Journal of Applied Statistics

ISSN: 0266-4763 (Print) 1360-0532 (Online) Journal homepage: http://www.tandfonline.com/loi/cjas20

Modeling of soybean yield using symmetric, asymmetric and bimodal distributions: implications for crop insurance

Gislaine V. Duarte, Altemir Braga, Daniel L. Miquelluti & Vitor A. Ozaki

To cite this article: Gislaine V. Duarte, Altemir Braga, Daniel L. Miquelluti & Vitor A. Ozaki (2017): Modeling of soybean yield using symmetric, asymmetric and bimodal distributions: implications for crop insurance, Journal of Applied Statistics, DOI: 10.1080/02664763.2017.1406902

To link to this article: https://doi.org/10.1080/02664763.2017.1406902

Published online: 28 Nov 2017.

JOURNAL OF APPLIED STATISTICS, 2017
https://doi.org/10.1080/02664763.2017.1406902

Modeling of soybean yield using symmetric, asymmetric and bimodal distributions: implications for crop insurance

Gislaine V. Duarte [a], Altemir Braga[a], Daniel L. Miquelluti [b] and Vitor A. Ozaki[b]

[a]Department of Agronomic Statistic and Experimentation, School of Agriculture 'Luiz de Queiroz', University of São Paulo, Piracicaba, São Paulo, Brazil; [b]Department of Economic, Management and Sociology, School of Agriculture 'Luiz de Queiroz', University of São Paulo, Piracicaba, São Paulo, Brazil

ABSTRACT

Over the years, many papers used parametric distributions to model crop yields, such as: normal (N), Beta, Log-normal and the Skew-normal (SN). These models are well-defined, mathematically and also computationally, but its do not incorporate bimodality. Therefore, it is necessary to study distributions which are more flexible in modeling, since most of crop yield data in Brazil presents evidence of asymmetry or bimodality. Thus, the aim of this study was to model and forecast soybean yields for municipalities in the State of Paran, in the period from 1980 to 2014, using the Odd log normal logistic (OLLN) distribution for the bimodal data and the Beta, SN and Skew-*t* distributions for the symmetrical and asymmetrical series. The OLLN model was the one which best fit the data. The results were discussed in the context of crop insurance pricing.

ARTICLE HISTORY
Received 8 March 2017
Accepted 14 November 2017

KEYWORDS
Asymmetry; bimodality distribution; crop insurance; heavier tails; risk management

1. Introduction

Agriculture is one of the oldest and most important activities to the human being, because it is a source of raw materials, food and energy. However, agriculture presents great risks in production on the basis of extreme weather events. Over the years, several tools for managing risk were developed to minimize the impact of weather on agricultural production, among them, the crop insurance [24]. In addition to minimize the risks of loss in agriculture, the crop insurance provides the recovery of the financial capacity of the producer in the event of damage caused by uncontrollable natural events.

However, crop insurance suffers from problems such as information asymmetry, thus causing adverse selection,[1] moral hazard,[2] the lack of appropriate methodologies for pricing and the absence of long historical series discourage the offer by the insurers [10,21,42,48].

In this way, it is common to have the presence of the State developing actions that seek to compensate for these deficiencies, such as premium subsidies. However, the absence of an actuarial methodology that accurate calculates the premium rate, and the lack of continuity

CONTACT Gislaine V. Duarte gislainevduarte@gmail.com Departament of Agronomic Statistic and Experimentation, School of Agriculture 'Luiz de Queiroz', University of São Paulo, Avenue Pádua Dias, 11, Piracicaba, São Paulo, Brazil

in the subsidy volume by the government, are pointed out as some of the main problems for the crop insurance market development.

Therefore, this paper aims to propose alternative methods for crop insurance pricing based on parametric distributions that capture the skewness or bimodality of the data. Considering Brazilian yield data, to obtain the actuarially fair and more accurate premium, such characteristics must be taken into account. We compare the rates calculated by the new methodology with the commercial fees charged by Brazilian insurance companies and discuss their implications for the market.

The rest of the paper is organized as follows. In Section 2 provides a background on the modeling of yield. In Section 3 is presented an introduction to agricultural insurance and the insurance premium calculation forms. The presents the description of the data used in the research in the Section 4. The methodology used to model the agricultural yield series is presented in the Section 5. In the Section 6 presents the results. In the Section 7 presents the discussions. Finally, in Section 8, the concluding remarks are addressed.

2. Background

In order to obtain the actuarially fair premium for crop insurance one needs to correctly model crop yields. The literature presents three approaches to this issue: the first involves the estimation of the parameters of a parametric distribution; the second is related to a wide variety of models using non-parametric distributions; and the third is composed by semi-parametric methods, wich combines elements of the first and second alternatives [28].

According to Goodwin and Mahul [23] non-parametric methods, despite being more flexible, require large samples, and also present difficulties in measuring rare or catastrophic events. Even so, they are the most used in studies with small samples. In the non-parametric methods, Goodwin and Ker [22], Turvey and Zhao [51], Ker and Goodwin [29], Ozaki *et al.* [39] applied the kernel estimator for the density function of crop yield and calculated the insurance premium rates.

Of particular note in the parametric methods, there are the studies of Just and Weninger [27], and Botts and Boles [7] who used the N distribution to estimate yield.

In the study of Just and Weninger [27] they studied evidence of non-normality on data of agricultural productivity in Kansas, in the United States. The authors identified three methodological common problems in the analysis of crop yield distribution: specification errors of non-random components in crop yield distribution, poor presentation and analysis of statistical significance and use of aggregated data to represent yield distributions at the farm level.

Even for Just and Weninger [27], the presence of one or more of these problems affect any evidence against normality. In addition, the use of the average yield of municipalities (aggregate data), removes the specific information of farms and emphasizes the temporal variability of producers from the region studied, supressing the farm variability.

Other studies such as Day [17], Taylor [49], Ramírez [44], Ramirez *et al.* [45] and Sherrick *et al.* [47] found evidence against normality, suggesting that the data is not symmetrical when compared to the average.

In addition to the N distribution, other parametric distributions were used to model crop yields such as: the gamma distribution in Gallagher [20], the inverse hyperbolic distribution in Moss and Shonkwiler [35], the Beta and Weibull distributions in Sherrick *et al*. [47] and the skew-normal (SN) distribution in Ozaki and Silva [38].

Among these models the most used is the Beta distribution. It is the study of Nelson and Preckel [36]; Babcock and Hennessy [6]; Coble *et al*. [11]; Hennessy *et al*. [25] that found strong evidence of skewness or kurtosis.

The study of Ramirez *et al*. [45] confirmed the evidences of non-normality and asymmetry of agricultural productivity found in Ramírez [44]. The productivity of corn and soybean was modeled in the 'Corn Belt' region and cotton in Texas, United States. The authors also tested for nonlinear tendency, heteroscedasticity, kurtosis and skewness. An expansion of Johnson SU distribution is used in modeling because of its flexibility. The tests for nonlinear trend and heteroscedasticity are conducted at the same time allowing the possibility of non-normal distributions using the additional information transmitted through a correlation matrix. It is concluded that different patterns of non-normal distributions could result in different factors that affect productivity at farm and municipalities level, depending on the crop, cultivation system and geographic region. The main recommendation of the study is that the researchers should recognize and consider in any risk analysis, being that the political, market, industry or farm, the application of non-normal distributions in modeling productivity.

In the study of Sherrick *et al*. [47], the productivity of corn and soybean was investigated on 26 farms in the United States, using the N, Logistic, Weibull, Beta and Lognormal distributions. In this study they calculated and compared the payments to agricultural crop insurance products for each of the parametric distributions within each farm and between them, which showed that the choice of distribution has a significant impact on the risk assessment of productivity and on the expected value of the insurance payment.

Bayesian hierarchical models were also applied to the modeling of agricultural productivity as shown by Ozaki and Silva [38]. This study proposed an alternative formula to calculate the insurance premium rate, on corn crops for the municipalities of the State of Paraná, through Bayesian hierarchical space-temporal models using the SN distribution. In this model the time trend was simultaneously modeled with the spatial correlation between the municipalities. Premium rates were calculated directly from the predicted distribution by Monte Carlo simulation through Markov Chain. The authors concluded that the proposed methodology improves the actuarial and statistical procedures frequently used in the calculation of premium rates, specially when working with a reduced number of observations. The authors concluded that the empirical rates,[3] commonly used by Brazilian insurers, overestimate the risk of producers in areas of low risk and underestimate the risk of ones with high risk.

Therefore, based on the studies we cited it is clear that short series of productivity limit the use of nonparametric techniques, however these reduce the errors associated with incorrect assumptions related to the distributions. The limitation of crop yield series and the results of several studies suggest using parametric distributions in productivity modeling.

In Brazil, there is no public database with information of agricultural productivity at farm level, only at the municipal level. In addition, agricultural productivity in Brazil has specific characteristics, such as asymmetry and bimodality. Therefore, for a more accurate calculation of the insurance premium rate, it is important to consider distributions that capture these characteristics, which has not yet been considered in existing literature, by means of a parametric approach.

In the study of da Silva Braga *et al.* [16] they propose a new extended N distribution with heavier tails called the Odd log normal logistic (OLLN) model. The distribution is symmetric, platykurtic, leptokurtic and may be unimodal or bimodal. In this paper is provided three applications to real data to prove empirically the flexibility of the OLLN distribution: the first application involves the temperature (°C) and overall daily radiation ($\text{cm}^{-2}\text{d}^{-1}$) variables correspond to daily data; the second data set refers to the experiment carried out to assess the effects of doses of an anthelmintic compound (ml) to control a parasite (fixed effects) using a CRD with five treatments and the data from the third application refer to weight gain of livestock. The purpose of this application is to fit a linear RCBD model using the OLLN distribution. In recent studies, Cruz *et al.* [13] obtained the Bivariate odd-log-logistic-Weibull regression model for oral health-related quality of life; da Silva Braga *et al.* [15] obtained the A new skew-bimodal distribution with applications and Alizadeh *et al.* [5] obtained the the odd log-logistic logarithmic generated family of distributions with applications in different areas.

3. Crop insurance

The agricultural insurance is an important risk transfer mechanism, however it also has other benefits, among them, the economic stability of producers and the development and adoption of new technologies. All of this has contributed to the decrease the rural exodus and reduce the producer loan debt in case of losses [9].

In this type of contract, the producer shall be indemnified if the observed yield Y , after the harvest, falls below the guaranteed yield Y_g in the contract. The guaranteed yield is defined as $Y_g = \lambda Y_e$, where λ is the coverage level (CL) chosen by the producer, $0 < \lambda < 1$, Y_e is the expected yield. In fact, the expected yield Y_e is the arithmetic average of the past five years.

The theoretical concepts related to the determination of the yield protection crop insurance premium rates were described in the study of Lawas [30] and Miqueleto [34]. In accordance with Ozaki [37], the premium is considered actuarially fair when the probability of the event of an accident is equal to the premium per unit of compensation, or when the premium is equal to the expected indemnity payment'. Thus, the fair premium rate is given by the expectation of loss, or expected loss, which is given by:

$$E(\text{loss}) = E(\max\{(y_g - Y); 0\}) = P(Y < y_g)(\lambda y_e - E(Y|Y < \lambda y_e)). \tag{1}$$

According to Goodwin and Ker [22], Goodwin and Mahul [23] and Ker and Goodwin [29], the fair premium rate of a yield protection insurance contract is given by:

$$\text{Premium rate} = \frac{F_Y(\lambda y_e)[\lambda y_e - E(Y|Y < \lambda y_e)]}{\lambda y_e}, \tag{2}$$

where the $F_Y(\lambda y_e)$ is the cdf of the variable Y, named 'probability of loss of the producer'. In this context, we denote the importance of estimating the yield pdf correctly, since inaccurate inferences could change the premium rate results.

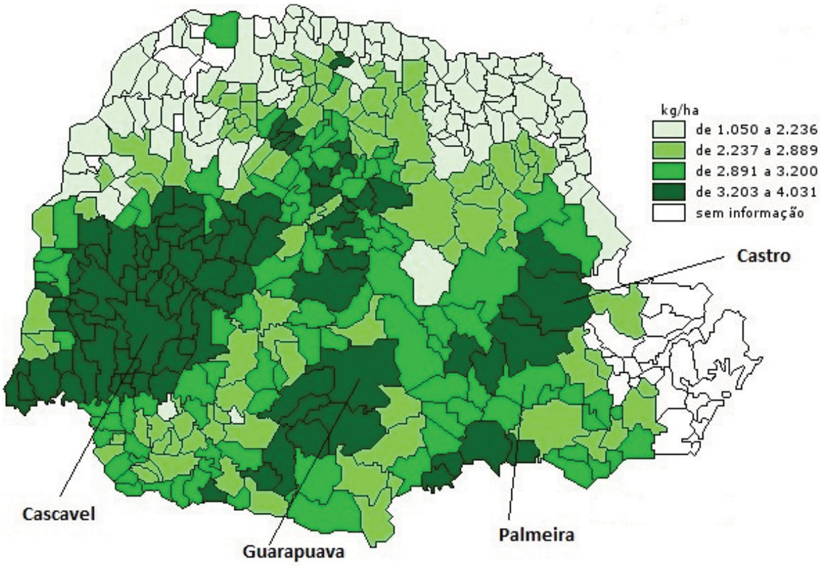


Figure 1. Soybean yield in Paraná in 2014.

4. Description of data

For the modeling of agricultural productivity, the annual soybean productivity data (kg/ha) of some important municipalities in Paraná, Brazil, was obtained from the Economic and Social Development Institute of Paraná [26]. The available observations are related to the 1979/1980 through 2013/2014 agricultural years, making a total of 35 observations. Figure 1 shows the soybean agricultural productivity in Paran for 2014, according to IPARDES.

The selected municipalities are Cascavel, Guarapuava, Castro and Palmeira. These municipalities, besides being large producers of soybean, were also selected because they are the State leaders in soybean planted area, in accordance with the Bulletin of Agricultural Monitoring – Conab – summer-season crops 2015/2016 [18]. The commercial insurance premium rates applied to the selected municipalities were provided by the Ministry of Agriculture, Livestock and Supply [33].

5. Methodology

In this section we will present the parametric models used in the modeling of productivity and the measures of adjustment for choosing the best model. The appendix presents the method of estimation by maximum likelihood applied to models used in this study. All analysis are conducted using package 'Optim' and 'AdequacyModel' of the R software [43].

The well-known N and Beta models were used as a first approach to modeling yields. As yields may present an asymmetric behavior the SN and ST models were also considered.

5.1. *Odd log-logistic – F distributions*

A recently introduced family of continuous distributions is the *odd log-logistic-F* (OLLF). This class is well defined and some of its mathematics properties have been demonstrated, such as ordinary and incomplete moments, quantiles functions, order statistics and two types of entropies. The inference and estimation by the maximum likelihood method is also covered for censored survivability data [14].

Let a cumulative probability function (cdf) of any baseline function $F(x;\xi)$ with a parameter vector ξ, the cdf of the OLLF, with a shape parameter $\alpha > 0$, is defined by

$$G(x,\xi) = \int_0^{F(x,\xi)/\bar{F}(x,\xi)} \frac{\alpha t^{\alpha-1}}{(1+t^\alpha)^2}\,\mathrm{d}t = \frac{F(x,\xi)^\alpha}{F(x,\xi)^\alpha + \bar{F}(x,\xi)^\alpha}, \tag{3}$$

where $\bar{F}(x,\xi) = 1 - F(x,\xi)$ and $\alpha = \log(G(x;\xi)/\bar{G}(x;\xi))/\log(F(x;\xi)/\bar{F}(x;\xi))$. The α parameter represents the log quotient of the odds ratio for the baseline distribution F. Multiple distributions may be generated from the above equation as described in the study of Alizadeh *et al.* [4], Cruz *et al.* [12] and da Silva Braga *et al.* [16].

The probability density function (pdf) of the new family is defined as:

$$g(x,\xi) = \frac{\alpha f(x;\boldsymbol{\xi})\{F(x;\boldsymbol{\xi})[1-F(x;\boldsymbol{\xi})]\}^{\alpha-1}}{\{F(x;\boldsymbol{\xi})^\alpha + [1-F(x;\boldsymbol{\xi})]^\alpha\}^2}. \tag{4}$$

The new OLLF distribution family allows a greater flexibility of the distribution's tails. If we consider $F(x;\boldsymbol{\xi})/\bar{F}(x;\boldsymbol{\xi}) = \Phi(x;\mu,\sigma)/\bar{\Phi}(x;\mu,\sigma)$, in Equation (3), the cdf of the OLLN model with an additional shape parameter $\alpha > 0$ is defined by

$$F(x;\mu,\sigma,\alpha) = \int_0^{\Phi(x;\mu,\sigma)/\bar{\Phi}(x;\mu,\sigma)} \frac{\alpha\, t^{\alpha-1}}{(1+t^\alpha)^2}\,\mathrm{d}t = \frac{\Phi^\alpha(\frac{x-\mu}{\sigma})}{\Phi^\alpha(\frac{x-\mu}{\sigma}) + [1-\Phi(\frac{x-\mu}{\sigma})]^\alpha}, \tag{5}$$

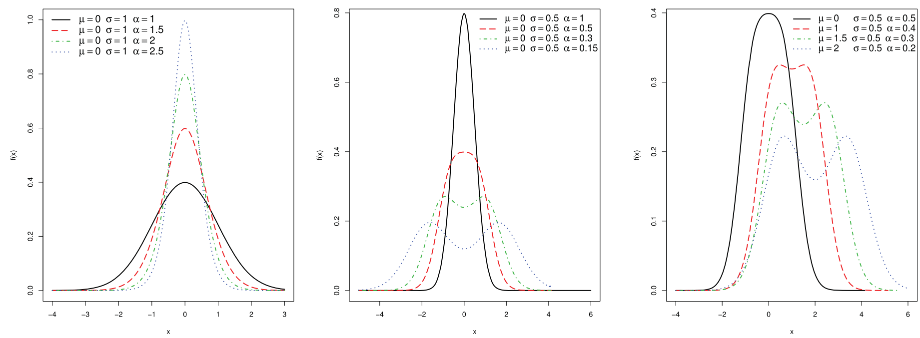
where $\bar{\Phi}(x;\mu,\sigma) = 1 - \Phi(x;\mu,\sigma)$. The OLLN density is given by

$$f(x;\mu,\sigma,\alpha) = \frac{\alpha\phi(\frac{x-\mu}{\sigma})\Phi(\frac{x-\mu}{\sigma})[1-\Phi(\frac{x-\mu}{\sigma})]^{\alpha-1}}{\sigma\{\Phi^\alpha(\frac{x-\mu}{\sigma}) + [1-\Phi(\frac{x-\mu}{\sigma})]^\alpha\}^2}, \tag{6}$$

respectively. Note that $\alpha > 0$ is a shape parameter. Henceforth, a random variable with density function (6) as above is denoted by $X \sim \text{OLLN}(\alpha,\mu,\sigma)$. For $\mu = 0$ and $\sigma = 1$, we obtain the standard OLLN distribution. Further, the OLLN distribution with $\alpha = 1$ reduces to the N distribution.

Note that the new model given by Equation (6) has only one extra parameter and two parameters of the N distribution. This compares favorably with other families, for example, the Generalized Beta family (B-G) proposed by Eugene *et al.* [19], which includes two extra shape parameters and its cdf depends on the incomplete beta function; the Kummer Beta generalized family (KB-G) [41], that involves three extra parameters and its cdf depends on the confluent hypergeometric function; the generalized McDonald class (Mc-G) [3], which includes three extra parameters and its cdf also depends on the incomplete beta function.

Figure 2 shows the density of the OLLN distribution for some values of the parameters μ,σ and α. It should be noted in the plots Figure 2(a, b) the contribution of the parameter α on the unimodality and bimodality of the distribution, when μ and σ are fixed. When the parameter α approaches zero, the pdf presents bimodality. On the other hand, when the value of α increases the function presents unimodality. It is observed that when μ varies, the plots is translated in the x-axis, regardless of the form Figure 2(c).



(a) Varying $\alpha > 0$ (b) Varying $\alpha < 0$ (c) Varying $\mu > 0$ e $0.2 < \alpha < 0.5$

Figure 2. Plots of the OLLN density function for different values of the parameters. (a) Varying $\alpha > 0$, (b) varying $\alpha < 0$, and (c) varying $\mu > 0$ e $0.2 < \alpha < 0.5$.

5.2. Model assessment

For the selection of the model that best fits the data, it is used some criteria or statistical tests. The most used criteria for model selection in practice are the Akaike's information criterion (Akaike AIC), proposed by Akaike [1] and the Bayes information criterion (BIC), Schwarz [46]. These criteria are based on the logarithm of the estimated likelihood function. The Akaike's criterion penalizes models with large number of parameters k, since it is expected that the logarithm of likelihood function grows with the number of parameters added to the model.

Several corrections of this criterion have been proposed, in order to reduce the probability of choice of models with greater order than desired. Among them, there is the corrected AICc, proposed by Akaike [2].

A different approach for choosing the best model is through the study of modified statistics, from Anderson–Darling (W^*) and Cramer–von Mises (A^*), proposed by Lin *et al.* [31] and Pakyari and Balakrishnan [40], respectively.

These statistics were calculated as follows: given the maximum likelihood estimator (MLE) $\hat{\theta}$, for a n-dimensional vector θ from the observed Type-II right censored sample $x_1 < \cdots < x_n$. Then, calculate $v_i = F(x_i;\theta)$, and convert the censored sample v_i to a complete sample of size r, $v_1^*, \ldots, v_r^*$. Then, compute $y_i = \Phi(v_i)^{-1}$, $\Phi(\cdot)^{-1}$ is the inverse cdf of the standard N distribution.

The model that presents the lowest value of these statistics should be chosen.

$$W^2 = \sum_{i=1}^{n} \left(\mu_i - \frac{(2i-1)}{2n} \right)^2 + \frac{1}{12n}, \quad W^* = W^2(1 + 0.5/n), \tag{7}$$

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^{n} ((2i-1)\log \mu_i + (2n+1-2i)\log(1-\mu_i)), \tag{8}$$

$$A^* = A^2(1 + 0.75/n + 2.25/n^2), \tag{9}$$

where $\mu_i = \Phi((y_i - \bar{y})/s_y)$, $\bar{y} = \sum_{i=1}^{n} y_i/n$, $s_y^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2/(n-1)$.

5.3. Crop yield trend

In the period of study (1980–2014), there was a great advance in the technologies employed in crops, such as planting techniques, inputs, machinery, among others, which suggests a trend in the data. Moreover, it can be expected time dependence and not constant variance over time. Thus, before adjusting any probabilistic model for the series of productivity, it is necessary to use statistical techniques that result in a series with no trend, independent and homoscedastic.

The range correction related to the trend is achieved by the same approach used in [20,23,50]. This procedure initially estimates a linear deterministic model of productivity and time, given by:

$$y_t = \alpha + \beta T + e_t, \quad e_t \sim N(\mu, \sigma), \tag{10}$$

where y_t is the productivity vector, T is the time vector, α and β the parameters of regression. It proceeds to take up the residuals $\hat{e}_t$ of the regression, the estimate of the last observation through the fitted model $\hat{y}_{2014}$ and removes the tendency according to the equation:

$$\tilde{y}_t = \hat{y}_{2014}\left(1 + \frac{\hat{e}_t}{\hat{y}_t}\right). \tag{11}$$

For the verification of temporal dependence is applied the test proposed by Ljung and Box [32], in which the null hypothesis is that there is independence in the series. In addition, to verify homocedasticy, it was applied the Breusch and Pagan [8] test, in which the null hypothesis is that the series is homoscedastic.

6. Results

Regarding the temporal dependence test for the municipalities of Cascavel (p-value $= 0.7416$), Guarapuava (p-value $= 0.1013$), Castro (p-value $= 0.1004$) and Palmeira (p-value $= 0.2597$) we conclude that there is no temporal dependence in the series.

Observing the p-values of the homocedasticy test for the municipalities of Cascavel (p-value $= 0.624$), Guarapuava (p-value $= 0.1785$), Castro (p-value $= 0.2412$) and Palmeira (p-value $= 0.0437$) we conclude the series are homoscedastic.

Table 1 presents some descriptive statistics for the selected series, corrected for tendency, temporal dependence and heteroscedasticity. Note that for all municipalities the coefficient of skewness is negative and the median is greater than the average, with the exception of Palmeira, suggesting a negative skewness. In addition, the municipality of Castro presents a leptokurtic curve. The municipalities of Guarapuava and Castro have a lower coefficient of variation when comparing with other municipalities, suggesting a lower production risk for these municipalities. For all municipalities the expected yield is more than 3300.00 kg/ha.

The parameter estimates for the parametric distributions were obtained by maximum likelihood. Details are presented in the appendix. Figure 3 shows the original and corrected productivity series (without trend). We note that the series without trend shows similar behavior to the original series.

Table 1. Descriptive statistics for the corrected series.

	Municipalities			
	Cascavel	Guarapuava	Castro	Palmeira
Average	3268.108	3380.111	3526.370	3244.340
Median	3310.269	3411.745	3547.058	3210.203
Standard deviation	438.727	282.186	231.310	328.003
Asymmetry	−0.075	−0.262	−0.920	−0.369
Kurtosis	2.908	2.440	5.431	2.219
Maximum	4277.174	3925.333	4014.058	3701.117
Minimum	2280.380	2703.796	2742.989	2473.699
Expected productivity	3403.802	3589.257	3589.257	3337.692
Coefficient of variation	0.134	0.083	0.066	0.101

Table 2. Statistics and information criteria for model selection.

Municá-pios	Modelo	A^*	W^*	AICc	BIC
	N	0.2190	0.0354	528.5575	531.2932
	SN	0.2190	0.0354	530.9567	534.8486
Cascavel	**OLLN**	0.2107	0.033	530.932	534.8240
	Beta	0.2220	0.0360	533.462	538.3502
	ST	0.8448	0.1066	534.2515	538.4039
	N	0.303	0.050	497.665	500.4015
	SN	0.2459	0.037	499.414	503.3063
Guarapuava	**OLLN**	0.1899	0.0254	498.2860	502.1779
	Beta	0.1950	0.026	500.8035	505.6915
	ST	0.8183	0.0893	503.3585	507.5109
	N	0.628	0.0999	483.7492	486.4849
	SN	0.4334	0.0646	483.5584	487.4503
Castro	**OLLN**	0.338	0.0557	482.6576	486.5494
	ST	0.8002	0.0948	484.1971	488.3496
	N	0.5713	0.0873	508.1976	510.9333
	SN	0.5713	0.0873	510.5968	514.4887
Palmeira	**OLLN**	0.3481	0.0397	506.3232	510.2150
	ST	0.9497	0.1078	513.8916	518.0440

Figure 4 shows the densities adjusted by the parametric models. For Castro and Palmeira municipalities there was no convergence for the Beta model, thus its not present in the figures. Also, the SN and ST models overlap the N model for Palmeira. Note that in series with bimodality, as Guarapuava and Palmeira, the OLLN distribution captures this characteristic. Therefore, a major breakthrough for parametric distributions studied in yield modeling applied to crop insurance.

In Table 2 are presented the AIC and BIC information criteria and the statistics A^* and W^*. It is observed that for all municipalities, the model that best fits the data is the OLLN model. Therefore, the rates adjusted by the OLLN model are more accurate than the rates adjusted by the other distributions.

After choosing the distribution that best fits the data, the next step is to calculate the premium rates for each municipality, according to Equation (2), and compare the results with those estimated by the N distribution, commonly used by the insurance market.

7. Discussions

Table 3[4] shows the actuarially fair rates (pure fees) of the calculated premium, using different distributions for the city of Cascavel. It is observed that for coverage levels ranging

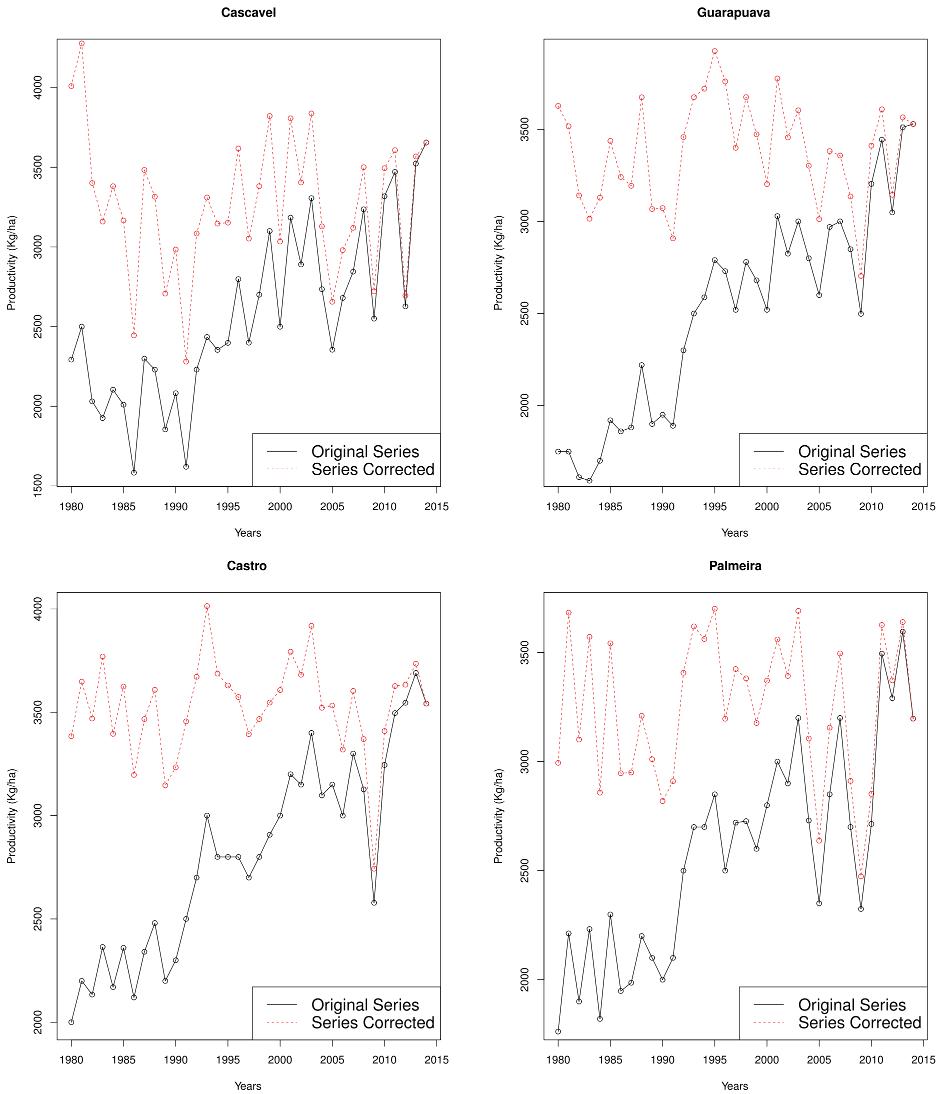


Figure 3. Original and corrected productivity series.

from 55% to 70%, the OLLN distribution overestimates the premium rate, compared to the N distribution, which is widely used by the insurance market. Equivalently, the N model underestimate the insurance rate. The implication for the insurer is reduce the total volume of premium.

However, for the coverage levels ranging from 75% and 80%, the OLLN distribution underestimates the premium rate compared to the N distribution. That is, the premium rate calculated considering the N distribution overestimates the OLLN-rate.The overpricing may hamper the massification of the insurance, and attract producers with higher risk profile, increasing the problem of adverse selection.

This analysis was also conducted for other municipalities, but we chose not to present the results as the primary objective is to compare the OLLN model comercial rates with

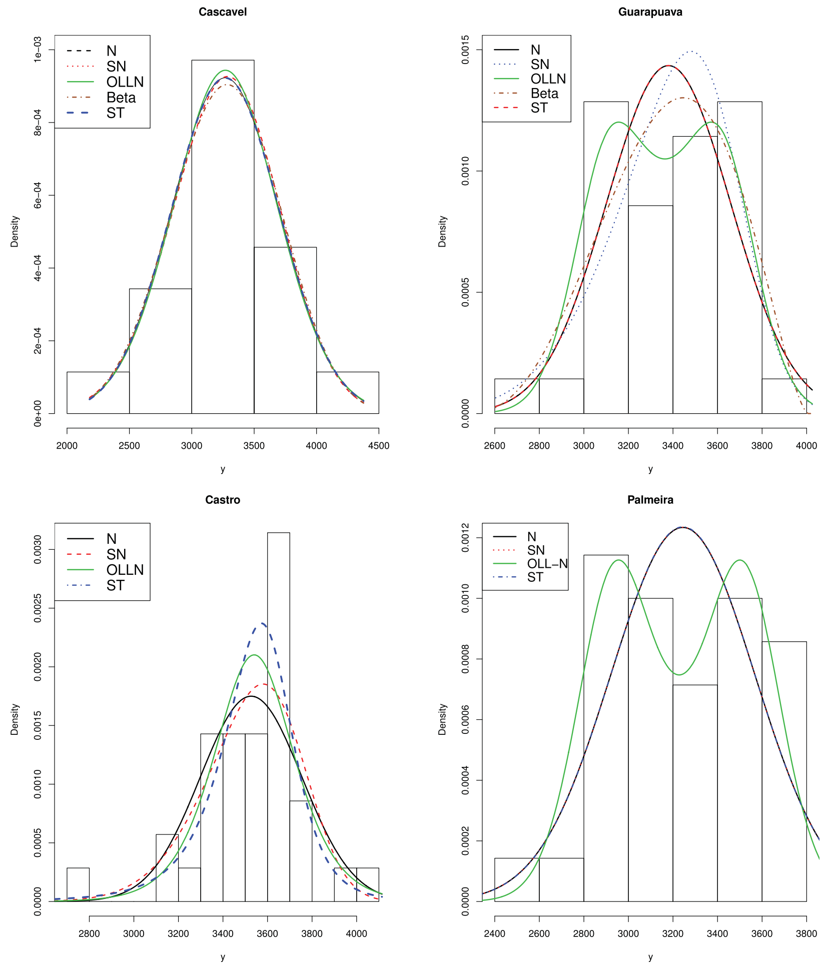


Figure 4. Adjusted distribution for the corrected series.

Table 3. Pure rate (in %) – Cascavel.

LC	N	SN	OLLN	Beta	ST
55%	0.007	0.007	0.008	0.003	0.006
60%	0.041	0.042	0.047	0.030	0.041
65%	0.178	0.181	0.189	0.154	0.176
70%	0.602	0.611	0.610	0.561	0.599
75%	1.686	1.705	1.653	1.610	1.678
80%	3.991	4.029	3.858	3.822	3.976

the other models comercial rates. Commercial rates for all municipalities are shown in Table 4.

The pure rate (PR) represents the intrinsic business risk, without including any additional costs. To compare it with the rates offered by insurance companies, we need to include average parameters of the market, regarding technical margin, administrative and business expenses and the insurer's profit margin.

Table 4. Commercial rate (in %) with a safety margin of 20% for the municipalities of Cascavel, Guarapuava, Castro and Palmeira.

Cascavel	LC	N	SN	OLLN	Beta	ST	A	B	C
	55%	0.014	0.014	0.016	0.006	0.012	7.320	6.290	7.330
	60%	0.082	0.084	0.094	0.060	0.082		6.660	
	65%	0.356	0.362	0.378	0.308	0.352			10.530
	70%	1.204	1.222	1.220	1.122	1.198			
	75%	3.372	3.410	3.306	3.220	3.356			
	79%	6.802	6.870	6.586	6.518	6.776			
	80%	7.982	8.058	7.716	7.644	7.952			
Guarapuava	LC	N	SN	OLLN	Beta	ST	B		
	65%	0.002	0.012	0.000	0.000	0.000	6.700		
	70%	0.020	0.080	0.000	0.000	0.010	6.480		
	75%	0.188	0.386	0.016	0.030	0.104			
	80%	1.200	1.458	0.354	0.784	0.738			
	85%	5.340	4.496	3.502	4.282	3.532			
	87%	8.808	6.704	6.962	7.004	5.964			
Castro	LC	N	SN	OLLN	ST	B	D		
	55%	0.000	0.000	0.000	0.000	6.400			
	60%	0.000	0.000	0.000	0.026	6.440			
	65%	0.000	0.000	0.002	0.076		5.240		
	70%	0.000	0.008	0.012	0.178	6.580			
	85%	1.474	2.668	1.492	2.228				
	87%	3.158	4.744	2.764	3.314				
Palmeira	LC	N	SN	OLLN	ST	B			
	70%	0.100	0.102	0.008	0.096	6.620			
	80%	2.268	2.316	1.668	2.218				
	83%	4.628	4.718	4.938	4.544				
	85%	7.074	7.200	8.854	6.958				
	87%	10.394	10.568	14.196	10.238				
	90%	17.242	17.504	23.774	17.018				

According to [9], some Brazilian insurers add a safety margin of 10–25% to pure rates, after the inclusion of administrative expenses, trade and profit margin, which correspond to an average of 10%, and 20%, respectively. Thus, in this study, we chose to use a loading factor of 40%, which means to divide the pure rate with the safety margin by 0.6. The market parameters adopted are 20% for the safety margin, 10% for business expenses, 20% for administrative expenses and 10% for profit margin.

The commercial rates (CR) with a 20% safety margin were obtained as follows:

$$\mathrm{CR} = \frac{\mathrm{TP} * 1.2}{[1-(0.1+0.2+0.1)]} = \frac{\mathrm{TP} * 1.2}{0.6}. \tag{12}$$

The pricing procedure used by each insurer is not publicly available, however, it is known that most insurers use the N distribution for crop yield modeling. Insurers also add a safety margin based on loss ratio historical data for the municipalities.

Table 4 presents the commercial rates with a safety margin of 20% for each model and the commercial rates charged by insurance companies A, B, C and D for the municipalities of Cascavel, Guarapuava, Castro and Palmeira.

When comparing the rates calculated in this study with those applied by insurers, it is noted that these are much below those practiced by the market. As an example, the commercial rate calculated by the OLLN distribution (Table 4) to Cascavel, with a coverage level (CL) of 55%, corresponds to 0.25% and 0.22% of the insurers B and C commercial rates, respectively. With a coverage level of 65%, the commercial fee calculated by the OLLN distribution corresponds to 3.58% of the insurer C commercial rate. Note in Table 4 that the rate of 6.66% offered by the insurer B, with a coverage level of 60%, represents approximately the commercial rate calculated by the model OLLN, with a 79% coverage level. Hence, insurers offer an inferior product with a high premium rate, making it difficult to massify insurance in Brazil.

For Guarapuava, the rates calculated by the N distribution are overpriced when compared to other distributions. Also, it is noted that the 6.70% rate offered by the insurance company B with 65% of coverage level represents, approximately, the commercial rate calculated by OLLN model with 87% coverage level. In Castro there is also one overpriced rate estimated by the N distribution, when compared to the OLLN distribution, for the 87% coverage level.

On the other hand, the municipality of Palmeira offers greater variability and, consequently, higher risks and rates. For coverage levels from 70% to 80% the N distribution overestimates the premium rate when compared to the OLLN distribution. However, to 83% or more coverage levels there is an underestimation of the rate. Moreover, the rate calculated by the OLLN distribution on the 70% coverage level equals to 0.12% of the rate used by the insurer B. This is evidence that Brazilian insurers use high rates on the yield protection product, when compared with the rates calculated in this study.

This detachment of the rates can be explained by the fact that the insurance is not mass marketed in Brazil, besides it is concentrated mainly in the southern region of the country, a region of higher occurrence of atypical climatic events. Thus resulting in a concentration of risk to the insurer who, consequently, needs to make an additional loading fee at the rate of pure risk, besides the natural loading fee, if insurance were massified in Brazil.

The differences in the premium rates can also be explained by the fact that it is assumed that all producers of this municipality bought insurance, that is , it is assumed that the insurance is massively sold in the city, causing a lower rate. In this study we used the average of municipal yields, reducing the individual variability as shown in the study of Just and Weninger [27].

8. Concluding remarks

In this study alternative probability distributions were evaluated for modeling soybean yields and estimate yield protection insurance premium rates. The series were adjusted considering the N, Beta, SN, OLLN and ST distributions for municipalities in the State of Paraná, Brazil.

We concluded that, for all series, the best fitted model was the OLLN model. Furthermore, the bimodality found in the municipalities of Guarapuava and Palmeira was considered in the calculation of the premium rate.

For the city of Cascavel the premium rate estimated by the N model is underestimated when compared to the OLLN model, with coverage levels ranging from 55% to 80%. However, for coverage levels from 83% to 90%, the N model overestimates the rate when compared to the OLLN model.

For the city of Guarapuava there is an overestimation of the rates calculated by the N model when compared to the OLLN model. As for the municipality of Palmeira, the premium rate is overpriced for the coverage levels from 79% to 80% and underestimated for levels of coverage higher than 83%. Finally, the premium rate calculated by the N model for the municipality of Castro is overestimated for coverage levels of 87% or more.

The underestimation of the rate may lead to serious damage to the insurer, because it offers a product that takes into account a smaller risk of yield than the one that should be taken. Moreover, overestimation of the premium rate hampers the massive sale of insurance in Brazil, as well as attract farmers with higher risks, strengthening the problem of adverse selection.

A large part of the distancing of the rate estimated in this study with those applied by insurers may be explained by the fact that insurance is not spread troughtout Brazil and by the delay of government subsidy payments to the insurers, which can encourage these to increase the rates thus making a provision of this cost. There also may be an excessive profit by the insurers.

Historically, in the Brazilian market, the crop insurance has been avoided by producers because they have high premium rates and insurers have avoided offering products of agricultural insurance, considering the high probability of receiving claims. Therefore, more precise statistical methods should be taken into consideration by insurers to better reflect the risk and improve the calculation of premium rates. For future research, we intend to study with regression models considering climatic variables to model the yields.

Notes

1. According to Quiggin *et al.* [42], 'Adverse selection means that people who are more likely to present claims will be more willing to insure at a given rate'.
2. Moral hazard refers to the fact that the insured may take certain actions which the insurer is unable to monitor, leading to an increase in production risk. For example, after buying insurance the producer may use less fertilizers or pesticides, causing the yield to decline [42].
3. According to Ozaki and Silva [38] these rates are based on the relationship between the average loss of the insured and this method does not take into account robust statistical analysis.
4. We present several coverage levels in order to compare with the equivalent coverage level in our model

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

We thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for financial support.

ORCID

Gislaine V. Duarte http://orcid.org/0000-0003-4372-7499.
Daniel L. Miquelluti http://orcid.org/0000-0002-6335-3618.

References

[1] H. Akaike, *A new look at the statistical model identification*, IEEE Trans. Autom. Control 19 (1974), pp. 716–723.

[2] H. Akaike, *Information theory and an extension of the maximum likelihood principle*, in *Selected Papers of Hirotugu Akaike*, Springer, New York, NY, 1998, pp. 199–213.
[3] C. Alexander, G.M. Cordeiro, E.M. Ortega, and J.M. Sarabia, *Generalized beta-generated distributions*, Comput. Statist. Data. Anal. 56 (2012), pp. 1880–1897.
[4] M. Alizadeh, M. Emadi, M. Doostparast, G.M. Cordeiro, E.M. Ortega, and R.R. Pescim, *A new family of distributions: The Kumaraswamy odd log-logistic, properties and applications*, Hacettepe J. Math. Stat. 44 (2015).
[5] M. Alizadeh, S. MirMostafee, E.M. Ortega, T.G. Ramires, and G.M. Cordeiro, *The odd log-logistic logarithmic generated family of distributions with applications in different areas*, J. Stat. Distrib. Appl. 4 (2017), p. 6.
[6] B.A. Babcock and D.A. Hennessy, *Input demand under yield and revenue insurance*, Am. J. Agric. Econ. 78 (1996), pp. 416–427.
[7] R.R. Botts and J.N. Boles, *Use of normal-curve theory in crop insurance ratemaking*, J. Farm Econ. 40 (1958), pp. 733–740.
[8] T.S. Breusch and A.R. Pagan, *A simple test for heteroscedasticity and random coefficient variation*, Econometrics J. Economia Soc. 47 (1979), pp. 1287–1294.
[9] C.S. Brisolara, *Proposições para o desenvolvimento do seguro de receita agrícola no brasil: do modelo teórico ao cálculo das taxas de prêmio*, Ph.D. diss., Escola Superior de Agricultura Luiz de Queiroz, 2013.
[10] K.H. Coble and T.O. Knight, *Crop Insurance as a tool for price and yield risk management*. In: R.E Just and R.D Pope, eds., A Comprehensive Assessment of the Role of Risk in U.S. Agriculture. Natural Resource Management and Policy, Vol. 23, Springer, Boston, MA, 2001.
[11] K.H. Coble, T.O. Knight, R.D. Pope, and J.R. Williams, *Modeling farm-level crop insurance demand with panel data*, Am. J. Agric. Econ. 78 (1996), pp. 439–447.
[12] J.N.D. Cruz, E.M. Ortega, and G.M. Cordeiro, *The log-odd log-logistic Weibull regression model: Modelling, estimation, influence diagnostics and residual analysis*, J. Statist. Comput. Simul. 86 (2016), pp. 1516–1538.
[13] J.N.D. Cruz, E.M. Ortega, G.M. Cordeiro, A.K. Suzuki, and F.L. Mialhe, *Bivariate odd-log-logistic-Weibull regression model for oral health-related quality of life*, Comm. Statist. Appl. Methods 24 (2017), pp. 271–290.
[14] J.N. da Cruz, *A nova famılia de distribuiçoes odd log-logıstica: teoria e aplicaçoes*, Ph.D. diss., Escola Superior de Agricultura Luiz de Queiroz, 2016.
[15] A. da Silva Braga, G.M. Cordeiro, and E.M. Ortega, *A new skew-bimodal distribution with applications*, Comm. Statist. Theory Methods (2017), 1–19.
[16] A. da Silva Braga, G.M. Cordeiro, E.M. Ortega, and J.N. da Cruz, *The odd log–logistic normal distribution: Theory and applications in analysis of experiments*, J. Statist. Theory Pract. 10 (2016), pp. 311–335.
[17] R.H. Day, *Probability distributions of field crop yields*, J. Farm Econ. 47 (1965), pp. 713–741.
[18] C.N. de Abastecimento, CONAB, *Monitoramento agrícola – cultivos de inverno (safra 2015) e de verão (safra 2015/16)* (2016). Available at http://www.conab.gov.br/OlalaCMS/uploads/arquivos/16_02_22_10_03_02_boletim_a16_v5_n01_e_02.pdf.
[19] N. Eugene, C. Lee, and F. Famoye, *Beta-normal distribution and its applications*, Comm. Statist. Theory 31 (2002), pp. 497–512.
[20] P. Gallagher, *Us soybean yields: Estimation and forecasting with nonsymmetric disturbances*, Am. J. Agric. Econ. 69 (1987), pp. 796–803.
[21] B.K. Goodwin, *Problems with market insurance in agriculture*, Am. J. Agric. Econ. 83 (2001), pp. 643–649.
[22] B.K. Goodwin and A.P. Ker, *Nonparametric estimation of crop yield distributions: Implications for rating group-risk crop insurance contracts*, Am. J. Agric. Econ. 80 (1998), pp. 139–153.
[23] B.K. Goodwin and O. Mahul, *Risk modeling concepts relating to the design and rating of agricultural insurance contracts*, World Bank Policy Research Working Paper, 2004.
[24] J.L. Harwood, R. Heifner, K. Coble, J. Perry, and A. Somwaru, *Managing risk in farming: concepts, research, and analysis*, US Department of Agriculture, Economic Research Service, Washington, DC, 1999.

[25] D.A. Hennessy, B.A. Babcock, and D.J. Hayes, *Budgetary and producer welfare effects of revenue insurance*, Am. J. Agric. Econ. 79 (1997), pp. 1024–1034.
[26] Ipardes, *Instituto paranaense de desenvolvimento econômico e social. base de dados do estado (bdeweb)*, 2015. Available at http://www.ipardes.pr.gov.br/imp/index.php.
[27] R.E. Just and Q. Weninger, *Are crop yields normally distributed?*, Am. J. Agric. Econ. 81 (1999), pp. 287–304.
[28] A.P. Ker and K. Coble, *Modeling conditional yield densities*, Am. J. Agric. Econ. 85 (2003), pp. 291–304.
[29] A.P. Ker and B.K. Goodwin, *Nonparametric estimation of crop insurance rates revisited*, Am. J. Agric. Econ. 82 (2000), pp. 463–478.
[30] C.P. Lawas, *Crop insurance premium rate impacts of flexible parametric yield distributions: an evaluation of the johnson family of distributions*, Ph.D. diss., Texas Tech University, 2005.
[31] C.T. Lin, Y.L. Huang, and N. Balakrishnan, *A new method for goodness-of-fit testing based on type-ii right censored samples*, IEEE Trans. Reliab. 57 (2008), pp. 633–642.
[32] G.M. Ljung and G.E. Box, *On a measure of lack of fit in time series models*, Biometrika 65 (1978), pp. 297–303.
[33] MAPA, *Ministério da agricultura, pecuária e abastecimento*, 2016. Available at http://www.agricultura.gov.br/politica-agricola/seguro-rural.
[34] G.J. Miqueleto, *Contribuições para o desenvolvimento do seguro agrícola de renda para o brasil: evidências teóricas e empíricas*, Ph.D. diss., Escola Superior de Agricultura Luiz de Queiroz, 2011.
[35] C.B. Moss and J.S. Shonkwiler, *Estimating yield distributions with a stochastic trend and nonnormal errors*, Am. J. Agric. Econ. 75 (1993), pp. 1056–1062.
[36] C.H. Nelson and P.V. Preckel, *The conditional beta distribution as a stochastic production function*, Am. J. Agric. Econ. 71 (1989), pp. 370–378.
[37] V.A. Ozaki, *Métodos atuariais aplicados à determinação da taxa de prêmio de contratos de seguro agrícola: um estudo de caso*, Ph.D. diss., Universidade de São Paulo, 2005.
[38] V. Ozaki and R.S. Silva, *Bayesian ratemaking procedure of crop insurance contracts with skewed distribution*, J. Appl. Stat. 36 (2009), pp. 443–452.
[39] V.A. Ozaki, B.K. Goodwin, and R. Shirota, *Parametric and nonparametric statistical modelling of crop yield: Implications for pricing crop insurance contracts*, Appl. Econ. 40 (2008), pp. 1151–1164.
[40] R. Pakyari and N. Balakrishnan, *A general purpose approximate goodness-of-fit test for progressively type-ii censored data*, IEEE Trans. Reliab. 61 (2012), pp. 238–244.
[41] R.R. Pescim, G.M. Cordeiro, C.G. Demétrio, E.M. Ortega, and S. Nadarajah, *The new class of Kummer beta generalized distributions*, SORT 36 (2012), pp. 153–180.
[42] J.C. Quiggin, G. Karagiannis, and J. Stanton, *Crop insurance and crop production: An empirical study of moral hazard and adverse selection*, Aus. J. Agric. Econ. 37 (1993), pp. 95–113.
[43] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2016. Available at https://www.R-project.org/.
[44] O.A. Ramírez, *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*, Am. J. Agric. Econ. 79 (1997), pp. 191–205.
[45] O.A. Ramirez, S. Misra, and J. Field, *Crop-yield distributions revisited*, Am. J. Agric. Econ. 85 (2003), pp. 108–120.
[46] G. Schwarz, *Estimating the dimension of a model*, Ann. Statist. 6 (1978), pp. 461–464.
[47] B.J. Sherrick, F.C. Zanini, G.D. Schnitkey, and S.H. Irwin, *Crop insurance valuation under alternative yield distributions*, Am. J. Agric. Econ. 86 (2004), pp. 406–419.
[48] V.H. Smith and B.K. Goodwin, *Crop insurance, moral hazard, and agricultural chemical use*, Am. J. Agric. Econ. 78 (1996), pp. 428–438.
[49] C.R. Taylor, *Two practical procedures for estimating multivariate nonnormal probability density functions*, Am. J. Agric. Econ. 72 (1990), pp. 210–217.
[50] H.A. Tejeda and B.K. Goodwin, *Modeling crop prices through a Burr distribution and analysis of correlation between crop prices and yields using a copula method*, Annual Meeting of the

Agricultural and Applied Economics Association, Orlando, FL, USA, Vol. 83, Citeseer, 2008, pp. 643–649.
[51] C. Turvey and J. Zhao, *Parametric and non-parametric crop yield distributions and their effects on all-risk crop insurance premiums*, Department of Agricultural Economics and Business, University of Guelph, Guelph,Ontario, 1999.

Appendix . Maximum likelihood estimation

To estimate the parameters of the distributions presented in Section 4, we used the method of maximum likelihood. This appendix describes the statistical inference used in this article for the skew-t Student models, being that procedure the same for the remaining distributions. Consider $y_1, \dots, y_n$, a random sample of size n of the distribution ST. Then, the logarithm of the likelihood function for the vector of parameters $\boldsymbol{\theta} = (\mu, \sigma, \nu, \tau)^{\mathrm{T}}$, is given by:

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log \left\{ \frac{2}{\tau + \frac{1}{\tau}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sigma(\pi\nu)^{1/2}} \times \left[\left(1 + \frac{z^2}{\nu}\right) \frac{1}{\tau^2} \right]^{-(\nu+1)/2} \right\}, \tag{A1}$$

where, $z = (y - \mu)/\sigma$.

The MLEs $\hat{\boldsymbol{\theta}}$ of the vector of parameters can be obtained by maximizing the log-likelihood ratio test (Equation (A1)). In this step, we used the 'Nelder-Mead' and 'L-BFGS B' methods, provided in the package 'Optim' of the R *software*, [43], in which the initial values to estimate the parameters can be obtained, first, estimating the parameters for the standard t-Student model and with these values it is estimated the parameters of the ST model.

In addition, the 'Optim' package provides the option 'gr' in which the user can provide the score vector, what makes the algorithm more efficient. In this case, the components of the $U(\boldsymbol{\theta})$ score vector are given by:

$$U_\nu(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left\{ \frac{\psi(\frac{\nu+1}{2})\sqrt{\pi\nu}\tau^2 + \Gamma(\frac{\nu}{2})[\psi(\frac{\nu+1}{2})\pi\nu - \psi(\frac{\nu}{2})\pi\nu - \pi]}{2\sqrt{\pi\nu}} \right\}$$

$$+ \sum_{i=1}^{n} \left\{ -\frac{1}{2} \ln\left[\frac{\nu\sigma^2 + (y_i - \mu)^2}{\nu\sigma^2\tau^2} \right] + \frac{(\nu+1) + (y_i - \mu)^2}{2\nu[\nu\sigma^2 + (y_i - \mu)^2]} \right\},$$

$$U_\mu(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left\{ \frac{(\frac{\nu+1}{2})(-2y_i + 2\mu)}{\nu\sigma^2 + (y_i - \mu)^2} \right\},$$

$$U_\tau(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left\{ \frac{-\tau^2 + \Gamma(\frac{\nu}{2})\sqrt{\pi\nu} - (\nu+1)[\tau^2 + \Gamma(\frac{\nu}{2})\sqrt{\pi\nu}]}{\tau(\tau^2 + \Gamma(\frac{\nu}{2})\sqrt{\pi\nu})} \right\},$$

$$U_\sigma(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left\{ \frac{1}{\sigma} - \frac{(\nu+1)(y_i - \mu)^2}{\sigma[\nu\sigma^2 + (y_i - \mu)^2]} \right\}.$$

Making these equations equal to zero and solving them, simultaneously, we obtain the MLEs of the parameters, using numerical methods.

Under certain regularity conditions the parameter vector $\boldsymbol{\theta}$, in its parameter space, has the $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ asymptotic distribution, N multivariate $N_4(0, K(\boldsymbol{\theta})^{-1})$, where $K(\boldsymbol{\theta})$ is the matrix of expected information.

The $K(\boldsymbol{\theta})^{-1}$ matrix of asymptotic covariances of $\hat{\boldsymbol{\theta}}$ can be approximated by the $(4) \times (4)$ inverse of the observed information matrix $-\ddot{\mathbf{L}}(\boldsymbol{\theta})$, so you can obtain an approximation of the covariance matrix through the inverse of the observed information matrix of the parameter function. So, the MLEs can provide trust regions using the asymptotic normality.

Then, the asymptotic inferences for the vector of parameters $\boldsymbol{\theta}$ can be performed using the N approximation $N_4(0, -\ddot{\mathbf{L}}(\boldsymbol{\theta})^{-1})$ for $\hat{\boldsymbol{\theta}}$ and the standard errors for the MLEs can be obtained from the square root of the elements in the main diagonal of the observed information's inverse matrix, and one may formulate hypotheses to be tested. The elements of the observed information matrix are given by:

$$\ddot{\mathbf{L}}(\boldsymbol{\theta}) = \begin{pmatrix} J_{\mu\mu} & J_{\mu\sigma} & J_{\mu\nu} & J_{\mu\tau} \\ \cdot & J_{\sigma\sigma} & J_{\sigma\nu} & J_{\sigma\tau} \\ \cdot & \cdot & J_{\nu\nu} & J_{\nu\tau} \\ \cdot & \cdot & \cdot & J_{\tau\tau} \end{pmatrix},$$

for $i = 1, \ldots, I$, which are obtained numerically. The N asymptotic distribution $N_4(0, -\ddot{\mathbf{L}}(\boldsymbol{\theta})^{-1})$ can be used to build approximate regions of trust for the vector of parameters $\boldsymbol{\theta}$.

The asymptotic normality is also useful for testing the quality of adjustment of some submodels. The likelihood ratio statistics can be used to compare the ST distributions $(\mu, \sigma, \lambda, \alpha)$ with the N, Student and Cauchy submodels.

So, we get the value of the w statistic of the log-likelihood function of the restricted model under the unrestricted model yet to be tested. In some cases, the hypothesis of type $H_0 : \boldsymbol{\psi} = \boldsymbol{\psi}_0$ versus $H : \boldsymbol{\psi} \neq \boldsymbol{\psi}_0$ is tested, where $\boldsymbol{\psi}$ is the vector formed with some components of $\boldsymbol{\theta}$ and $\boldsymbol{\psi}_0$ is a specific vector. For example, the hypothesis $H_0 : \tau = 1$ versus $H : H_0$ is not true is equivalent to compare $\mathrm{ST}(\mu, \sigma, \tau, \nu)$ to the Student's t distribution, and the statistic W is obtained by:

$$w = 2\{\ell(\hat{\mu}, \hat{\sigma}, \hat{\tau}, \hat{\nu}) - \ell(\tilde{\mu}, \tilde{\sigma}, 1, \tilde{\nu})\},$$

in which $\hat{\alpha}$, $\hat{\mu}$ and $\hat{\sigma}$ are the MLEs under H and $\tilde{\mu}$ and $\tilde{\sigma}$ are the estimates under H_0.