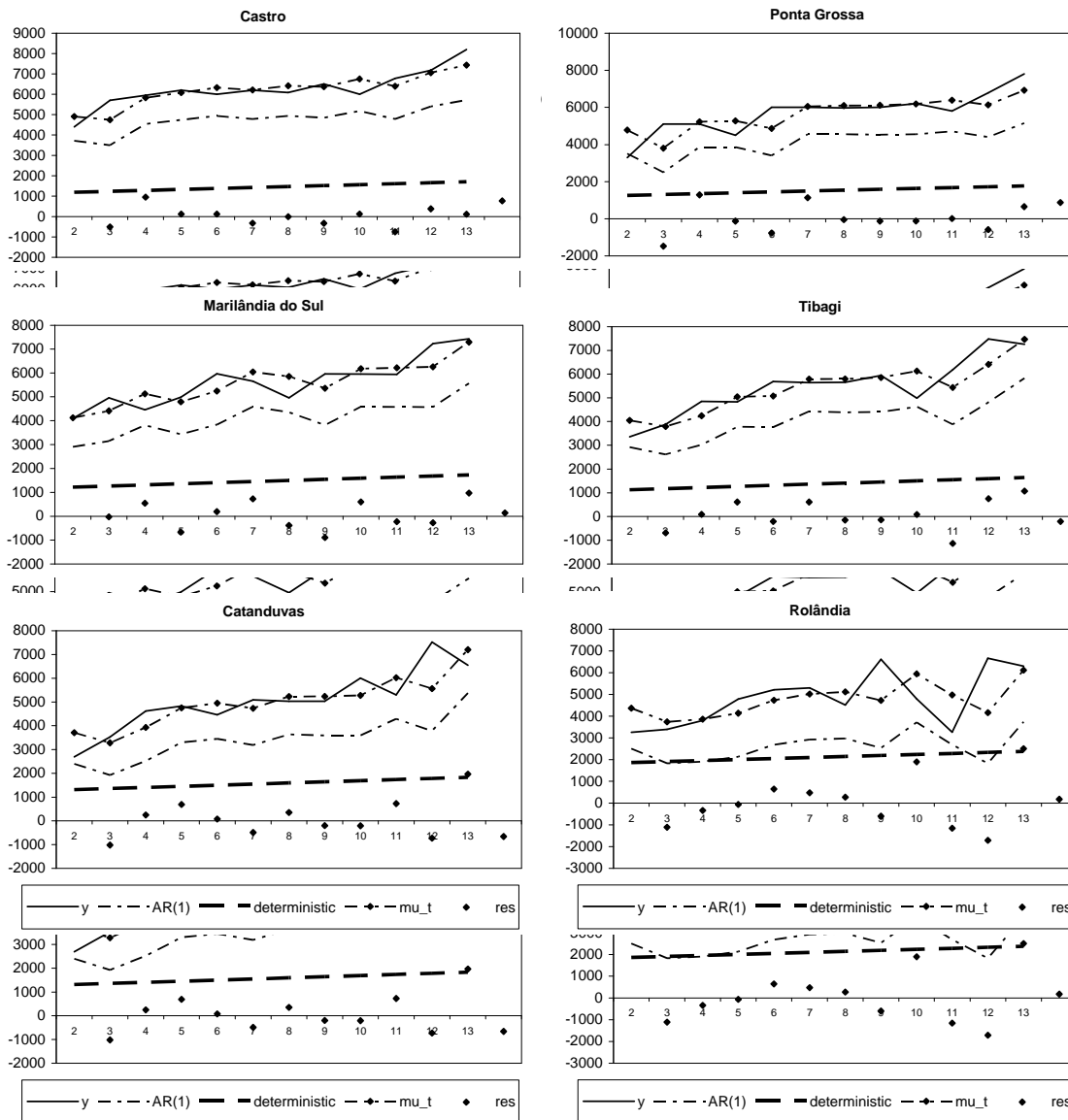


FIGURE 6. Posterior densities of β_1 , β_2 and ρ , respectively, for Castro, Ponta Grossa, Marilândia do Sul, Tibagi, Catanduvas and Rolândia.

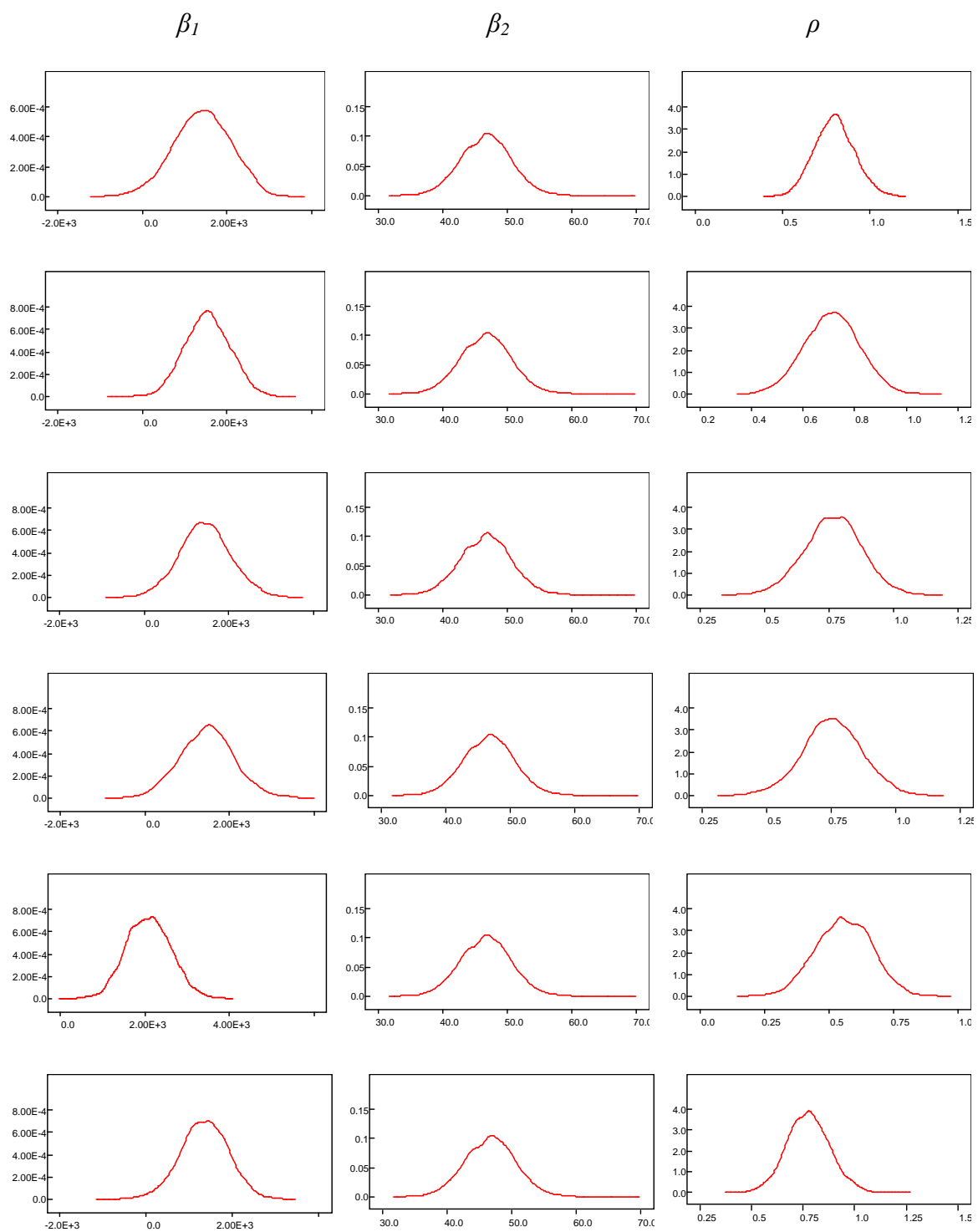


FIGURE 7. Premium rates (%) aggregated by regions in the state of Paraná

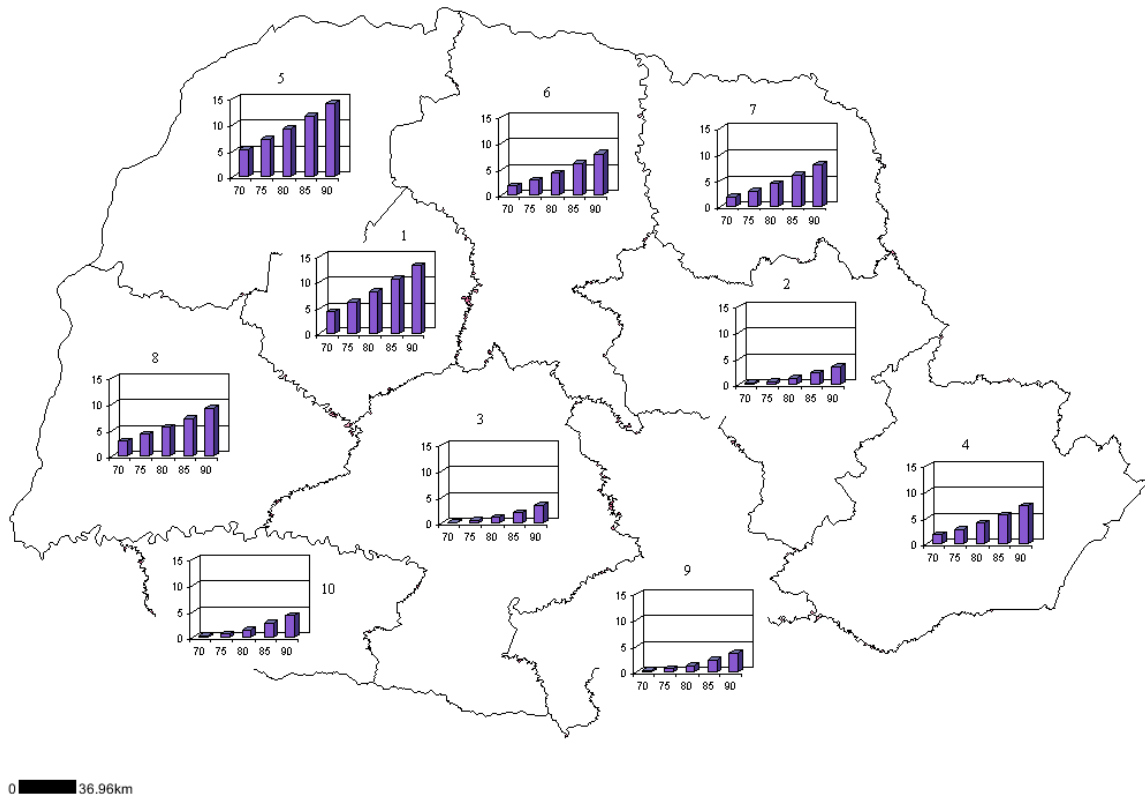


FIGURE 8. Corn yields in counties 1, 2, 3 and 4 (kg/hectare), 1990 and 2002.

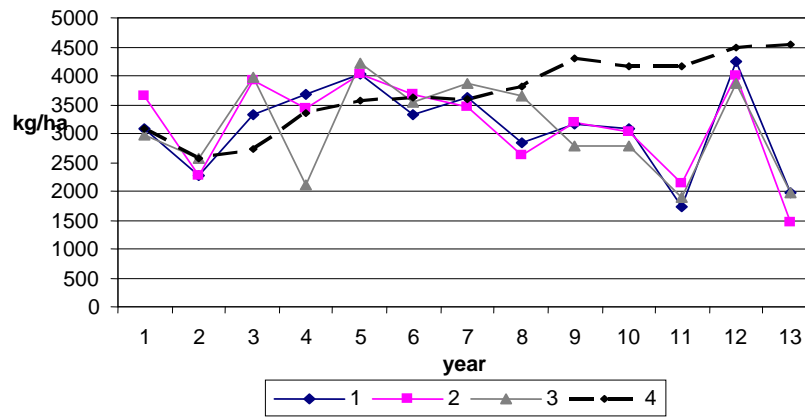


FIGURE 9. Premium rates (%) in counties 1, 2, 3 and 4 with coverage levels of 70, 75, 80, 85 and 90%.

