



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

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

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

aesearch@umn.edu

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

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

Estimation of the Agricultural Probability of Loss: evidence for soybean in Paraná State

Vitor Augusto **Ozaki**¹, Ricardo **Olinda**²,
Priscila Neves **Faria**³ and Rogério Costa **Campos**⁴

Abstract: In any agricultural insurance program, the accurate quantification of the probability of the loss has great importance. In order to estimate this quantity, it is necessary to assume some parametric probability distribution. The objective of this work is to estimate the probability of loss using the theory of the extreme values modeling the left tail of the distribution. After that, the estimated values will be compared to the values estimated under the normality assumption. Finally, we discuss the implications of assuming a symmetrical distribution instead of a more flexible family of distributions when estimating the probability of loss and pricing the insurance contracts. Results show that, for the selected regions, the probability distributions present a relative degree of skewness. As a consequence, the probability of loss is quite different from those estimated supposing the Normal distribution, commonly used by Brazilian insurers.

Key-words: Risk, extreme value theory, probability of loss, crop insurance.

Resumo: Em todo programa de seguro agrícola, a quantificação da probabilidade da perda da produtividade agrícola é de grande importância. A fim de estimar esta quantidade, é necessário supor alguma distribuição de probabilidade paramétrica. O objetivo deste trabalho é estimar a probabilidade da perda usando a teoria dos valores extremos para modelar a cauda esquerda da distribuição. Os valores estimados foram comparados aos valores estimados sob a suposição da normalidade. Por fim, são discutidas as implicações de se supor uma distribuição simétrica em vez de uma família mais flexível de distribuições para estimar a probabilidade da perda e fixar a taxa de prêmio dos contratos de seguro. Os resultados mostraram que, diferente dos procedimentos adotados no mercado segurador, que supõem normalidade, as distribuições de probabilidade apresentam um relativo grau de assimetria que modifica o valor das probabilidades de perda e, conseqüentemente, as taxas de prêmio.

Palavras-chave: Teoria dos valores extremos, risco, probabilidade de perda, seguro agrícola.

JEL Classification: C00.

1. Professor no Departamento de Economia, Administração e Sociologia da ESALQ/USP e coordenador do GESER/ESALQ. E-mail: vitorozaki@yahoo.com.br
2. Professor no Departamento de Estatística – UEPB. E-mail: ricardo.estat@yahoo.com.br
3. Professor na Faculdade de Matemática – UFU. E-mail: prineves.est@gmail.com
4. Professor no Departamento de Estatística – UFPEL. E-mail: rogerio.c.campos@hotmail.com

1. Introduction

Historically, Brazilian agricultural insurance market faces some well recognized problems. More specifically, insurance high costs, systemic nature of risk, adverse selection, lack of accurate farm-level information, and moral risk can be pointed out as the main drawbacks.

Recently, the Federal Government has implemented some initiatives to develop crop insurance in Brazil. For instance, by the Federal Law 10,823 part of the total insurance costs will be subsidized. Now, the following step relies on providing information to drive rules by which the insurance companies will offer their products. It implies in getting around the problem concerning the lack and inaccuracy of the information exhibited in some areas.

Remarkably, the insurance market is based on individual information which is important to show the risk profile of each one. In the absence of reliable and accurate information insurers will avoid offer their contracts. In the agricultural sector farm-level yield data is almost inexistent. Some cooperatives have gathered yield information from their associated producers, but it is still far from being enough to support the spatial density of information which is demanded by the crop insurance. Municipality-level yield has been recorded and released by

Brazilian Institute for Geography and Statistics (IBGE) and used as an alternative. However, such aggregation is not desirable for local (farming fields) risk analysis (OZAKI, 2008).

According to the IBGE the Parana State is the major grain producer in Brazil. In 2011, the IBGE estimated a high production of soybean (13.4 millions of tons). This amount is 41% larger, when compared to the total reached in the previous harvest year. However, less than 10% of the planted area is covered by crop insurance. In this context, the analyses of the probability of loss can support insurance companies to deal with seasonal fluctuation of grain production. Traditionally, the Normal distribution assumption is commonly used by the insurers to quantify and price the risk. Nevertheless, by assuming the Normal distribution it is not possible to take into account the skewness and bimodalities present in the probability distributions of the agricultural yields (GOODWIN and KER, 1998).

Moreover, 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. When modeling agricultural yields one must look at the density concentrated at the left tail of the distribution. When yields events are assumed as normally distributed the probability of loss will be underestimated if the true distribution exhibits

heavier tails as in extreme values distributions (OZAKI et al., 2008).

In the literature, Just and Weninger (1999) suggest the normality assumption, whilst others point out evidences against it (DAY, 1965; TAYLOR, 1990; RAMIREZ et al., 2003). Alternatively, Beta distribution (NELSON and PRECKEL, 1989; TIRUPATTUR, HAUSER and CHAHERLI, 1996; BABCOCK, HART and HAYES, 2004; and COBLE et al., 1997), Inverse Hyperbolic Sine Transformations (MOSS and SHONKWILER, 1993; RAMIREZ, 1997), Log-Normal (GOODWIN, ROBERTS and COBLE, 2000), Skew Normal distributions (OZAKI and SILVA, 2009) and Gamma (GALLAGHER, 1987) have been proposed.

This study applies the extreme value theory to model the left tail of the agricultural yield probability distribution of the mesoregions⁵ in Parana state. Finally, the estimates are compared to those supposing the Normal distribution that is commonly used by insurance companies.

The paper is organized as follows: in Section 2 we briefly review some aspects of crop insurance. In Section 3 we show the extreme value theory and in Section 4 we describe the Brazilian yield data for soybean. In Section 5 we present our empirical findings and discuss their implications, and in Section 6 we conclude the paper.

2. Crop insurance

Basically, the compensation mechanism is triggered by the farm-level yield. Producers are indemnified when the agricultural yield observed in the end of the harvest (in the unit or farm) falls below the yield guaranteed in the contract. This type of agricultural insurance is called individual yield crop insurance. The indemnity I for each farm i can be expressed as follows:

$$I_i = \phi_i \max[(x_i^c - x_i), 0]$$

In which:

ϕ_i is the deductibility, $0 < \phi_i < 1$;

x_i^c is the critical yield;

x_i is the observed (final) yield;

The critical yield is defined according to the equation: $x_c = \alpha_i \mu_i$. In which: α_i is the level of coverage chosen by the agricultural producer, $0 < \alpha_i < 1$; and, μ_i is the farmer expected yield. In what follows, I_i represents the indemnity due to each producer when the agricultural yield x_i falls below the guaranteed yield x_i^c .

3. Methodology

Extreme value theory (EVT) is widely applied in financial, economical and insurance areas. One of the main challenges to the risk manager is to implement risk management models which allow for rare and damaging events with perverse consequences (KOEDIJK et al., 1990; LORETAN and PHILLIPS, 1994; LONGIN, 1996; EMBRECHTS et al., 1997; DANIELSSON and DE VRIES, 2000; NEFTCI, 2000; MCNEIL and FREY, 2000; GENCAY et al., 2003; DIEBOLD et al., 1998).

The regulator agents of insurance companies expect the companies be able to honor their contracts even under crises scenarios. Thus, it is mandatory to keep a reasonable fund to avoid insolvency in case of catastrophic events.

EVT has become one of the main theories in developing statistical models for extreme insurance losses. This approach is focused on a special class of probability distributions called Generalized Extreme Value distribution (GEV) which encompass distributions like Gumbel, Fréchet and Weibull. PGD (Pareto Generalized Distribution) distribution such as Exponential, Pareto and Beta are also used in the EVT approach as well. In the standard format GEV and PGD depend only on the parameter which is called tail index.

There are two main approaches to deal with extreme random variables: POT (Picks Over Threshold) approach concerns on fitting the probability distribution (usually a PGD) by taking

5. Groups of municipalities within a State with common characteristics.

values over some threshold; and, the Block Maxima (or Gumbel method) approach addresses the set of maximum values coming from a block of observations. In order to assess the market risk, for instance, the Block Maxima is used to estimate the return probability of maximum (minimum) event for a given time interval (e.g., months, years) (DE HAAN and PEREIRA, 2006).

3.1. Extreme value theory

Suppose that the sequence X_1, X_2, \dots, X_n is independent and identically distributed (i.i.d.) with distribution function $F(X)$ and let $Y_n = \max(X_1, X_2, \dots, X_n)$ with distribution function:

$$\Pr\{Y_n \leq x\} = \Pr\{X_1 \leq x, \dots, X_n \leq x\} = (F(x))^n$$

Then Y_n has the distribution function

$$\lim_{n \rightarrow \infty} \Pr(Y_n \leq a_n x + b_n) = \lim_{n \rightarrow \infty} \Pr\left[\frac{(Y_n - b_n)}{a_n} \leq x\right] = Hx \quad (1.1)$$

In which $a_n > 0$ and b_n are normalized constants.

If (1.1) holds, we say F (or X) belongs to the (maximum) domain of attraction (MDA) of H and write $F \in MDA(H)$ (or $X \in MDA(H)$). Note that H has one of the following three parametric forms (which are generally called Extreme Value Distributions - EVD):

Type I: $H(x) = \exp\{-\exp(-x)\}$, all x

$$\text{Type II: } H(x) = \begin{cases} 0, & \text{if } x \leq 0 \\ \exp(-x^{-\alpha}), & \text{if } x > 0 \end{cases} \quad (1.2)$$

$$\text{Type III: } H(x) = \begin{cases} \exp(-(x)^\alpha), & \text{if } x < 0 \\ 1, & \text{if } x \geq 0 \end{cases}$$

In II and III α is any positive number. The three types are also often called the Gumbel, Fréchet and Weibull distributions, respectively.

3.1.1. Fisher-Tippett Theorem

Let $Y_n = \max(X_1, X_2, \dots, X_n)$ be a sequence of independent and identically-distributed random variables. For some $a_n > 0$ let:

$$\Pr\{a_n^{-1}(Y_n - b_n) \leq x\} \rightarrow H(x) \quad (1.3)$$

for some non-degenerate H then it belongs to the three types described in 1.2.

Based on the Fisher-Tippett theorem is possible to estimate the asymptotic distribution of $\frac{(Y_n - b_n)}{a_n}$ directly from the family H without the distribution of X . The three types of extreme value distributions can be written into a generalized extreme value (GEV) distribution form given by

$$H(x; \mu, \sigma, \xi) = \exp\left\{-\left[1 + \frac{\xi(x - \mu)}{\sigma}\right]^{-1/\xi}\right\}$$

in which μ, σ and ξ are the parameters of location, scale and shape, respectively. Moreover, $1 + \xi(x - \mu)/\sigma > 0, \sigma > 0$. The case where $\xi = 0$ is interpreted as the limit case $\xi \rightarrow 0$, that is

$$H(x; \mu, \sigma, 0) = \exp\left\{-\exp\left[-\frac{(x - \mu)}{\sigma}\right]\right\}$$

Type II and III correspond to $\xi > 0$ ($\xi = \frac{1}{\alpha}$) and $\xi < 0$ ($\xi = -\frac{1}{\alpha}$) respectively⁶.

The Fisher-Tippett theorem gives a limit distribution for the maximum collected in a block of size n . Let x_1, x_2, \dots, x_k be observations of the random variable X . The sample data is partitioned into n blocks such that $nk \leq m$, and let

$$\begin{aligned} Y_1 &= \max\{x_1, \dots, x_k\} \\ Y_2 &= \max\{x_{k-1}, \dots, x_{2k}\} \\ &\vdots \\ Y_n &= \max\{x_{nk-k+1}, \dots, x_{nk}\} \end{aligned}$$

The estimators of $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\xi}$ are estimated using this new sample, Y_1, Y_2, \dots, Y_n .

From the GEV we can get the probability density function of the GEV distribution given by:

$$f_X(x; \mu, \sigma, \xi) = \frac{1}{\sigma} \left[1 + \xi \left(\frac{x - \mu}{\sigma}\right)\right]^{\frac{1+\xi}{\xi}} \exp\left\{-\left[1 + \xi \left(\frac{x - \mu}{\sigma}\right)\right]^{\frac{1}{\xi}}\right\}$$

6. Smith (1990) detailed the statistical treatments, applications and estimations of the GEV.

in which $-\infty < x < (\mu - \sigma)/\xi$ for $\xi < 0$, and $(\mu - \sigma)/\xi < x < +\infty$ for $\xi > 0$ in which x is a random variable associated to the maximum values. Suppose now that $X_i, i = 1, 2, \dots$ is a sequence of *i.i.d.* random variables with a continuous marginal distribution function $F(x)$, and $\hat{X}_i, i = 1, 2, \dots$ is the so-called associated sequence of *i.i.d.* random variables with the same marginal distribution function F . Note that Y_n stands for the maximum as usual, defined by (1.1), while \hat{Y}_n denotes the corresponding maximum of $\{\hat{X}_1, \dots, \hat{X}_n\}$. The limiting distribution of Y_n can be related to the limiting distribution of \hat{Y}_n via a quantity called the extremal index of the sequence $\{X_n\}$ (CARTWRIGHT, 1958; NEWELL, 1964; O'BRIEN, 1974).

3.2. Selection of the extreme value distribution

The statistics of likelihood ratio (T_{LR}) is defined by

$$T_{LR} = -2[l(\hat{\theta}_G) - l(\hat{\theta}_{GEV})] = 2[l(\hat{\theta}_{GEV}) - l(\hat{\theta}_G)] \quad (1.4)$$

In which $l(\hat{\theta}_{GEV})$ and $l(\hat{\theta}_G)$ are the maximum of the logarithm of the maximum likelihood function of GEV and the Gumbel distribution in which $\hat{\theta}_{GEV} = (\hat{\mu}, \hat{\sigma}, \hat{\xi})$ and $\hat{\theta}_G = (\hat{\mu}, \hat{\sigma})$ are vectors of the estimated parameters μ, σ and ξ with asymptotic distribution χ^2 with one degree of freedom. Hosking et al. (1985) suggest the use of a modified test statistics to improve the approximation of the asymptotic distribution of (1.4) given by

$$T_{LR}^* = \left(1 - \frac{2.8}{n}\right) \times T_{LR}$$

In which n is the length of the sample.

To test the null hypothesis $H_0 : \xi = 0$ versus $H_1 : \xi \neq 0$, one must compare the test statistics T_{LR}^* with the tabulated value of χ^2 distribution with one degree of freedom and 9 significance level. If $T_{LR}^* \geq \chi_{[0.1]}^2$, H_0 is rejected. In other words, there is strong evidence that the distribution is not the type I (Gumbel).

3.3. Diagnostic of the extreme value distribution

In order to test the assumption that the data follow a GEV distribution, it is possible to use the Kolmogorov-Smirnov test (SANSIGOLO, 2008; BAUTISTA et al., 2004). The D statistic of the Kolmogorov-Smirnov tests is defined by

$$D = \max |F(x_{(i)}) - \hat{F}(x_{(i)})|, \quad i = 1, 2, \dots, n$$

in which $F(x_{(i)})$ is the theoretical distribution of GEV and $\hat{F}(x_{(i)})$ is the empirical distribution. The test procedure consists in sorting the data in ascending order. The distribution function assumed for the data is defined by $F(x_{(i)})$ and the empirical distribution function of X is described as follows:

$$\hat{F}(x_{(i)}) = \frac{i}{n+1}, \quad i = 1, 2, \dots, n$$

The hypothesis that the data follow the GEV distribution (H_0) will be rejected if the test statistics $D \geq D_{[n, 0.1]}$, where $D_{[n, 0.1]}$ is the critical value. One should also use the graphical interface via the *qq*-plot (quantile-quantile chart).

In order to determine the probability distribution assuming a Normal distribution we used the maximum likelihood method to estimate the two first moments of the distribution. In what follows we fitted the Normal distribution and it was compared to the GEV distribution.

The next step is to calculate the probability of loss by integrating the area below the curve less than a predetermined level (a percentage of the average yield). In the crop insurance market, this percentage – often called level of coverage – range between 60 up to 80% of the average yield for each mesoregion. In this study we consider only three levels, 60, 70 and 80%.

The shape of the distribution is of great importance when calculating the probability of loss. Considering the situation where the true distribution is symmetric and one fits an

asymmetrical distribution to the data. In this context if the asymmetry is negative, then the probability of loss will be underestimated in relation to the true distribution. The opposite is also true.

4. Data description

The municipality-level soybean yield data were acquired from the Secretary of Agriculture of the State of Parana (Seab), from 1981 to 2007, in kg/hectare. The municipality-level yield is an average for crop yield in the municipality and is based on the subjective methodology created by the Geography and Statistics Brazilian Institute (IBGE). These statistics are based on a consensus among agricultural players in a municipality (farmers, bank managers, crop extensionists, etc). Thus, the IBGE releases the average yield for each municipality. Given a mesoregion, we select the minimum value of the soybean yield in a group of municipalities within the mesoregion. In other words, if there are ten mesoregions, we have ten observations in each year. This data are collected and released annually (two year lag).

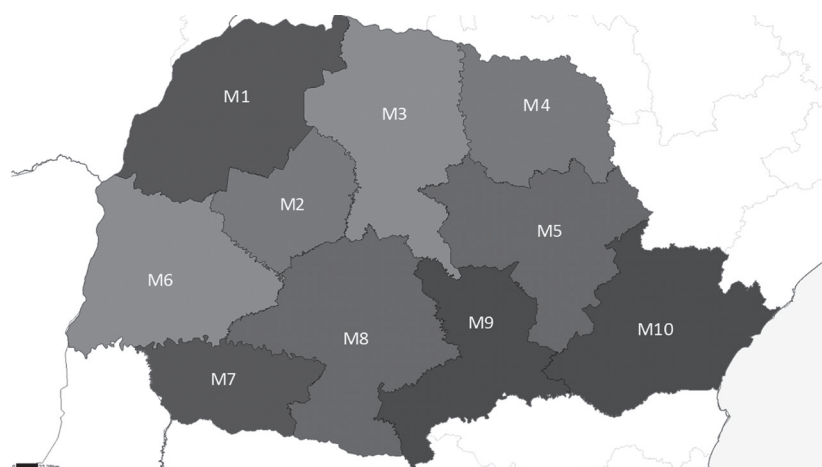
Thus, the empirical application considers mesoregions (a set of municipalities) defined by the Brazilian Institute of Geography and Statistics (IBGE) (Figure 1). Considering the fact that in each mesoregion there exist a set of municipalities then the minimum value of this set of municipalities in a year is used to create the time series of minimum values for each mesoregion. We utilized this procedure for each year since 1980. Figure 2 shows the evolution of the average of the minimum values, from 1981 to 2007.

5. Results and discussion

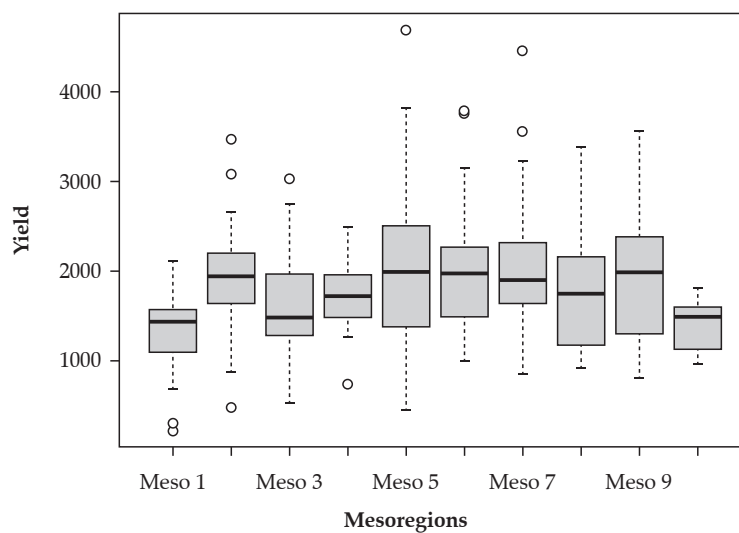
Table 1 and Figure 3 show, respectively, descriptive statistics and the boxplot of the soybean yield for all mesoregions. The descriptive statistics show that the median is systematically higher than average or smaller, which suggests that the distributions are asymmetric to the left or to the right, respectively.

The coefficients of asymmetry and kurtosis can be used in the exploratory analysis to recognize the shape (skewness, kurtosis) of the distribution, or even to recognize mixtures of distributions.

Figure 1. Mesoregions in Paraná State



Source: Results of the study.

Figure 2. Boxplot of the yields for each mesoregion

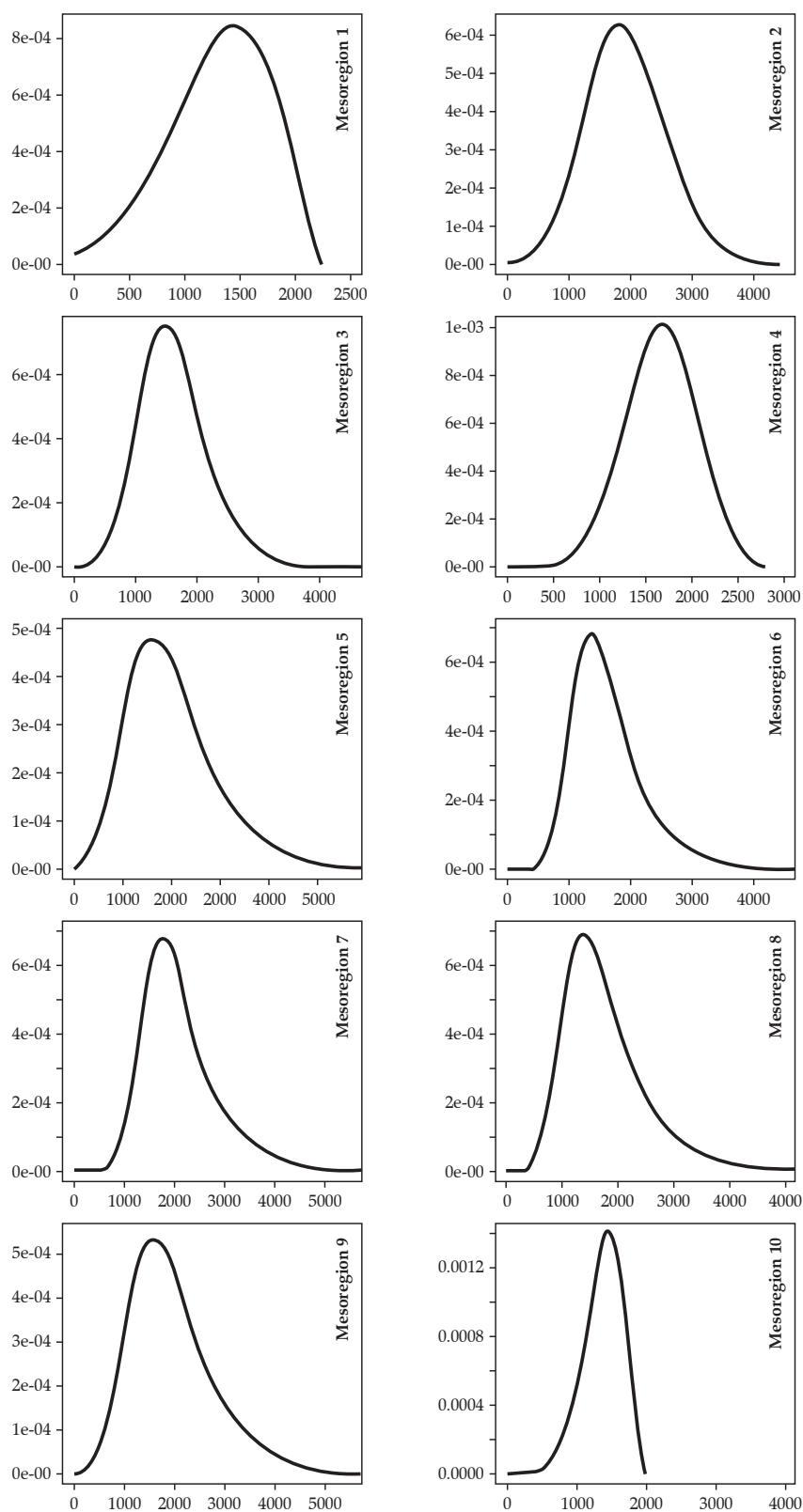
Source: Results of the study.

Table 1. Soybean yield descriptive statistics

Mesoregion	Number of Counties	Mean	Median	Standard Deviation	Skewness	Excess Kurtosis	CV
M ₁	61	1302.5	1427	471.0	-0.487	-0.321	36,2
M ₂	25	1931.3	1941	626.3	0.146	0.051	32,4
M ₃	79	1639.9	1476	560.4	0.501	0.038	34,2
M ₄	46	1712.6	1712	373.4	-0.209	0.117	21,8
M ₅	14	2022.7	2000	974.2	0.827	0.019	48,2
M ₆	50	2006.7	1983	719.9	0.898	0.043	35,9
M ₇	37	2096.7	1907	752.8	1.279	1.741	35,9
M ₈	29	1769.1	1740	706.9	0.517	-0.794	40,0
M ₉	21	1968.8	1983	868.1	0.465	-1.003	44,1
M ₁₀	37	1383.7	1495	284.3	0.051	-1.501	20,5

Source: Results of the study.

Figure 3. Density estimation



Source: Results of the study.

Table 2. Estimates of the location, scale and shape parameters

Mesoregion	location	scale	Shape (tail index)
M ₁	1184.327	498.578	-0.4793
M ₂	1671.326	604.537	-0.2243
M ₃	1416.531	498.270	-0.1446
M ₄	1544.749	385.109	-0.3140
M ₅	1593.662	765.571	-0.0241
M ₆	1678.665	541.199	0.0262
M ₇	1770.485	548.770	0.0182
M ₈	1425.315	536.631	0.0588
M ₉	1569.713	687.459	-0.0011
M ₁₀	1307.421	293.905	-0.4446

Source: Results of the study.

Table 2 shows the point estimates of the shape parameter $\xi = \frac{1}{\alpha}$. Results show that the shape parameters were negative for 7 regions and positive for 3 regions suggesting that the extreme distributions are either Weibull or Gumbel.

Comparing the value of the statistics (T_{LR}^*) presented in Table 3 with the χ^2 statistics with one degree of freedom and 5% level of significance ($\chi^2_{1;0.05} = 3.84$), one can note that the distribution to be fitted is the Gumbel distribution in the mesoregions 2, 3, 5, 6, 7, 8, 9 and 10. On the other side the mesoregions 1 and 4 will be modeled using the Weibull distribution. This fact

is confirmed by the D statistics (Kolmogorov-Smirnov test) presented in Table 3. According to the test most of the mesoregions could be fitted by the Gumbel distribution.

It is important to note that the Normal distribution could also be used as an alternative probability distribution according to the Shapiro-Wilk normality test. In Table 4 one can note that in mesoregions 2, 3 and 4 the Normal distribution could not be rejected at the level of 10%. On the other hand, Atwood et al. (2003) raise critical issues on the use of the Normal distribution when modeling crop yields.

Table 3. Intervals of 95% confidence for the shape parameter (ξ) and values of the Modified Likelihood Ratio Test (T_{LR}^*)

Mesoregion	Limits of 95% confidence intervals for ξ		(T_{LR}^*)	Distribution
	Lower	Upper		
M ₁	-0.721	-0.211	10.240	Weibull
M ₂	-0.421	0.003	3.783	Gumbel
M ₃	-0.382	0.093	1.353	Gumbel
M ₄	-0.545	-0.123	5.975	Weibull
M ₅	-0.288	0.248	0.032	Gumbel
M ₆	-0.294	0.346	0.030	Gumbel
M ₇	-0.207	0.243	0.026	Gumbel
M ₈	-0.496	0.554	0.082	Gumbel
M ₉	-0.571	0.471	0.001	Gumbel
M ₁₀	-0.961	0.073	0.030	Gumbel

Source: Results of the study.

Table 4. Shapiro-Wilk normality test

Mesoregion	statistic W	p-value
M ₁	0.9653	0.0483
M ₂	0.9751	0.7390
M ₃	0.9662	0.5044
M ₄	0.9775	0.8012
M ₅	0.9385	0.0121
M ₆	0.9159	0.0314
M ₇	0.8904	0.0081
M ₈	0.9334	0.0837
M ₉	0.9205	0.0404
M ₁₀	0.9092	0.0218

Source: Results of the study.

Figure 4 shows the distributions adjusted for each mesoregion. Most of the distributions are asymmetric to the right. In this context, the probability of loss is supposed to be higher compared to the Normal distribution case, commonly used by the insurance companies. The Figure 5 shows the diagnostics of the fitted distribution through the *qq*-plot. Considering the level of significance of 1%, Table 4 shows that the

data follow the GEV distribution. In other words, the null hypothesis cannot be rejected.

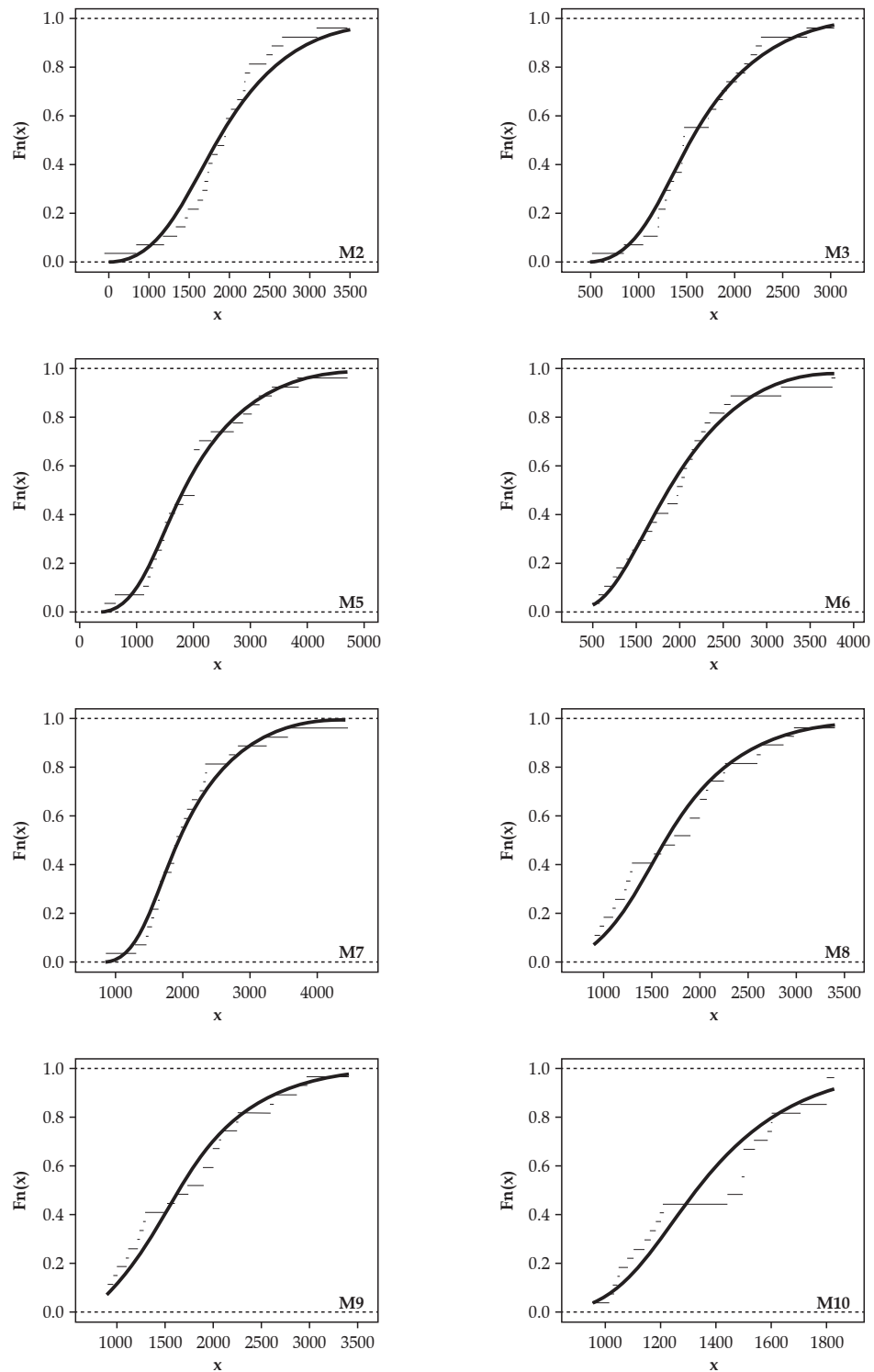
Once the distribution of interest is chosen, the next step is to determine the probability of loss for each mesoregion and compare the results assuming the Normal distribution. Tables 5 and 6 show the probabilities of loss of the selected municipalities using both distributions (Weibull-Gumbel and Normal).

Table 5. Kolmogorov-Smirnov test – diagnostic of the fitted distribution (Gumbel)

Mesoregion	statistic D	p-value
M ₁	0.2696	0.0194
M ₂	0.1446	0.5752
M ₃	0.1257	0.7405
M ₄	0.2262	0.0380
M ₅	0.1073	0.9150
M ₆	0.1092	0.8698
M ₇	0.1123	0.8484
M ₈	0.1335	0.7218
M ₉	0.1185	0.8426
M ₁₀	0.1166	0.1586

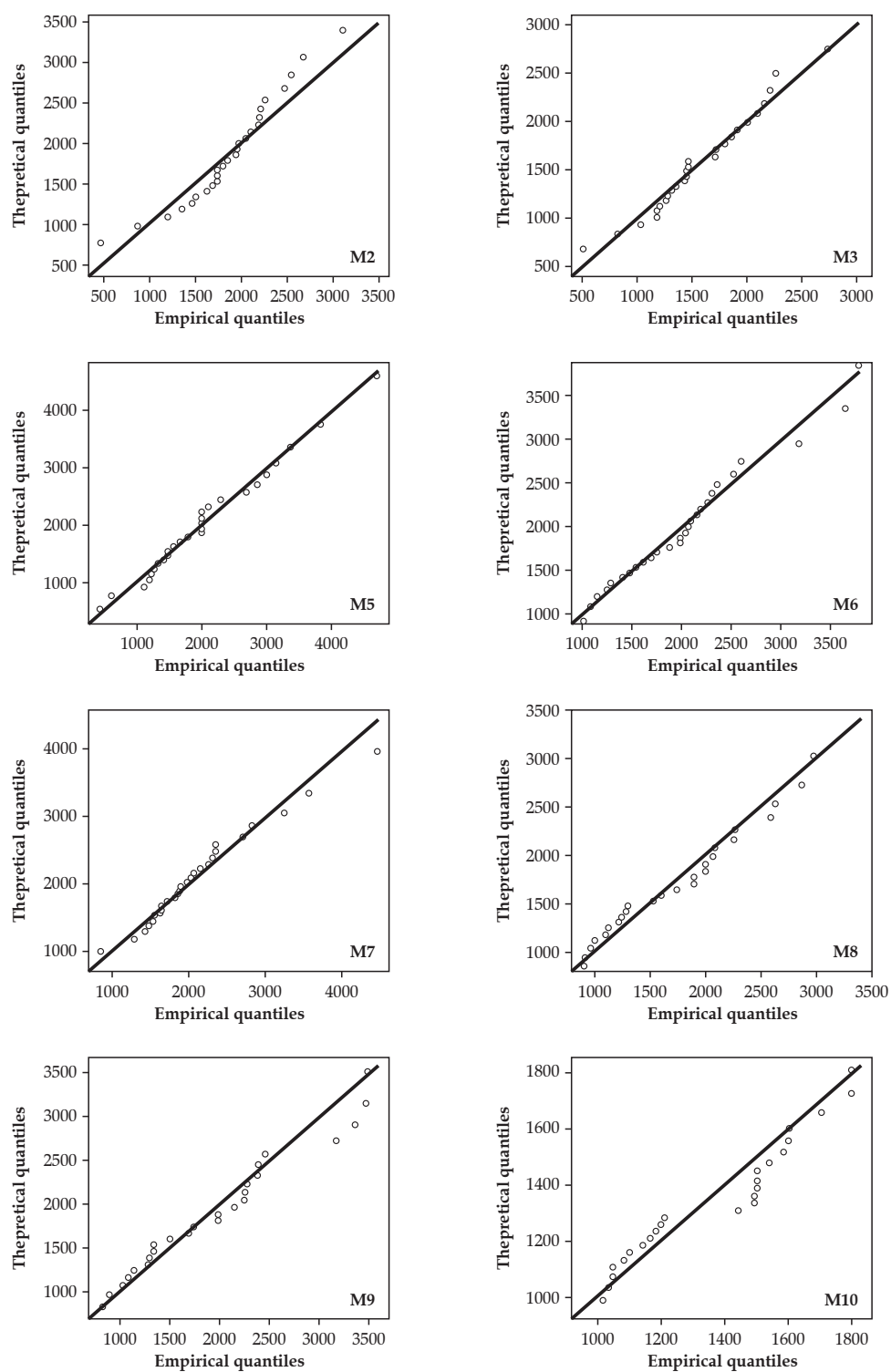
Source: Results of the study.

Figure 4. Kolmogorov-Sminorv test of the empirical cumulative distribution (dotted line) and theoretical distribution (continuous line)



Source: Results of the study.

Figure 5. QQ-plot –Gumbel distribution



Source: Results of the study.

Table 6. Probability of loss assuming the Weibull and Gumbel distributions – levels of coverage of 60, 70 and 80%

Mesoregion	Level of coverage (9)		
	60%	70%	80%
M ₁	13.8	19.7	27.1
M ₂	11.4	19.3	29.3
M ₃	10.4	18.6	29.2
M ₄	4.7	11.0	21.7
M ₅	19.4	28.3	38.0
M ₆	8.8	18.9	31.8
M ₇	7.7	17.5	30.6
M ₈	13.6	24.1	36.1
M ₉	17.2	26.7	37.1
M ₁₀	3.4	7.9	16.3

Source: Results of the study.

Table 7. Probability of loss assuming the Normal distribution - levels of coverage of 60, 70, and 80%

Mesoregion	Level of coverage (9)		
	60%	70%	80%
M ₁	21.0	29.2	38.8
M ₂	19.8	29.9	41.7
M ₃	19.3	29.9	41.7
M ₄	9.0	18.5	32.5
M ₅	31.0	40.8	51.3
M ₆	19.0	30.6	44.6
M ₇	17.5	29.1	43.3
M ₈	24.9	36.4	49.3
M ₉	28.6	39.0	50.3
M ₁₀	5.2	12.5	24.8

Source: Results of the study.

One must note that the level of coverage is a parameter chosen by the insured at the beginning of the insurance contract. In this context, the agricultural producer can choose the levels from 60 to 80% of the average yield of the municipality i (μ_i) resulting in the critical yield. If the final yield - in the end of the harvest, y_i , is lesser than the critical, the insured receives the indemnity. In this study, because of the fact that we aggregate the information into mesoregions, the critical yield is related to each mesoregion instead of municipalities. Thus, the premium rate is calculated for each mesoregion.

In Tables 5 and 6 one can notice that for all levels of coverage, the probability of loss estimated using the Normal distribution is greater than the estimated probability using the Weibull-Gumbel

distributions. A direct implication of this fact is the overpricing of the premium rate, given by (Goodwin and Ker, 1998):

$$\text{Premium Rate} = \frac{F_Y(\alpha\mu)E_Y[\alpha\mu - (Y | y < \alpha\mu)]}{\alpha\mu}$$

in which E is the expectation operator and F is the distribution function.

The probability of loss is represented by the distribution function in the premium rate formulae. In the calculation of the rate, the operator of expectation and the term in the denominator are constant.

In what follows, we present a hypothetical example to illustrate this fact. Consider a soybean producer in Mesoregion 1. The premium rate calculated for the level of 80% is equal 2% and

3,18%, respectively, considering the distributions Weibull and Normal.

In the insurance market this difference represents a large variation in the total premium. Considering for example, that a insurance company sells to a soybean producer a insurance contract in the Mesoregion 1 charging 3,18% (Normal rate) instead of 2% (Weibull rate). Suppose, for instance, the average liability is equal to US\$ 1 mi in a pool (20 thousand producers). The average premium charged is approximately US\$ 31,800 instead of US\$ 20,000. The soybean producers will be overcharged on average in US\$ 11,800 and US\$ 236 mi, in total.

6. Conclusion

In this research we analysed alternative statistical assumptions for estimating the probability of loss with SEAB's agricultural yield data and analyse its consequences in the premium rate. The soybean data were adjusted considering the Weibull-Gumbel and Normal distributions for all mesoregions in Paraná state.

Moreover, the probability of loss in Table 5 is quite different from the results in Table 6. The normality assumption is commonly used by most of the crop insurance companies in Brazil because of its mathematical tractability. Looking more carefully at the results, for all levels of coverage, the probabilities of loss are higher in the Normal case. On average, 77, 54 and 41% for the 60, 70 and 80% level of coverage, respectively.

It means that insurance companies are ignoring the skewness of the distribution and overpricing the risk. The pure premium rate is actually smaller than the premium rate charged. The consequence for the insurance market is that high risk producers may find attractive to demand the insurance increasing the probability to receive the indemnity. This classical problem, known as adverse selection, is well documented in the economical and actuarial literature and it could cause severe problems of financial losses in the agricultural insurance market.

From the insurance company point of view, supposing a Normal distribution, the probability of loss is greater than the rate estimated using the Weibull-Gumbel distributions. It means that if the true distribution is non-symmetric then the Normal case will overestimate the true rate. When the crop insurance is not compulsory, higher rates implies that the insurance company will adversely select only the farmers with higher risks. Otherwise, when the crop insurance is compulsory, the insurance company increases their revenue by charging a higher premium rate dissatisfying the farmer. However, in this situation, by spreading out the crop insurance in several regions the risk is better managed by the insurance company. On the other hand if the true distribution is symmetric than the Normal distribution could be used to estimate the premium rate. In this case, estimating the rates by the Weibull-Gumbel distributions will underestimate the rates. Farmers will be undercharged and the insurance companies will lose revenue.

In a market where historically most of the producers avoid demanding insurance products because of its high premium rate, and high probability to have indemnities paid higher than premiums collected by insurers, better statistical assumptions should be taking into account to better reflect the agricultural risk and the calculation of the premium rate.

7. References

- ATWOOD, J., SHAIK, S. and WATTS, M. Are crop yields normally distributed? A reexamination. *American Journal of Agricultural Economics*, v. 85, p. 888-901, 2003.
- BABCOCK, B. A., HART, C. E. and HAYES, D. J. Actuarial fairness of crop insurance rates with constant rate relativities. *American Journal of Agricultural Economics*, v. 86, p. 563-575, 2004.
- BAUTISTA, E. A. L., ZOCCHI, S. S. and ANGELOCCI, L. R. A distribuição generalizada de valores extremos aplicada ao ajuste dos dados de velocidade máxima de vento em Piracicaba, SP. *Revista de Matemática e Estatística*, v. 22, p. 95-111, 2004.

- CARTWRIGHT, D. E. On estimating the mean energy of sea waves from the highest waves in a record. *Proceedings of the Royal Society of London*, v. 247, p. 22-28, 1958.
- COBLE, K. H., KNIGHT, T. O., POPE, R. D. and WILLIAMS, J. R. An expected indemnity approach to the measurement of moral hazard in crop insurance. *American Journal of Agricultural Economics*, v. 79, p. 216-26, 1997.
- DANIELSSON, J. and VRIES, C. de. Value-at-Risk and extreme returns. *Annales d'Economie et de Statistique*, v. 60, p. 239-270, 2000.
- DAY, R. H. Probability distributions of field crop yields. *Journal of Farm Economics*, v. 47, p. 713-741, 1965.
- DIEBOLD, F. X., SCHUERMANN, T. and STROUGHAIR, J. D. Pitfalls and opportunities in the use of extreme value theory in risk management. In: REFENES, A. P., A. BURGESS and J. MOODY (Ed.). *Decision Technologies for Computational Finance*, Kluwer Academic Publishers, 1998.
- EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH, T. *Modelling extremal events for insurance and finance*. Springer Verlag, Berlin, 1997.
- FISHER, R. A. and TIPPETT, L. H. C. Limiting forms of the frequency distribution of the largest and smallest member of a sample, *Proc. Cambridge Phil. Soc.*, v. 24, p. 180-190, 1928.
- GALLAGHER, P. U.S. Soybean yields: estimation and forecasting with nonsymmetric disturbances. *American Journal of Agricultural Economics*, v. 69, p. 796-803, 1987.
- GENCAY, R., SELCUK, F. and ULUGÄLYAGCI, A. High volatility, thick tails and extreme value theory in value-at-risk estimation. *Insurance: Mathematics and Economics*, v. 33, p. 337-356, 2003.
- GOODWIN, B. K. and KER, A. P. Nonparametric estimation of crop yield distributions: implications for rating group-risk crop insurance contracts. *American Journal of Agricultural Economics*, v. 80, p. 139-153, 1998.
- GOODWIN, B. K., ROBERTS, M. C. and COBLE, K. H. Measurement of price risk in revenue insurance: implications of distributional assumptions. *Journal of Agricultural and Resource Economics*, v. 25, p. 195-214, 2000.
- HOSKING, J. R. M., WALLIS, J. R. and WOOD, E. F. Estimation of the generalized extreme value distribution by the method of probability-weighted moments. *Technometrics, Alexandria*, v. 27, p. 251-261, 1985.
- DE HAAN, L. and PEREIRA, A. E. Spatial extremes: Models for the stationary case. *Annals of Statistics*, Philadelphia, v. 34, p. 146-168, 2006.
- JUST, R.E. and WENINGER, Q. Are crop yields normally distributed? *American Journal of Agricultural Economics*, v. 81, p. 287-304, 1999.
- KOEDIJK, K. G., SCHAFGANS, M. and VRIES, C. de. The tail index of exchange rate returns. *Journal of International Economics*, v. 29, p. 93-108, 1990.
- LONGIN, F. M. The asymptotic distribution of extreme stock market returns. *Journal of Business*, v. 69, p. 383-408, 1996.
- LORETAN, M. and PHILLIPS, P. Testing the covariance stationarity of heavy-tailed time series. *Journal of Empirical Finance*, v. 1, p. 211-248, 1994.
- MCNEIL, A. J. and FREY, R. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, v. 7, p. 271-300, 2000.
- MOSS, C. B. and SHONKWILER, J. S. Estimating yield distributions with a stochastic trend and non Normal errors. *American Journal of Agricultural Economics*, v. 75, p. 1056-1062, 1993.
- NEFTCI, S. N. Value at risk calculations, extreme events, and tail estimation. *Journal of Derivatives*, p. 23-37, 2000.
- NELSON, C. H. and PRECKEL, P. V. The conditional beta distribution as a stochastic production function. *American Journal of Agricultural Economics*, v. 71, p. 370-378, 1989.
- NEWELL, G. F. Asymptotic extremes for m-dependent random variables. *Annals of Mathematical Statistics*, v. 35, p. 1322-1325, 1964.
- O'BRIEN, G. L. Limit theorems for the maximum term of a stationary process. *Annals of Probability*, v. 2, p. 540-545, 1974.
- OZAKI, V. A. and SILVA, R. S. Bayesian ratemaking procedure of crop insurance contracts with skewed distribution. *Journal of Applied Statistics*, v. 36, p. 443-452, 2009.
- OZAKI, V. A. A digression about the subvention program for rural insurance premium and the implications for the future of the rural insurance market. *Brazilian Review of Risk and Insurance*, v. 2, p. 79-94, 2008.
- OZAKI, V. A., GOODWIN, B. K. and SHIROTA, R. Parametric and nonparametric statistical modeling of crop yield: implications for pricing crop insurance contracts. *Applied Economics*, v. 40, p. 1151-1164, 2008.

RAMIREZ, O. A. Estimation and use of a multivariate parametric model for simulating heteroskedastic, correlated, nonNormal random variables: the case of corn belt corn, soybean and wheat yields. *American Journal of Agricultural Economics*, v. 79, 191-205, 1997.

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

SANSIGOLO, C. A. A distribuição de extremos e precipitação diária, temperatura máxima e mínima

e velocidade do vento em Piracicaba, SP (1917-2006). *Revista de Matemática e Estatística*, v. 23, p. 341-346, 2008.

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

TIRUPATTUR, V., HAUSER, R. J. and CHAHERLI, N. M. Crop yield and price distributional effects on revenue hedging. *Office of Futures and Options Research*, v. 5, p. 1-17, 1996.