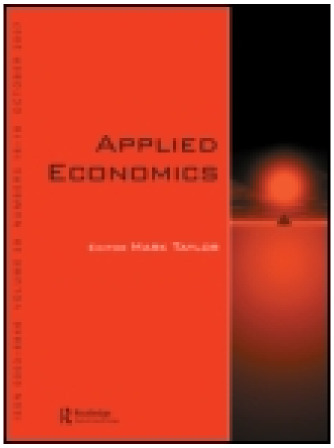

Applied Economics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/raec20

Parametric and nonparametric statistical modelling of crop yield: implications for pricing crop insurance contracts

Vitor A. Ozaki [a], Barry K. Goodwin [b] & Ricardo Shirota [c]

[a] Department of Mathematics and Statistics, Universidade de São Paulo/ESALQ, Av. Pádua Dias, 11 Agronomia/CEP 13418-900, Piracicaba, SP, Brazil

[b] Department of Agricultural and Resource Economics, North Carolina State University, Raleigh, NC, USA

[c] Department of Economics, Business and Sociology, Universidade de São Paulo/ESALQ, Piracicaba, SP, Brazil

Published online: 11 Apr 2011.

To cite this article: Vitor A. Ozaki, Barry K. Goodwin & Ricardo Shirota (2008) Parametric and nonparametric statistical modelling of crop yield: implications for pricing crop insurance contracts, Applied Economics, 40:9, 1151-1164, DOI: 10.1080/00036840600749680

To link to this article: http://dx.doi.org/10.1080/00036840600749680

Applied Economics, 2008, **40**, 1151–1164

Parametric and nonparametric statistical modelling of crop yield: implications for pricing crop insurance contracts

Vitor A. Ozaki[a,*], Barry K. Goodwin[b] and Ricardo Shirota[c]

[a]*Department of Mathematics and Statistics, Universidade de São Paulo/ESALQ, Av. Pádua Dias, 11 Agronomia/CEP 13418-900, Piracicaba, SP, Brazil*

[b]*Department of Agricultural and Resource Economics, North Carolina State University, Raleigh, NC, USA*

[c]*Department of Economics, Business and Sociology, Universidade de São Paulo/ESALQ, Piracicaba, SP, Brazil*

This article considers alternative methods to calculate the fair premium rate of crop insurance contracts based on county yields. The premium rate was calculated using parametric and nonparametric approaches to estimate the conditional agricultural yield density. These methods were applied to a data set of county yield provided by the Statistical and Geography Brazilian Institute (IBGE), for the period of 1990 through 2002, for soybean, corn and wheat, in the State of Paraná. In this article, we propose methodological alternatives to pricing crop insurance contracts resulting in more accurate premium rates in a situation of limited data.

I. Introduction

Protection against climate changes has been an important issue in agriculture. Because of the weather variation, agricultural production varies substantially, reducing the producer income. Producers usually bear the risk with little or even without any support.

Over the years, several risk management tools were created by producers to reduce these type of losses, including insurance (Harwood *et al.*, 1999). Traditionally, the private insurance market offers protection to individuals against insurable risks. However, in the agricultural sector, risks may not be completely insurable (Trowbridge, 1989; Redja, 1995; Hart *et al.*, 1996; Booth *et al.*, 1999; Skees and Barnett, 1999; Ozaki, 2005).

Several reasons are pointed out to explain the nonemergence of a private crop insurance market in these countries: adverse selection (Skees and Reed, 1986; Quiggin *et al.*, 1994), moral hazard (Chambers, 1989; Goodwin and Smith, 1996), systemic risk (Miranda and Glauber, 1997) and the absence of actuarial methods to accurately calculate the fair premium rate. In Brazil, the absence of a suitable methodology is pointed as one of the main problems

*Corresponding author. E-mail: vaozaki@esalq.usp.br

Applied Economics ISSN 0003–6846 print/ISSN 1466–4283 online © 2008 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/00036840600749680

for the development of an agricultural insurance market (Rossetti, 1998, 2001).

This article suggests alternative actuarial methods for pricing crop insurance contracts based on regional agricultural yield (area-yield insurance), considering the small number of observations available. The area-yield insurance has been used in the United States, India, Sweden and Canada (Miranda *et al*., 1999). In Brazil, it is currently offered in the State of the Rio Grande do Sul. This type of insurance was first studied by Halcrow (1949) and later, formally presented by Miranda (1991). Smith *et al*. (1994) and Mahul (1999), based on Miranda's paper, generalized the model. Skees *et al*. (1997) showed some empirical work on area-yield insurance.

II. Statistical Modelling of Agricultural Yield

Over many years, statistical aspects underlying agricultural yields have been a controversial point in the literature.[1] Precisely, the shape of the distribution has been discussed extensively. On one hand, Just and Weninger (1999) concluded that agricultural yields follow normal distribution. However, others found evidences against normality (Day, 1965; Taylor, 1990; Ramirez, 1997; Ramirez *et al*., 2003). Alternatively, beta distribution (Nelson and Preckel, 1989), inverse hyperbolic sine transformations (Moss and Shonkwiler, 1993) and Gamma distributions (Gallagher, 1987) have been proposed.

The shape of the distribution is particularly important in the context of crop insurance studies, because it reflects the risk (probability of loss) of the producer. In other words, when modelling agricultural yields, one must look at the mass concentrated at the left tail of the distribution. Taking this fact into account, several statistical models, have been proposed in the crop insurance literature by different authors to better reflect the innovation of the agricultural yield, such as, parametric (Sherrick *et al*., 2004; Ozaki, 2005), semiparametric (Ker and Coble, 2003), nonparametric (Goodwin and Ker, 1998; Turvey and Zhao, 1999), empirical Bayes nonparametric approaches (Ker and Goodwin, 2000) and spatial-temporal models (Ozaki, 2005).

The choice of a statistical model that better reflects the conditional density of yields is an important aspect in the actuarial calculation of the premium rate. In doing this, one must try to recover the generating process of data when modelling yield data. The agricultural yield follows a spatial-temporal process, in the sense that if we take the average in a region conditional to the temporal process, one can recover the conditional density yield $f(y|\Omega_t)$ at certain moment in time and region, where Ω_t is the minimum σ – algebra generated by the information known at moment t (Ker and Goodwin, 2000). In several empirical works, the only information known at time t is the time itself. Thus, in these analyses, the conditional density is based only on the temporal generating process of the data.

Problems in modelling yield data

Fair premium rate (or in other words, the expected loss) correct calculation is an important factor to insurers so that a crop insurance programme can be actuarially sound. The fact is that pricing against the risk of an average farmer will overcharge low-risk farmers and undercharge high-risk farmers. Thus, low-risk individuals are less likely to buy, remaining are only those riskier in the pool. Indemnities rise and the insurer loses money. Raising rates drives out low-risk end of pool, becoming smaller and riskier, increasing losses even more.

To accurately estimate the individual risk of each producer, insurers must take into account farm-level yield data. However, this estimation demands a reliable and long series of data set for each producer. In practice, this is not possible in developing countries due to the nonexistence of such information.[2] To partially overcome such restriction, index-based crop insurance were developed using aggregate and longer series, such as, county-level crop yields (Miranda, 1991).

Another problem related to empirically modelling agricultural yields is the spatial dependence across counties. The average value of independent and identically distributed observations, according to the central limit theorem (CLT), follows a normal distribution. But in the case of average observations of county yields, these observations are spatially correlated.

For spatial processes, alternative versions of the CLT, supporting normality assumption are available.[3] However, spatial correlation may or may not

[1] The small number of yield observations, in aggregate level and smaller in the individual level (in several countries) makes any type of statistical analysis a troublesome task.

[2] Moreover, with small number of observations it is hard to detect structural changes in yields and the occurrence of catastrophic events in the series.

[3] For the dependent spatial process the CLT the normality assumption is acceptable when the dependence dies off quickly as the distance increase (Guyon, 1995).

die relatively fast to assure that the average yield follows a normal distribution (Goodwin and Ker, 2002).

Goodwin (2001)[4] argues that, when regular years are considered, the spatial correlation dies faster when the distance increases, stabilizing in 0.10 when the distance between areas is approximately 200 miles. However, if years with droughts are considered, spatial correlation dies slower, reaching the same level of 0.10 with 400 miles. Consequently, for years in which extreme weather events occur, the systemic risk problem is higher than normal years.

Wang and Zhang (2003), using correlograms to estimate the spatial dependence for corn, soy and wheat, using 26 years, for, respectively, 2591, 2000 and 2641 cities in USA, concluded that the correlation dies faster when the distance increases. The maximum distance estimated for the positive correlation is equal to 570 miles.

In Brazil, the spatial correlation was found to reduce quickly when the distance increases supporting the assumption of normality for yield data. Ozaki (2005) estimated the parameters of the variogram (in particular, the range parameter) through the spherical correlation function, and showed that the average distance in which the correlation tends to zero is, respectively, 91 and 82 miles for soybean and corn, in the period of 1990 through 2002, much less than the values for the United States. Another important aspect of the agricultural yield modelling is the trend and the heteroskedasticity. Those problems appear when the data generating processes are not constant or stable over time. The incorporation of new technologies and more suitable and efficient methods by farmers increase the level of agricultural yields over time. Thus, yields observed in the early 80s cannot be compared with yields in 2004. If this is the case, yield data must be detrended according to some polynomial function. Several linear and nonlinear methods are proposed in the literature for this purpose. Autoregressive moving average models, local nonparametric smoothing and splines are some examples.

In this article, due to the shortness of our data set, detrend is performed with a first order deterministic trend model (in t), according to $y_t = \beta_0 + \beta_1 t + u_t$, where y is the agricultural yield and $t = 1, \ldots, 13$. Series with significant slope coefficient (at the 5% level) were detrended and the detrended yield was represented by the error term u_t.

The other point, previously mentioned, is the situation where the yield variability changes over time (heteroskedasticity). Nonconstant variability must be diagnosed, according to parametric (Goldfeld–Quandt) and nonparametric tests and corrected if necessary.

If the heteroskedasticity is assumed as deviations from the trend in relation to the level of the agricultural yields, then the assumption of constant coefficient of variation is supported. Proportional errors ε_t will be calculated dividing the error term u_t by its respective predicted value. The resulting values are homoskedastic (Goodwin and Ker, 1998). Goodwin and Mahul (2004) suggest multiplying $(1 + \textit{proportional error})$ by the yield observed in 2004, resulting in *normalized yields* y^n, such that:

$$y^n = (1 + \varepsilon_t) y_{2004} \tag{1}$$

In doing this we can express yields in terms of 2004 technology.

On the opposite side, if errors are not proportional to the level of yields, then normalized yields will be calculated adding the error term to yields observed in 2004.

$$y^n = u_t + y_{2004} \tag{2}$$

In this study, normalized yields are calculated according to both methods (additively and multiplicatively), to check for possible differences in premium rates. Moreover, because the last observation released by IBGE is the one observed in 2002, all residuals will be normalized to 2002 technology level.

III. Description of the Data Set

In Brazil, county-level data series are available for 13 years from the Statistics and Geography Brazilian Institute (IBGE), in kilograms per hectare, from 1990 to 2002. This article uses yields in counties with more than 10 years of observations for soybean, corn and wheat, resulting in 267, 366 and 266 counties in the State of Paraná, respectively. Figure 1 shows the evolution of the agricultural yield in the State of Paraná, for soybean, corn and wheat, in 1990, 1996 and 2002.

In 1990, high yields were presented in the west, north and middle east region of the state for corn. This trend remained for 1996 and 2002. In the last year, counties in the south region improved the performance ranging in this year between 4800 to 8500 kg/ha. In the case of soybean, the pattern remained the same for the same period, but the number of producer counties increased dramatically during this period.

[4] Studying the spatial correlation of corn in USA in the three largest producer States.

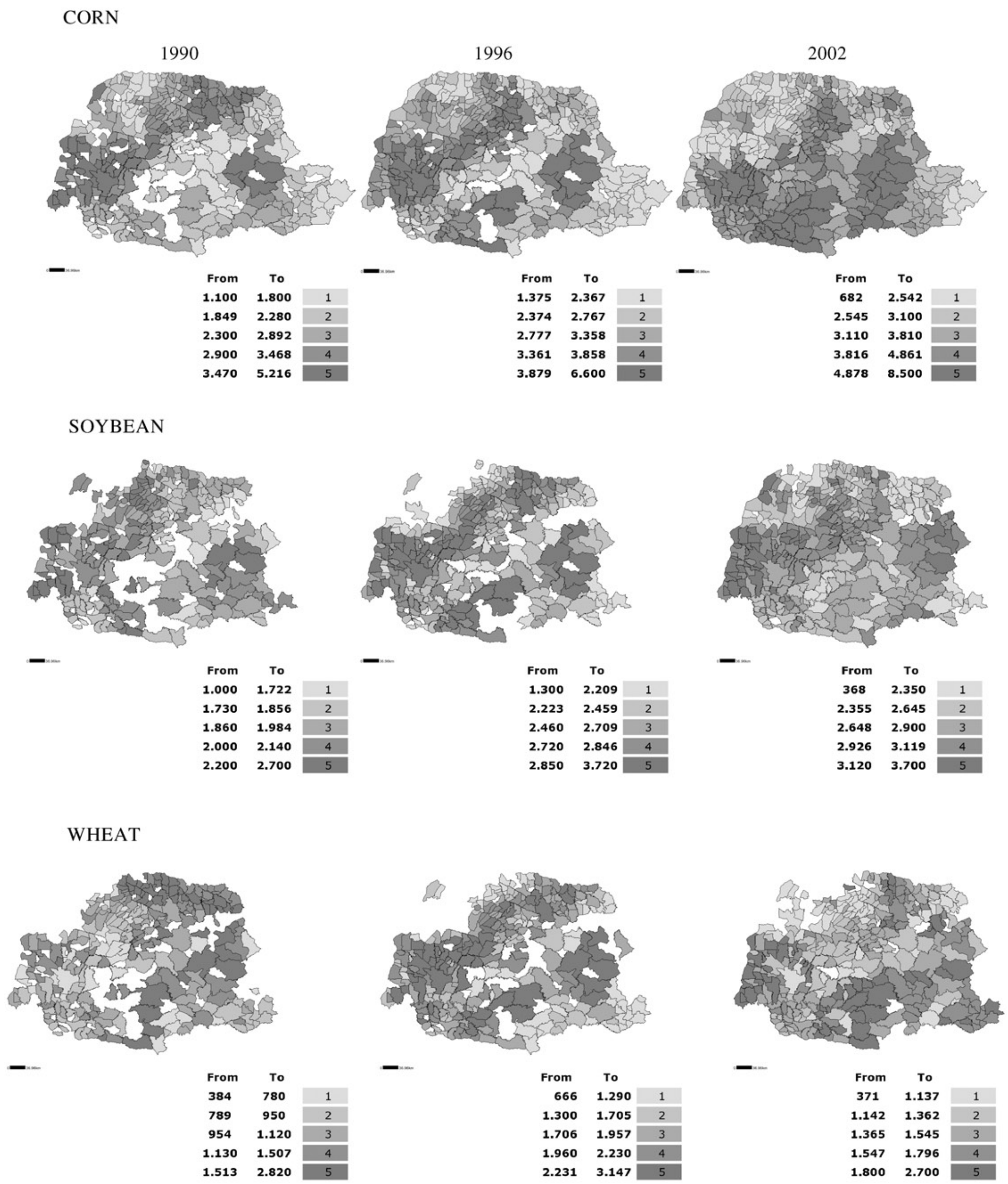

Fig. 1. Evolution of corn, soybean and wheat yields in the State of Paraná, in 1990, 1996 and 2002, in kilograms per hectare (kg/ha)
Source: IBGE (2004).

Different patterns can be observed in the case of wheat. In 1990, counties with high yields were concentrated in the north and middle East. This situation changed in 1996 with most of the high yields counties located in the middle west and west part of the state. In 2002, a better performance was achieved basically in the extreme west, south and east.

IV. Parametric Analysis

The methods commonly used to pricing crop insurance contracts are based on the relation of the average loss over liability, called empirical rates, and do not take into account any more advanced statistical analysis.[5] One of the main disadvantages

[5] Information obtained during personal interview with several private insurance companies operating in Brazil.

of the empirical rates method is its dependence on a large number of observations to accurately reflect the probability distribution.

This study fits two parametric distributions to the data set, normal and beta distributions, with all parameters estimated by the maximum likelihood method. The normal distribution does not allow for the positive or negative skewness, or in other words, it is symmetric.

The alternative beta distribution has the advantage of being flexible, assuming several different shapes according to the values of the parameters (α, β). The distribution can be strictly increasing ($\alpha > 1$, $\beta = 1$), strictly decreasing ($\alpha = 1$, $\beta > 1$), U form ($\alpha < 1$, $\beta < 1$) or unimodal ($\alpha > 1$, $\beta > 1$) (Casella and Berger, 1990). Thus, it can accommodate positive or negative skewness. Both parametric distributions have the disadvantages of not being able to model bimodality or multimodality.

V. Nonparametric Analysis

Another actuarial method presented is the nonparametric analysis of yield data.[6] In this case, no prior specification is given to define the shape of the distribution, or in other words, it lets the data reveal the shape of the density.

For actuarial reasons, there are two important variables to consider when pricing an agricultural insurance contract: the expected loss and the probability of loss. The latter is represented by the probability of yields less than a predetermined yield value equal to a percentage of the expected yield.

In the nonparametric analysis, some characteristics of the distribution can be shown, such as, positive and negative skewness and bimodality. Moreover, this method does not make any prior assumption on the shape of the density. Amongst the several types of the nonparametric density estimators, the histogram is by far the most popular. To estimate histograms one must take into account the choice of the origin and the width of bins h (also known as 'windows').

Several histograms can be estimated according to different values of the origin. Thus many different interpretations might arise. The width of bins determine the amount of smoothness of the series. Despite its simplicity, this estimator presents disadvantages in relation to the Kernel estimator. It is straightforward to estimate histograms in only one dimension, despite the choice of the origin and the smoothing parameter. But working on an m-dimensional space, the estimation becomes a troublesome work. However, one may find it useful to work with histograms to make exploratory analysis of the data.

Goodwin and Ker (1998) and Turvey and Zhao (1999) used the Kernel estimator to estimate the shape of the conditional yield density and pricing a crop insurance contract. The Kernel estimator of the density $\hat{f}(y)$ can be represented as a convolution of the sample distribution, using some kernel function, according to Goodwin and Ker (2002), such that:

$$\hat{f}(y) = \int K_h(y - v)\mathrm{d}F_n(v) \tag{3}$$

where $K_h(v) = 1/hK(v/h)$ and $F_n(v)$ is the sample distribution function.

Basically, the kernel estimator is the sum of 'jumps' (bumps) located in each observation, where the kernel function determines the shape of these jumps, and the smoothing parameter its width. The larger the window value, the larger the smoothness and the details tend to disappear. On the opposite side, the closer the window to zero, 'jumps' will have a peak shape, enhancing the details in the density.

Some assumptions must be made with respect to K. The kernel function must be a density function, nonnegative and symmetric, such that: $\int K(v)\mathrm{d}v = 1$. Moreover, $\int vK(v)\mathrm{d}v = 0$ and $\int v^2K(v)\mathrm{d}v = \theta_2 \neq 0$. If one considers the Gaussian kernel function, then θ_2 will be the variance of the distribution.

As a discrepancy measure of the density estimator $\hat{f}$ with relation to the true density, Silverman (1986) adopts the mean integrated squared error (MISE) given by:

$$\int E\left[\hat{f}(y) - f(y)\right]^2 \mathrm{d}x \tag{4}$$

The MISE can be decomposed in two components, the integrated squared bias and the integrated variance. Under some assumptions, the former can be approximated by $1/4h^4\theta_2 \int f''(y)^2\mathrm{d}y$ and the variance by $(nh)^{-1}\int k(v)^2\mathrm{d}v$.

Yet, if we choose a high value of the smoothing parameter to minimize the MISE, the random variation (variance) is reduced leading to the increase of the systematic error (bias). On the opposite side, a low value of the parameter results in higher integrated variance and lower integrated squared bias.

[6] This subsection was based on the results of Silverman (1986).

Looking for the optimum smoothing parameter and the kernel that minimizes the MISE, Silverman shows that if h is the optimum, then:

$$h_{\text{opt}} = \theta_2^{-2/5}\left(\int K(v)^2 \mathrm{d}v\right)^{1/5}\left(\int f''(y)^2 \mathrm{d}y\right)^{-1/5} n^{-1/5} \tag{5}$$

Thus the approximate value for the MISE will be given by:

$$\frac{5}{4} W(K)\left(\int f''(y)^2 \mathrm{d}y\right)^{1/5} n^{-4/5} \tag{6}$$

where $W(K) = \theta_2^{2/5}(\int K(v)^2 \mathrm{d}v)^{4/5}$.

The problem of minimizing the MISE is the choice of minimum $W(K)$, given the smoothing parameter. Under some constraints, if the kernel function chosen is the *Epanechnikov* kernel $W(K_e)$, then the MISE will be minimized. Comparing several symmetric kernel functions $W(K)$, an efficiency index was calculated to be equal to $[W(K_e)/W(K)]^{5/4}$. Using the Biweight, Triangular, Gaussian and Rectangular densities, it was showed that all kernel functions have a very close index. Thus, the choice of the kernel function has little influence under the criteria described earlier.

The choice of the optimum smoothing parameter, in the case of the Gaussian distribution, will be equal to $1.06\sigma n^{-1/5}$, where σ is the SD of the yields. If one wants to consider the deviation from normality, Silverman (1986) suggests that the following estimate must be used $\sigma = \min(SD\ interquartiel/1.34)$.

Moreover, if the factor 1.06 is reduced to 0.9 this will lead to better empirical results. Consequently, this estimate of the smoothing parameter was considered in the analysis.

VI. Empirical Analysis

Because of the small number of observations used to calculate the premium rate and considering the sensitive and stability of the nonparametric methods to small samples, neighbour counties[7] information were also used. Consequently, considering a central county i observations of the jth neighbour counties in relation to i were incorporated in order to increase the number of observations used, to estimate the conditional yield density and reduce the spatial dependence between counties. Accordingly, weights were respectively assigned to a central and m neighbour counties by:

$$\left.\begin{array}{r} \dfrac{(m+1)}{(2m+1)} \\ \text{and } \dfrac{1}{(2m+1)} \end{array}\right\} \tag{7}$$

In the analysis, counties within 30 and 40 miles (approximately 48 and 64 km) of distance from a central county were considered. Differences in premium rates will be analysed, considering the distance of the neighbour counties.

The largest corn producer was Guarapuava, with almost 172 000 ton. Soybean and wheat's largest producers were respectively, Cascavel and Tibagi, with approximately 210 000 and 60 000 ton.

The effect of the distance 'd' on the densities of yield distribution is shown in Fig. 2. Figure 3 presents the distribution of yields using the corrected and original series.

Figure 2 shows significant differences in the mass concentration in the left tail of the density. This difference is larger in the case of corn and wheat. In both cases, the probability of loss (the area under the density where yields are less than the trigger yield) is larger for $d = 30$ than $d = 40$. Additionally, because the premium rate is directly proportional to the probability of loss, higher rates are expected in the case of corn and wheat in the counties of Guarapuava and Tibagi, compared to soybean. In fact, this is confirmed in the next section.

Some other interesting features can be pointed in the densities. Bimodality is considerably visible in the case of corn, hence one can expect that higher yields can be considered more frequently than lower yields in Guarapuava county. In the case of soybean, bimodality is close to the mean, for $d = 30$, but not for $d = 40$. Moreover, the low variance suggests that values closer to the mean are more likely than extreme values. The wheat in Tibagi County, shows little evidence of bimodality, but for higher levels of coverage, the premium rate tends to be higher than for low levels of coverage, due to the fact that the probability of loss is larger when $d = 30$.

In addition to the distance, another point must be investigated. The fact that yield series is relatively short creates difficulties to check the presence of heteroskedasticity. Consequently, the next paragraphs study the densities in both cases. First, when yield series are adjusted correcting the residuals, resulting in proportional errors. Second, when series are nonadjusted for heteroskedasticity. In the latter

[7] We consider neighbour counties not only the contiguous counties to a central county i but all those counties located within a circular area with radius equal to d. Consequently all counties whose centroids were located within this circle were considered neighbours to i.

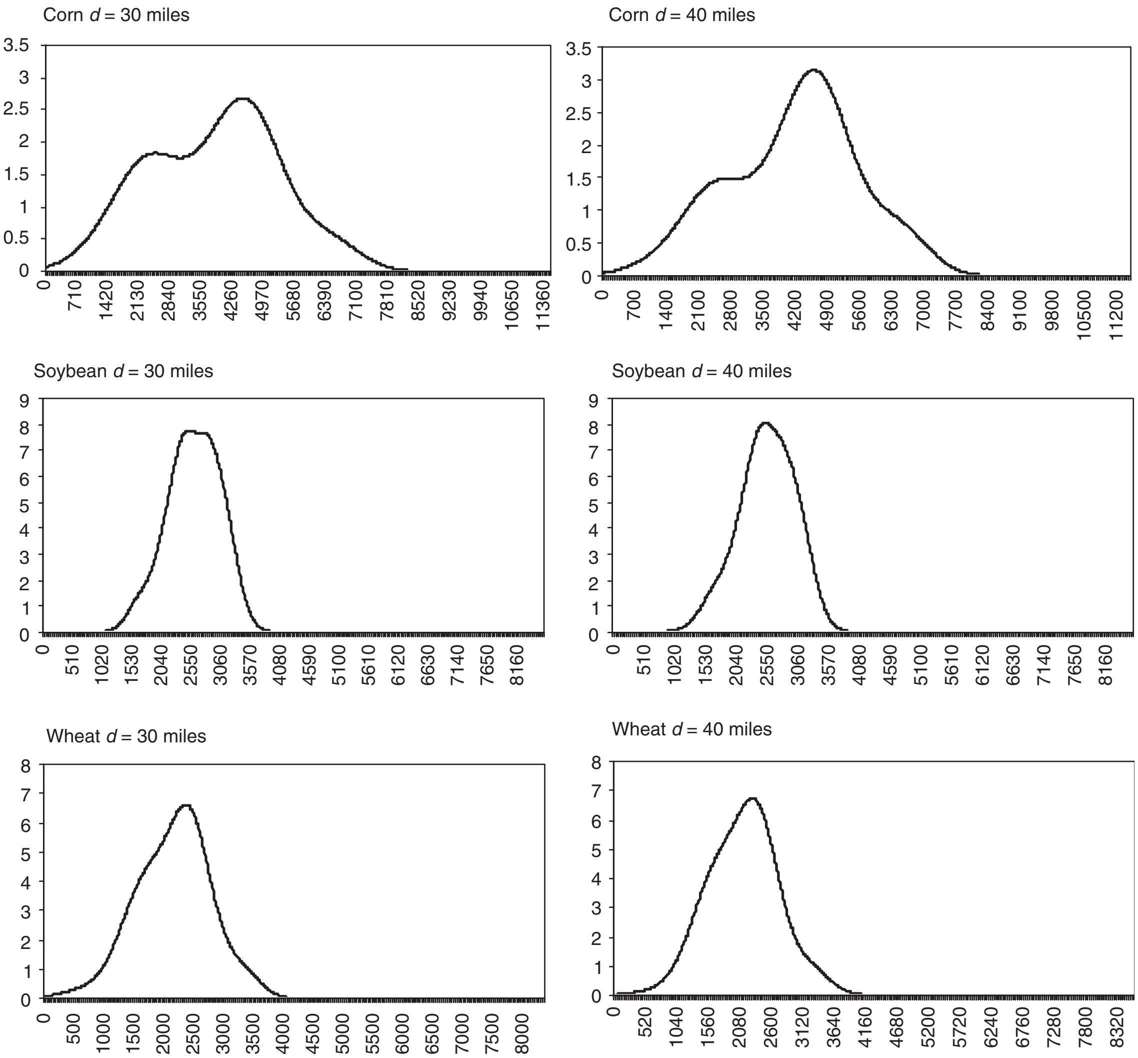

Fig. 2. Nonparametric densities (10^{-4}) for corn, soybean and wheat, according to the distance *d*, for corn (Guarapuava County), soybean (Cascavel County) and wheat (Tibagi County)

case, we assume that series are homoskedasticity. In both cases, the observations were detrended and normalized to reflect the 2002 technology.

Corn, in Fig. 3, shows a clear bimodality in both cases. But, there are little differences in the probability of loss, such that the probability in the adjusted series is smaller than the probability considered in the other case when series are adjusted. In the case of soybean, differences are more enhanced. In the original series, the density is less peaked than the density estimated considering the adjusted case. In the latter case, the density shows bimodality (Fig. 2) and smaller variance. Further yields above the average, around 2514 kg/ha, for Cascavel County, are more frequently than low yields, consequently for this county one can expect lower premium rates in relation to other counties.

In the case of wheat (Tibagi County), one can notice some differences in the probability of loss regarding the different levels of coverage between the two graphs. The value of this probability for the level of coverage of 70, 80 and 90% for the nonadjusted series are, respectively, 0.21, 0.32 and 0.45 and equal 0.15, 0.23 and 0.33 for the adjusted series. The former shows average values around 30% larger than the adjusted series.

VII. Application: Pricing Crop Insurance Contracts

The probability of loss is equal in the area under the curve when yield is smaller than the guarantee yield. Thus, let λ be the level of coverage, such that

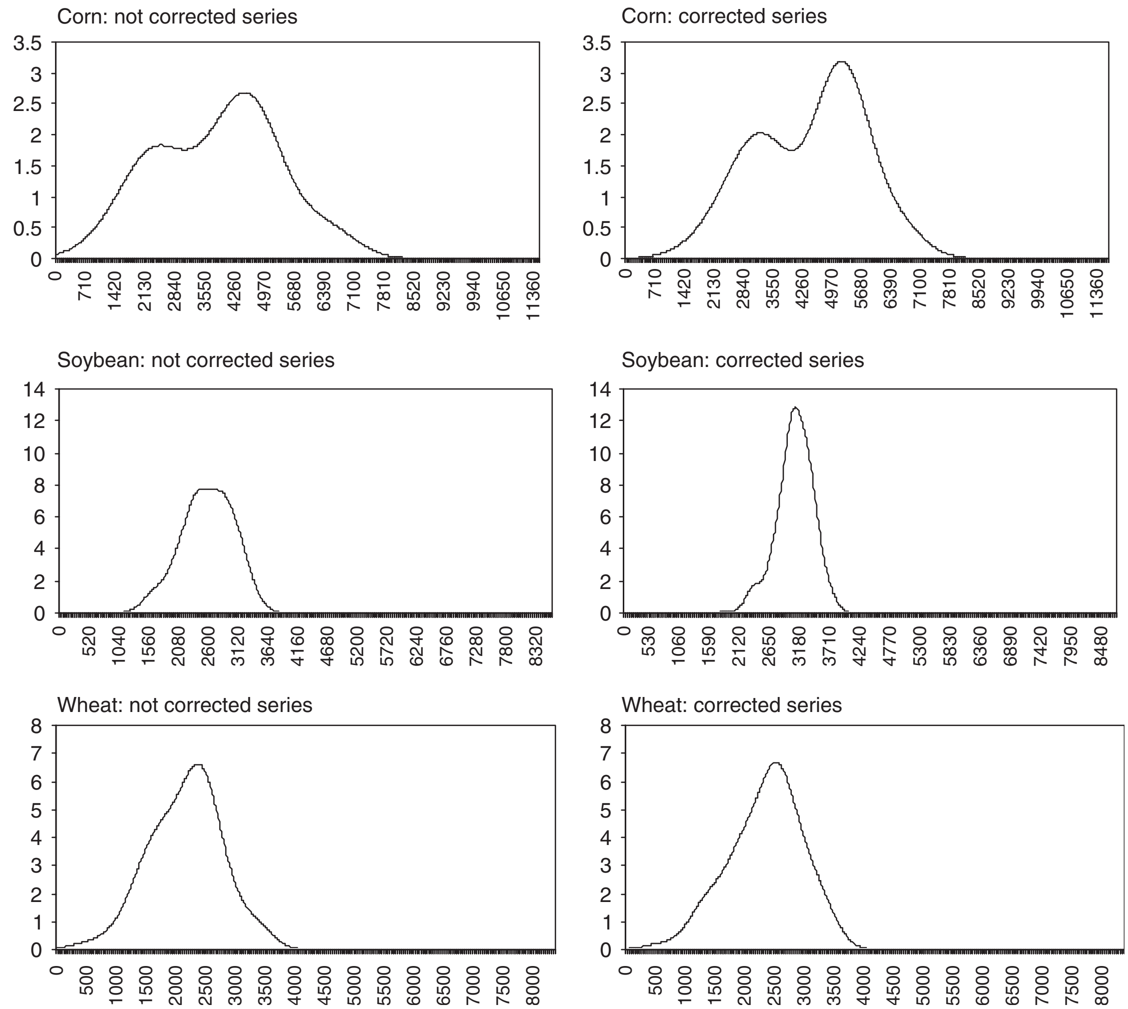

Fig. 3. Nonparametric densities (10^{-4}) for corn, soybean and wheat, for adjusted and nonadjusted series (distance fixed at 30 miles)

$0 < \lambda < 1$ and y^e be the expected yield. The probability of loss will be calculated through the area under the density, using the trapezoidal rule to estimate it numerically. The premium rate will be given by (Goodwin and Ker, 1998):

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

where E is the expectation operator and F is the distribution function.

Next, in the tables, we will show the premium rates, separated by methods of estimation, considering the correction of the series, the distance in relation to a central county and the level of coverage for the largest counties producers of corn, soybean and wheat (respectively, Cascavel, Guarapuava and Tibagi Counties). Figures 4 through 9 illustrate the premium rates for each county in the State of Paraná for corn, soybean and wheat calculated through the nonparametric method, at the level of coverage 90% with both distances and series correction changing.

In Tables 1–3, are presented the premium rates of corn, soybean and wheat, for different levels of coverage. These tables also present the rates calculated through the parametric method; specifically we adjusted the normal and beta distribution to the data.

Table 1 shows rates for soybean in Cascavel county. Notice that beta rates are much larger than empirical, nonparametric and normal rates, suggesting that soybean yields in Cascavel have positive skewness considering the beta distribution. Moreover, nonadjusted series rates are much higher than adjusted series for 30 and 40 miles, suggesting that heteroskedasticity, if present, can affect the premium rates calculation in the case of soybean in Cascavel.

The choice of the distance '*d*' can influence the calculation of the premium rate. The larger the value

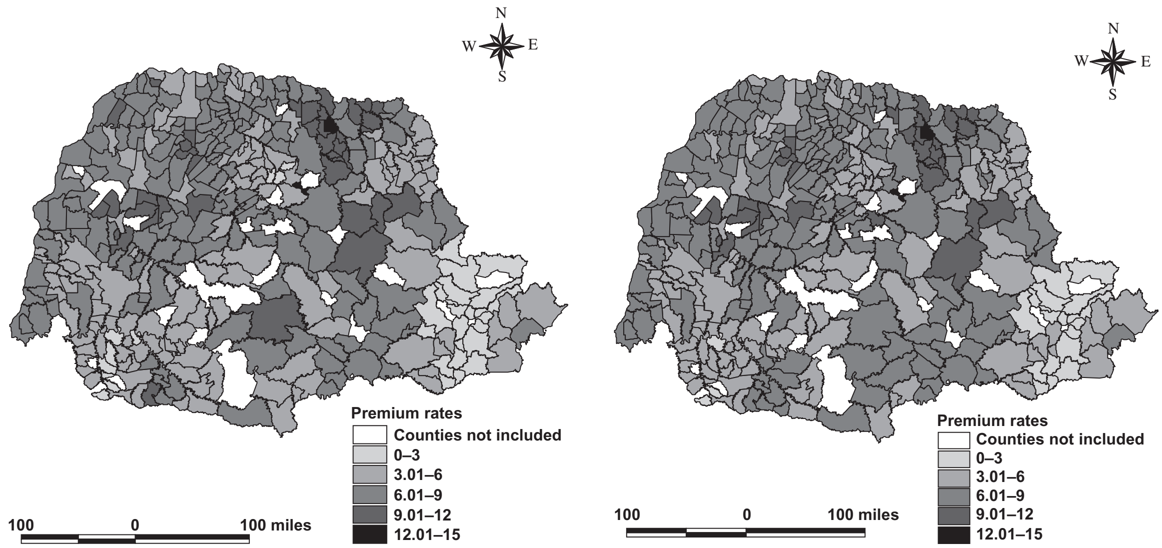

Fig. 4. Nonparametric rates for corn – series *adjusted* for heteroskedasticity (level of coverage at 90%), distance fixed at 30 miles (left) and 40 miles (right)

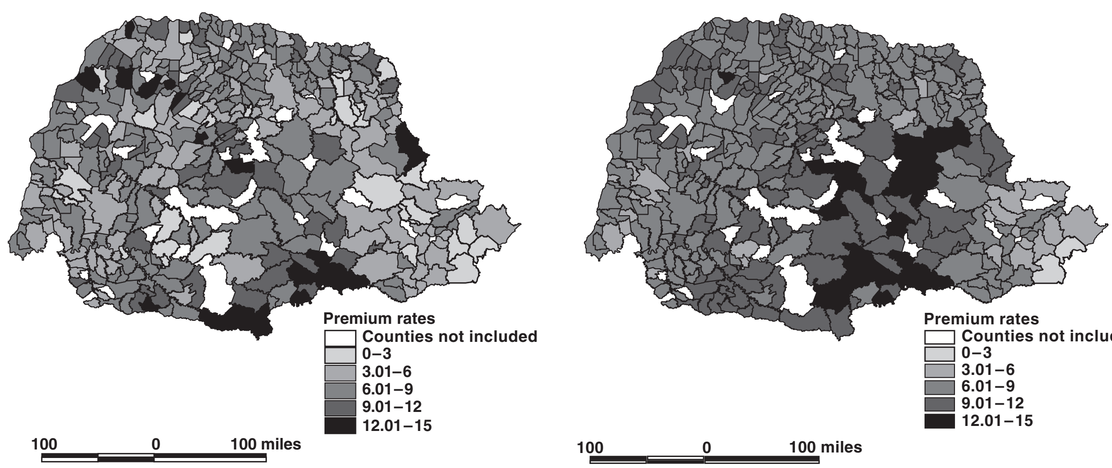

Fig. 5. Nonparametric rates for corn with nonadjusted series for heteroskedasticity (level of coverage at 90%), distance fixed at 30 miles (left) and 40 miles (right)

of '*d*', the greater the influence of distant counties on the density estimation (affecting premium rates). Specifically, in the case of soybean the larger the distance, the larger the premium rate for both series.

Table 1 shows that rates for the adjusted series are, on an average, 35% higher for $d = 40$ than $d = 30$ and 12% larger for the nonadjusted series. Rates are higher, on average, 28 and 18%, respectively, for the adjusted and nonadjusted series.

Table 2 presents corn rates in Guarapuava County. Contrarily to soybean, in the case of corn, the larger the distance, the smaller rates considering the adjusted series. Using nonadjusted series, rates increase with the distance.

The normal rates are smaller than the beta rates, suggesting that the beta distribution has positive skewness. Moreover, when comparing both series, for the distance $d = 30$, the adjusted series show higher rates than the nonadjusted series for all levels of coverage.

Rates for the nonadjusted series (when $d = 40$) are higher than beta rates for both methods. This situation suggests that the incorrect specification of the parametric distribution can have a tremendous impact on the calculation of the premium rate.

Table 2 shows that rates for the adjusted series are 21% higher on average for $d = 30$. Rates are 24% higher for $d = 30$ than for $d = 40$ considering the adjusted series. For the nonadjusted series, rates are

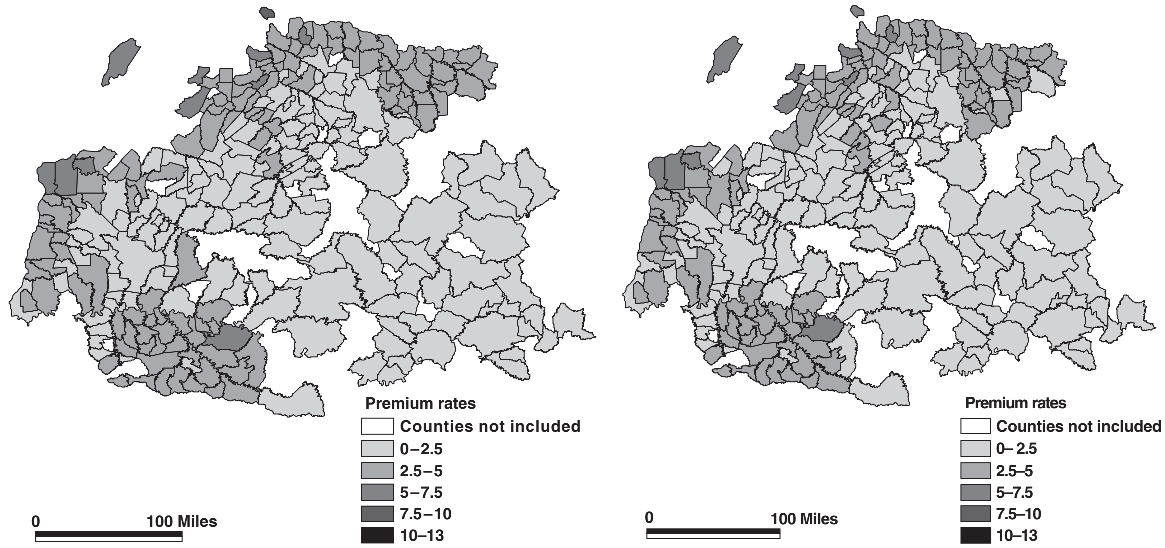

Fig. 6. Nonparametric rates for soybean with series adjusted for heteroskedasticity (level of coverage at 90%), distance fixed at 30 miles (left) and 40 miles (right)

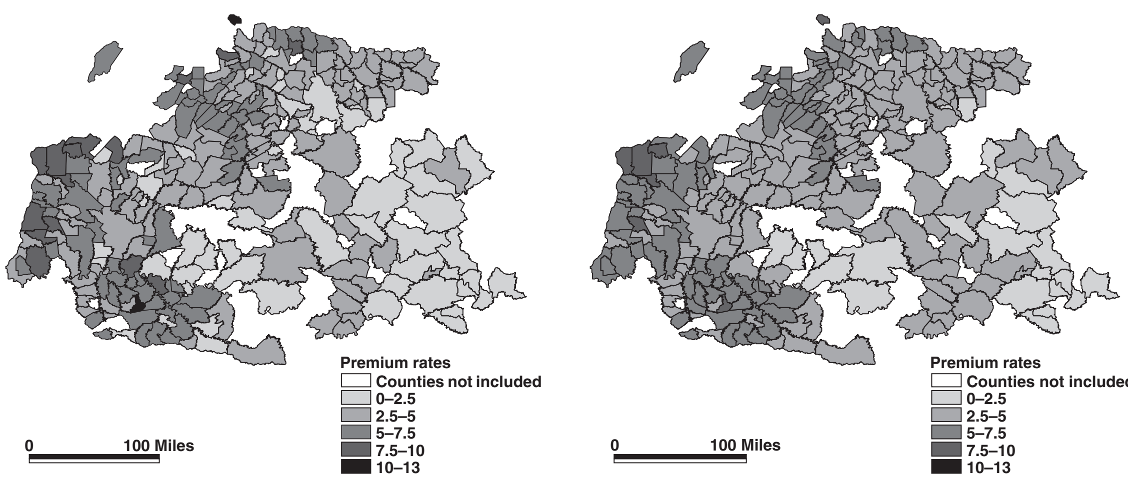

Fig. 7. Nonparametric rates for soybean with nonadjusted series for heteroskedasticity (coverage at 90%), distance of 30 miles (left) and 40 miles (right)

46 and 48% higher, on average, for $d=40$, respectively.

In Table 3, we show the premium rates for wheat in Tibagi County. Although there are differences in the rates, such differences are less expressive than the corn and soybean case.

Comparing the two series, one can notice that rates for the adjusted series in are higher for the level of coverage of 70 and 75% for both distances and lesser for the other levels of coverage. Rates for the adjusted series are lesser for all levels of coverage.

In relation to the normal rates, the beta rates are also higher. In this case, when distance is considered, the situation is inverse the one observed in the case of soybean, or in other words, the larger the distance the lesser the rate. Rates for the nonadjusted series are, on average, 15% higher for $d=30$ than for $d=40$. Additionally, when the levels of coverage increase, rates also rise but at decreasing rates, or in other words, for the levels 70 to 90%, the differences are, respectively, 21, 18, 16, 13 and 9%. Rates for the adjusted series are, on average, 13% higher when $d=30$. Looking at the nonadjusted series, one can notice that rates are higher for $d=30$, on average, 14 and 12%, respectively.

Looking at Tables 1–3, one can realize that rates calculated by the nonparametric approach are higher than empirical rates. This situation happens because the nonparametric rates are smoothed versions of the empirical rates. The smoothing process tends to add

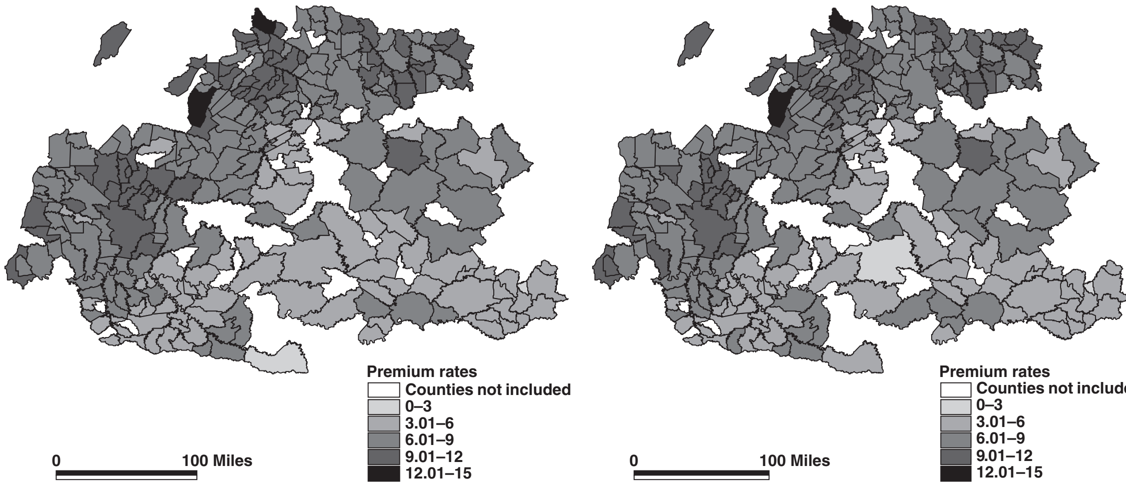

Fig. 8. Nonparametric rates for wheat, series adjusted for heteroskedasticity (coverage at 90%), distance of 30 miles (left) and 40 miles (right)

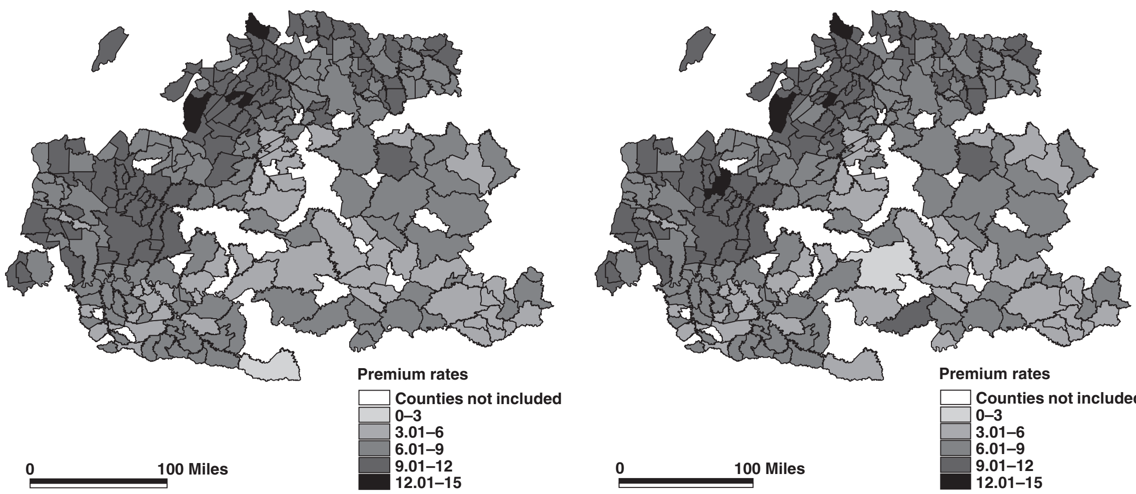

Fig. 9. Nonparametric rates for wheat, nonadjusted series for heteroskedasticity (coverage at 90%), distance of 30 miles (left) and 40 miles (right)

Table 1. Premium rates (%) for soybean, in Cascavel County

			Empirical premium rates				Nonparametric premium rates			
			Adjusted series		Nonadjusted series		Adjusted series		Nonadjusted series	
LC	Normal rates	Beta rates	$d=30$	$d=40$	$d=30$	$d=40$	$d=30$	$d=40$	$d=30$	$d=40$
70	0.3501	1.1364	0.0513	0.1169	0.6120	0.6759	0.0689	0.1400	0.7198	0.8979
75	0.6965	1.8961	0.0830	0.2118	1.0840	1.1121	0.1807	0.2819	1.1814	1.4252
80	1.2288	2.9264	0.3813	0.5099	1.4970	1.6848	0.4244	0.5561	1.7041	2.0766
85	1.9712	4.4149	0.7263	0.9049	2.0600	2.5404	0.8518	1.0213	2.4641	3.0094
90	3.1059	6.3614	1.3001	1.5077	2.9120	3.5944	1.5062	1.7025	3.4764	4.2388

Table 2. Premium rates (%) for corn, in Guarapuava County

			Empirical premium rates				Nonparametric premium rates			
			Adjusted series		Nonadjusted series		Adjusted series		Nonadjusted series	
LC	Normal rates	Beta rates	$d=30$	$d=40$	$d=30$	$d=40$	$d=30$	$d=40$	$d=30$	$d=40$
70	2.083	3.635	3.224	2.260	2.7038	5.4815	4.518	3.270	2.7993	6.4311
75	2.791	4.739	4.355	3.470	3.5492	6.8271	5.706	4.280	3.6722	7.5549
80	3.778	6.217	5.797	4.960	4.2889	8.0417	7.021	5.400	4.5412	8.6799
85	4.901	7.963	7.643	6.280	5.0048	9.1484	8.431	6.540	5.5478	9.9206
90	6.330	9.815	9.403	7.4500	6.3087	10.1873	9.906	7.7900	6.7250	11.2357

Table 3. Premium rates (%) for wheat, in Tibagi county

			Empirical premium rates				Nonparametric premium rates			
			Adjusted series		Nonadjusted series		Adjusted series		Nonadjusted series	
LC	Normal rates	Beta rates	$d=30$	$d=40$	$d=30$	$d=40$	$d=30$	$d=40$	$d=30$	$d=40$
70	1.1889	2.4382	2.9897	2.3477	2.7410	2.1906	3.4987	2.8571	3.7526	3.1213
75	1.7825	3.4711	3.7435	3.0671	3.6470	3.0301	4.3912	3.7330	4.7152	4.0751
80	2.5453	4.7907	4.6911	3.9635	4.7255	4.1162	5.4253	4.6800	5.8450	5.1020
85	3.5976	6.4816	5.7267	4.9860	6.0738	5.4861	6.5087	5.8878	7.0129	6.4051
90	4.9642	8.4299	7.1103	6.4530	7.5325	6.9222	7.8486	7.2912	8.4467	7.8854

mass in the inferior tail of the distribution and consequently increase rates.

For illustrative purpose, we will show a comparison between the premium rate charged by a private insurance company in the state of São Paulo[8] and rates calculated in this research, for corn and soybean. It is important to point out that the crop insurance offered by this company covers the credit loan and not the yield, although the indemnity is based on the reduction of the agricultural production. Further, the expected loss and premium rate are based on county yields.

For soybean (Cascavel County) and corn (Guarapuava County) rates are, respectively, 4 and 4.5%. We found in our empirical analysis (nonparametric approach) that rates for soybean are on an average 1.6 and 3.9% for the adjusted and nonadjusted series, respectively (level of coverage of 90%). Average rates for corn are equal to 8.85% and 8.95%.

One can notice the great variability on premium rates. Rates charged by this company are much higher for soybean, in the adjusted case and much lesser for corn, in both cases (adjusted and nonadjusted). In the latter case, producers are being undercharged.

VIII. Conclusion

In this research, we analysed alternative statistical methods for pricing crop insurance contracts based on area-yield with IBGE's aggregate yield data. Corn, soybean and wheat series were adjusted through parametric analysis in Cascavel, Guarapuava and Tibagi Counties. We used the normal and beta parametric distributions.

Results showed that beta rates are higher than normal rates for all levels of coverage considering the empirical rates and the nonparametric rates for corn, soybean and wheat. For these three counties, results suggest some positive skewness in the beta distribution.

Rates in the empirical and nonparametric cases are quite different in both cases: the adjusted and nonadjusted series in Tables 1, 2 and 3. Empirical rates method is commonly used by most of the crop insurance companies in Brazil. Looking more carefully at the results, for all levels of coverage, rates are higher in the nonparametric approach. It means that insurance companies are underpricing the insurance contract. The pure premium rate is actually higher than the premium rate charged. The consequence for the insurer company is the financial loss due to the

[8] Rates charged are based only on five years of observations using the empirical rates method.

lower rate charged and high-risk producers may find this situation attractable to demand the insurance contract increasing the probability to receive the indemnity.

Considering, e.g. that a insurance company sells to a soybean producer a insurance contract in the Cascavel County charging 3.5944% (nonadjusted series using the empirical rates method with distance equal 40 miles) instead of 4.23885% (nonparametric method). Suppose, for instance, the average liability is equal to US$ 1 million in a pool (20 000 producers). The average premium charged is approximately US$ 36 000 instead of US$ 42 400. The average loss is equal to US$ 6400, but the total loss is approximately US$ 128 million.

In a market where historically total loss ratio (indemnity divided by total premium) is greater than one, better actuarial methods (the nonparametric approach proposed in this article) should be taking into account by insurance companies to calculate the premium rate.

Futures researches will study the conditional yield density modelling at the farm level and rating an insurance contract based on individual agricultural yields using some of the methods discussed in this research and alternative approaches that take the spatial correlation between farms into account.

Acknowledgements

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

References

Booth, P., Chadburn, R., Cooper, D., Haberman, S. and James, D. (1999) *Modern Actuarial Theory and Practice*, Chapman and Hall, London, p. 716.

Botts, R. R. and Boles, J. N. (1958) Use of normal-curve theory in crop insurance ratemaking, *Journal of Farm Economics*, **40**, 733–40.

Casella, G. and Berger, R. L. (1990) *Statistical Inference* (Ed.) Brooks/Cole Publishing Company, California, p. 650.

Chambers, R. G. (1989) Insurability and moral hazard in agricultural insurance markets, *Americal Journal of Agricultural Economics*, **71**, 604–16.

Day, R. H. (1965) Probability distributions of field crop yields, *Journal of Farm Economics*, **47**, 713–41.

Gallagher, P. (1987) US Soybean yields: estimation and forecasting with nonsymmetric disturbances, *American Journal of Agricultural Economics*, **69**, 796–803.

Goodwin, B. K. (2001) Problems with market insurance in agriculture, *American Journal of Agricultural Economics*, **83**, 643–9.

Goodwin, B. K. and Ker, A. P. (1998) Nonparametric estimation of crop yield distributions: implications for rating group-risk crop insurance contracts, *American Journal of Agricultural Economics*, **80**, 139–53.

Goodwin, B. K. and Ker, A. P. (2002) Modeling price and yield risk, in *A Comprehensive Assessment of the Role of Risk in US Agriculture* (Eds) R. E. Just and R. D. Pope, Kluwer Academic Publisher, Massachusetts, p. 586.

Goodwin, B. K. and Mahul, O. (2004) Risk modeling concepts relating to the design and rating of agricultural insurance contracts, World Bank.

Goodwin, B. K. and Smith, V. (1996) Crop insurance, moral hazard, and agricultural chemical use, *American Journal of Agricultural Economics*, **78**, 428–38.

Guyon, X. (1995) *Random Fields on a Network: Modeling, Statistics and Applications*, Springer Verlag, New York.

Halcrow, H. G. (1949) Actuarial structures for crop insurance, *Journal of Farm Economics*, **31**, 418–43.

Hart, D. G., Buchanan, R. A. and Howe, B. A. (1996) *The Actuarial Practice of General Insurance*, The Institute of Actuaries of Australia, Australia, p. 592.

Harwood, J., Heifner, R., Coble, K., Perry, J. and Agapi, S. (1999) Managing risk in farming: concepts, research and analysis. Agricultural Economic Report 774, No. Economic Research Service, US Department of Agriculture.

Ibge (2004) Available at www.ibge.gov.br (accessed January 2004).

Just, R. E. and Weninger, Q. (1999) Are crop yields normally distributed?, *American Journal of Agricultural Economics*, **81**, 287–304.

Ker, A. P. and Coble, K. (2003) Modeling conditional yield densities, *American Journal of Agricultural Economics*, **85**, 291–304.

Ker, A. P. and Goodwin, B. K. (2000) Nonparametric estimation of crop insurance rates revisited, *American Journal of Agricultural Economics*, **83**, 463–78.

Mahul, O. (1999) Optimum area yield crop insurance, *American Journal of Agricultural Economics*, **81**, 75–82.

Miranda, M. J. (1991) Area-yield crop insurance reconsidered, *American Journal of Agricultural Economics*, **73**, 233–42.

Miranda, M. J. and Glauber, J. W. (1997) Systemic risk, reinsurance, and the failure of crop insurance markets, *American Journal of Agricultural Economics*, **79**, 206–15.

Miranda, M., Skees, J. and Hazell, P. (1999) Innovations in agricultural and natural disaster insurance for developing countries. Working paper, Dep. of Agr., Envir. and Development. Econ., The Ohio State University.

Moss, C. B. and Shonkwiler, J. S. (1993) Estimating yield distributions with a stochastic trend and nonnormal errors, *American Journal of Agricultural Economics*, **75**, 1056–62.

Nelson, C. H. and Preckel, P. V. (1989) The conditional beta distribution as a stochastic production function, *American Journal of Agricultural Economics*, **71**, 370–8.

Ozaki, V. A. (2005) Métodos atuariais aplicados à determinação da taxa de prêmio de contratos de seguro agrícola: um estudo de caso, Thesis (Ph.D), Piracicaba – Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, p. 324.

Quiggin, J., Karagiannis, G. and Stanton, J. (1994) Crop insurance and crop production: an empirical study of moral hazard and adverse selection, in *Economics of Agricultural Crop Insurance: Theory and Evidence* (Eds) D. L. Hueth and W. H. Furtan, Kluwer Academic Publishers, Boston, p. 380.

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

Ramirez, O. A., Misra, S. and Field, J. (2003) Crop-yield distributions revisited, *American Journal of Agricultural Economics*, **85**, 108–20.

Redja, G. E. (1995) *Principles of Risk Management and Insurance*, HarperCollins College Publishers, New York, p. 784.

Rosseti, L. A. (1998) Seguridade e zoneamento agrícola no Brasil. Novos rumos, in I SIMPÓSIO INTERNACIONAL DE SECURIDADE E ZONEAMENTO AGRÍCOLA DO MERCOSUL. Brasília, 1998, Anals, Ministério da Agricultura e Abastecimento/CER/PROAGRO/GM, Brasília, pp. 1–10.

Rosseti, L. A. (2001) Zoneamento agrícola em aplicações de crédito e securidade rural no Brasil: aspectos atuariais e de política agrícola, *Brazilian Journal of Agrometeorology*, **9**, 386–99.

Sherrick, B. J., Zanini, F. C., Schnitkey, G. D. and Irwin, S. H. (2004) Crop insurance valuation under alternative yield distributions, *American Journal of Agricultural Economics*, **86**, 406–19.

Silverman, B. W. (1986) *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London, p. 175.

Skees, J. R. and Barnett, B. J. (1999) Conceptual and practical considerations for sharing catastrophic risks, *Review of Agricultural Economics*, **21**, 424–41.

Skees, J. R., Black, J. R. and Barnett, B. J. (1997) Designing and rating an area yield crop insurance, *American Journal of Agricultural Economics*, **79**, 430–8.

Skess, J. R. and Reed, M. R. (1986) Rate making for farm-level crop insurance: implications for adverse selection, *American Journal of Agricultural Economics*, **68**, 653–9.

Smith, V. H., Chouinard, H. H. and Baquet, A. E. (1994) Almost ideal area yield crop insurance contracts, *Agricultural and Resource Economics Review*, **23**, 75–83.

Taylor, C. R. (1990) Two practical procedures for estimating multivariate nonNormal probability density functions, *American Journal of Agricultural Economics*, **72**, 210–7.

Trowbridge, C. L. (1989) *Fundamental Concepts of Actuarial Science*, Actuarial Education and Research Fund, p. 79.

Turvey, C. and Zhao, C. (1999) 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, Ontario.

Wang, H. W. and Zhang, H. (2003) On the possibility of private crop insurance market: a spatial statistics approach, *The Journal of Risk and Insurance*, **70**, 111–24.