African Journal of Agricultural Research Vol. 3 (12), pp. 818-824, December, 2008
Available online at http://www.academicjournals.org/AJAR
ISSN 1991-637X © 2008 Academic Journals

Full Length Research Paper

Estimating the threshold value and loss distribution: Rice damaged by typhoons in Taiwan

Li-Hua Lai[1]* and Pei-Hsuan Wu[2]

[1]Department and Graduate Institute of Risk Management and Insurance, National Kaohsiung First University of Science and Technology, Taiwan.
[2]Institute of Management, National Kaohsiung First University of Science and Technology, Taiwan. Department of Public Finance, Meiho Institute of Technology, Taiwan

Accepted 24 November, 2008

We apply extreme value theory to determine the over-threshold peaks of the data and then use the Kolmogorv-Smirnov and Anderson-Darling goodness of fit tests to show that the generalized Pareto distribution fits the heavy-tailed distribution better than the Lognormal, Gamma, Weibull and Normal distributions in rice damaged by typhoons. The appropriate of the threshold value and probable maximum loss can be calculated as one of reference indexes on risk retention or/and crop insurance associated with the natural systematic risk of major agricultural disasters. The properties we found are useful in crop loss assessment and in the decision making of government's risk financing for major agricultural disasters. Our method may be applied to other disasters and other countries.

Key words: natural disasters, generalized Pareto distribution, heavy-tailed distribution, threshold value, probable maximum loss.

INTRODUCTION

Natural disasters have caused extensive damage to worldwide crop production (Adams et al., 1998; Rosenzweig et al., 2002; Chang, 2002; Larsson, 2005). Rice production is largely concentrated in Asia, where it is considered to be the major source of food (Luo et al., 1998; Aggarwal et al., 2006). Of all the natural disasters occurring in Asia, tropical cyclones (typhoons) are the most serious (Lansigan et al., 2000; Ji et al., 2002). Over a 30 years period, Taiwan was hit by an average of 3.3 typhoons which is expected to cost taxpayers about $1,260 million for crop loss each time, and they brought abundant rainfalls and strong winds, leading to sever damage to crops and great property losses. Taiwan's agricultural natural disaster assistance program (TANDAP) was passed implementation in 1991. The program provides government's substantial funds (cash succor, subsidy or/and loan of preferential interest rate) to farmers to compensation for yield loss resulting from natural disaster. To receive compensation, a farmer has to meet TANDAP's "area-damaged" triggered loss criteria in individual loss. We take the case of "cash succor" as an example, an "area-damaged" must incur a loss of at least 30% for any given crop before any farmer of the crop in that area is eligible for a "cash succor" payment. Once this area-damaged is met, the individual eligibility criterion requires that the individual's loss must excess 50% of his or her normal yield to receive a payment. Although it is fact but it does not seem high in the proportion of the agricultural loss of the whole that the cash is succored, only about average 5.34% in one year. It is important to note that the best-fitted loss distribution and the threshold value or limit which can be deemed a part of measuring the probable maximum loss (PML) of risk retained by the government, or/and to make crop insurance for exceeding the loss, but the normalized assumption has been widely applied to estimate the loss of crop in earlier studies, in which the extreme values of the crop loss distribution are usually ignored (Hansen, 2004; Larsson, 2005; Muralidharan and Pasalu, 2006). Furthermore, it is not practical to obtain a distribution under assumption of frequency and severity distributions separately, and only aggregate information is available

---

*Corresponding author. E-mail:lihua@ccms.nkfust.edu.tw. Tel.: 886-7-6011000 EXT 3020.

for analysis. In this paper, the testing about the shape of aggregate loss distribution becomes very important, especially in the "tail" of the distribution of loss due to typhoons. From the crop loss assessing point of view, most prior research was based upon theoretical considerations, and did not consider directly the goodness of fit of various approximation distributions with the heavy-tailed. Therefore it is interesting to analyze the tail region instead of the center region of the distribution of rice damaged by typhoons, as suitable estimates for the tails of loss severity distributions are essential for risk financing or positioning of high-excess loss layers in private or governmental insurance and disaster risk management programs (Rosenzweig et al., 2002). Hence, it is more accurate to estimate the number of events that exceed the economic damage threshold from a model (Hansen, 2004; Muralidharan and Pasalu, 2006), and the choice of the threshold value above which losses are insured or retained by the government also becomes a critical decision (Hansen, 2004; Larsson, 2005). These methods revolve around the generalized Pareto distribution (GPD) and are supported by generalized extreme value theory (GEV) (Hosking and Wallis, 1987; Embrechts et al., 1999; Brabson and Palutikof, 2000). The GPD was introduced by Pickands (1975) as a two-parameter distribution and has been widely applied by many investigators (Hosking and Wallis, 1987; Prudhomme et al., 2003) in flood frequency analysis. We are using the generalized extreme value theory and the generalized Pareto distribution, building a sample empirical analysis of rice damaged by typhoons, and then directly testing the goodness of fit of various approximations to this observed sample. In addition, we tested five widely used two-parameter distributions, to test their fits to the loss distributions constructed in each of the same dataset.

This paper is organized as follows. Section 2 summarizes the theoretical materials and methods concerning the estimation of the extreme observations. Section 3 interprets and discusses the results which estimate the parameters of the loss distributions using maximum likelihood and analyses theoretical model that can fit the data. We draw conclusions in Section 4.

MATERIALS AND METHODS

This paper will present the question what type of probability distribution is the most appropriate to use to approximate a distribution of rice damaged by typhoons. We continue the research into the accuracy of different approximations of the distribution which to get at the first factor in risk assessment of rice damaged, and measure the likelihood natural events with the extreme value and express the range of magnitudes of events observed. However, there are five aspects that differentiate it from previous investigations. First, we use the Kernel density estimation to explore the tail behavior of the data. Second, we concentrated our consideration to two-parameter probability distributions, and try to find parameters that cause the function's statistical properties to match those of the empirical distribution. Third, the choice of the threshold u or, likewise, the number k of upper ordered values can be supported visually by a diagram. The advantages of peaks over threshold methods and more data can be used as estimators and are not affected by the small “rice damaged”. Therefore, estimates of u are plotted against the number k of upper ordered values. If k is appropriate for the data, the values of the estimates stabilize around the true extremes and a plateau becomes visible. Fourth, the goodness-of-fit is measured using the Kolmogorov-Smirnov (K-S test) and Anderson-Darling test (A-D test). The theoretical distribution with the smallest K-S and A-D values is determined to be the best fit to the empirical distribution of rice damaged by typhoons. Finally, we explore and discuss the relation linking of the best-fitted loss distribution, the threshold value and the probable maximum loss (PML) that is how grouping losses and threshold value of risk financing would be affected under the agricultural natural disaster assistance program.

Detection of the heavy-tailed characteristic of the data

Kernel density estimation is a non-parametric way of estimating the probability density function of a random variable (Silverman, 1986). Let $\mathbf{X} = (X_1, \cdots, X_n)$ be a random sample from a university distribution with unknown density f. Let K be a symmetric probability density function to be used as a kernel and $h > 0$ its scaling parameter, or bandwidth; write $K_h(.) = h^{-1}K(h^{-1})$. Then the standard kernel estimator of the density f at a point x is given by

$$\hat{f}(x) = n^{-1}\sum_{i=1}^{n} K_h(x - X_i) \tag{1}$$

If kernel density function exhibits the right-skewed tendency, it stresses the long-tailed behavior of the underlying data (McNeil, 1997).

Threshold value model

When the heavy tailed character of the data is fulfilled, the data are appropriate by using GPD modeling fitting, if enough data are available above a high enough threshold. The practical problem comes to how to determine a “high enough threshold” or, likewise, the number k of upper extremes corresponds to upper ordered values. Reiss and Thomas (1997) proposed a heuristic method of choosing the number of extremes to estimate the tail index estimates. Let $\hat{\xi}_{k,n}$ be estimates of the shape parameter ξ based on the k upper extremes. The selection procedures choose k as the value that minimizes:

$$\frac{1}{k}\sum_{i\le k} i^{\beta}\,|\hat{\xi}_{i,n} - med(\hat{\xi}_{1,n},\ldots,\hat{\xi}_{k,n})| \quad ,0 \le \beta \le \frac{1}{2} \tag{2}$$

where $med(\hat{\xi}_{1,n},\ldots,\hat{\xi}_{k,n})$ denotes the median of $\hat{\xi}_{1,n},\ldots,\hat{\xi}_{k,n}$ and β is the coefficient of automatic selection in calculating system. The procedures are solved reasonably well by the “Xtremes” package. Moreover, we can use a graphic tool, GPD

Table 1. Summary of property for other distributions.

Type of Distribution	Parameters	Probability Density Function	Mean	Variance
Gamma	$\alpha > 0$, $\beta > 0$	$f(x) = 1/(\Gamma(\alpha)) \times \beta^{-\alpha} x^{\alpha-1} \exp(-x/\beta)$	$\alpha\beta$	$\alpha\beta^2$
Lognormal	$-\infty < \mu < \infty$ $\sigma > 0$	$f(x) = 1/(\sigma x\sqrt{2\pi}) \times \exp(-(\ln x - \mu)^2/(2\sigma^2))$	$e^{\mu+\sigma^2/2}$	$(e^{\sigma^2} - 1)e^{2\mu+\sigma^2}$
Weibull	$\alpha > 0$, $\beta > 0$	$f(x) = \alpha\beta^{-\alpha} x^{\alpha-1} \times \exp(-(x/\beta)^{\alpha})$	$(\beta/\alpha) \times \Gamma(1/\alpha)$	$(\beta^2/\alpha) \times \left[\begin{matrix} 2\Gamma(2/\alpha) - \\ (1/\alpha)\Gamma(1/\alpha)^2 \end{matrix}\right]$
Normal	$-\infty < \mu < \infty$ $\sigma > 0$	$f(x) = 1/(\sqrt{2\pi\sigma^2}) \times \exp(-((x-\mu)^2/(2\sigma^2)))$	μ	σ^2

index plot, to identify the optimal threshold value. In the plot of the index, maximum likelihood estimators are plotted against the number k of upper ordered values. If k is appropriate for data, the values of the estimators stabilize around the true extremes and a plateau becomes visible.

Generalized Pareto distribution

The Generalized Pareto is a right-skewed distribution. If we consider an unknown distribution function F of a random variable X, we are interested in the behavior of large observations which exceed a high threshold. Given a high threshold u, the distribution of excess values of x over threshold u is defined by

$$F_u(y) = P(X - u \leq y \mid X > u) = \frac{F(y+u) - F(u)}{1 - F(u)} \quad (3)$$

which represents the probability that the value of x exceeds the threshold u by at most an amount y given that x exceeds the threshold u. Pickands (1975), Balkema and de Haan (1974) posed that for a large class of underlying distribution function F the conditional excess distribution function $F_u(y)$, for u large, is well approximated by

$$F_u(y) \approx G_{\xi,\sigma}(y), \quad u \to \infty,$$

where

$$G_{\xi,\sigma}(y) = \begin{cases} 1 - (1 + \frac{\xi}{\sigma} y)^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - e^{-y/\sigma} & \text{if } \xi = 0 \end{cases} \quad (4)$$

For $y \in [0, (x_F - u)]$ if $\xi \geq 0$ and $y \in [0, -\frac{\sigma}{\xi}]$ if $\xi < 0$. $G_{\xi,\sigma}$ is the so-called generalized Pareto distribution where ξ is the shape parameter and σ is the scale parameter. To apply such formula, we will use an approach based on the GPD approximation of the over-threshold losses. The ultimate objective for determining the best-fit distribution of a risk is therefore to help stabilize the overall loss results of a group of risks during an accounting period.

Other right-skewed distributions and normal distribution

Several different distributions can often fit the same data set. If the histogram indicates a right-skewed data set, the recommended right-skewed distribution could be gamma, lognormal, or Weibull distributions and the statistics for the normal distribution are evaluated. We summarize the property of gamma, lognormal, Weibull and normal distributions in Table 1. We have chosen the same threshold value which is the best-fitting distribution for rice damaged by typhoons. Each of these four distributions was an appealing candidate to provide a good approximation. The following table lists the four distributions used.

Test of the appropriateness of model selection

Once the calculated sample of losses and approximation distributions were constructed, we tested the goodness-of-fit. While the usual deviation tests (K-S test and A-D test) provide a general measurement of how close above distributions. In this paper, we conduct K-S test and A-D test to test of the appropriateness of model selection. The K-S test has the advantage of making no assumption about the distribution of data (Technically speaking it is non-parametric and distribution free) which is based on the largest vertical difference between the theoretical and the empirical cumulative distribution function. The A-D test is a modification of the

Table 2. Summary Statistics for Rice Loss in Taiwan (n = 117), 1971 - 2005 Unit: NTD$1,000.

Sample Size	Mean	Standard Deviation	Minimum	25% quartile	50% quartile	75% quartile	Maximum	Skewness	Excess Kurtosis
117	197,732	482,417	21	6,506	24,849	185,190	3,863,707	5.4	35.3

Source: Taiwan Agriculture Year Book.

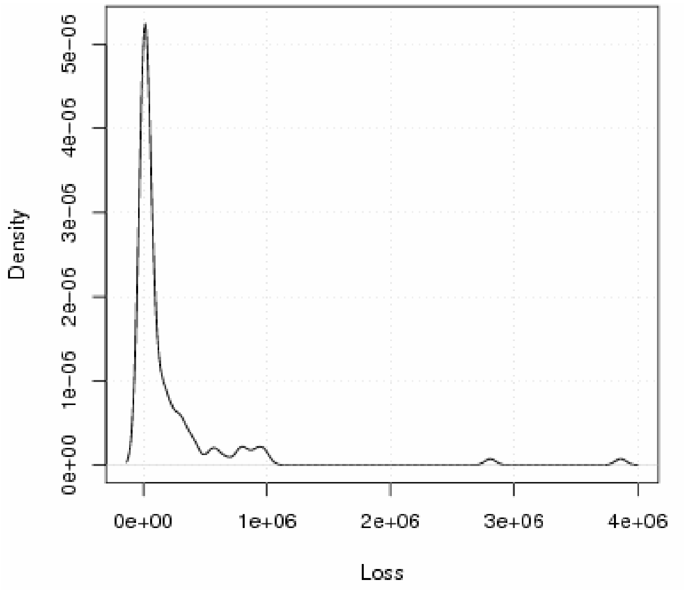

Figure 1. Plot of the Kernel Density for loss.

K-S test and gives more weight to the tails than does the K-S test. These two tests both show how well the distribution we selected fits to the data.

Kolmogorv-Smirnov statistic

Let $\mathbf{X}=(X_1,\cdots,X_n)$ be a random sample from some distribution with cumulative distribution function [CDF, $F(X)$]. The empirical CDF is denoted by $F_n(x)=n^{-1}\times\left[\text{number of observations}\le \text{x}\right]$. The Kolmogorov-Smirnov statistic (D) is based on the largest vertical difference between the theoretical and the empirical cumulative distribution function (Bain and Engelhardt, 1991):

$$D=\max_{1\le i\le n}\left(F(x_i)-\frac{i-1}{n},\frac{i}{n}-F(x_i)\right) \qquad (5)$$

If the test statistic, D , is greater than the critical value obtained from a table, we reject the data following the specified distribution.

Anderson-Darling statistic

The Anderson-Darling test is a form of minimum distance estimation, which assesses whether a sample comes from a specified distribution. The Anderson-Darling goodness of fit test is designed to detect differences in the tails between the fitted distribution and the data. The Anderson-Darling statistic (A^2) is defined as (Anderson and Darling, 1952; Luceno, 2006):

$$A^2=-n-\frac{1}{n}\sum_{i=1}^{n}(2i-1)\times\left[\ln F(X_i)+\ln(1-F(X_{n-i+1}))\right] \qquad (6)$$

If the test statistic, A^2 , is greater than the critical value obtained from a table, we reject the data following the specified distribution.

RESULTS AND DISCUSSION

In this paper we apply the method a dataset consisting of 117 inflation-adjusted typhoons losses in Taiwan, from published government statistics which is included in Taiwan Agricultural Yearbook, Production Cost and Income of Farm Products Statistics, and Taiwan Area Agricultural Products Wholesale Market Yearbook, which record rice losses due to major natural disasters in Taiwan over the years 1971 - 2005. Although there were 10 types of natural disasters damaged rice, majority (about 60%) of losses were caused by typhoons. Hence we only consider the rice loss data due to the typhoon disasters. We summarize the data characteristics for rice losses in Table 2. The table shows that, on average, loss amounts are nearly 20 million NT dollars per typhoon. The inter-quartile range is too large (about 180 millions NTD) and the data contain a significant number of very high losses (the maximum loss observed is 3,863,707 thousands NTD). The distribution of rice loss is considerably skewed to the right (here, the skewness coefficient is larger than 0). In Figure 1 we plot the Kernel density for losses. The right-skewness therein reveals the heavy-tailed behavior of the underlying data which might cast doubt on an assumption of a normal distribution. We have seen that, rice losses data are appropriate by using GPD modeling fitting, if enough data are available above a high enough threshold. Figure 2 is a plot of the GPD estimate $\hat{\xi}$ versus number k of extremes. It shows a

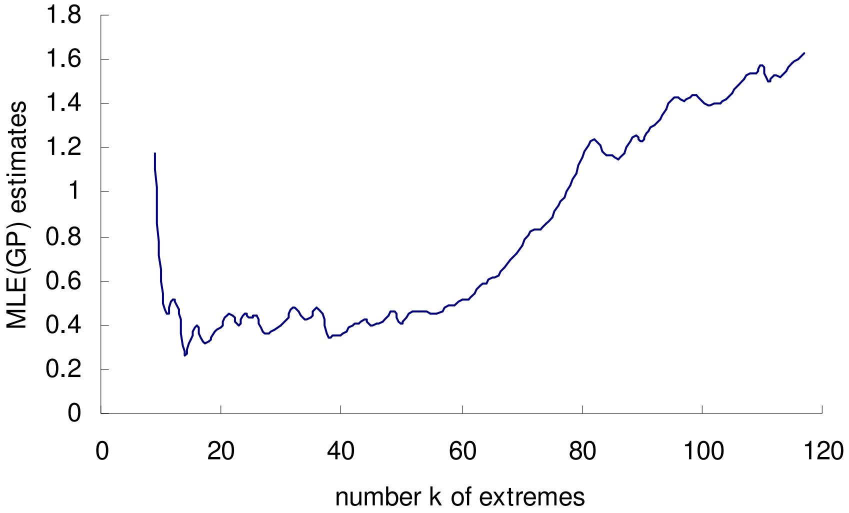

Figure 2. Plot of the GPD estimate $\hat{\xi}$ versus number k of extremes.

Table 3. Maximum Likelihood Estimates for Excess.

Threshold u	$\hat{\sigma}$	$\hat{\xi}$	LR Test (p-value)
30,502	155,247	0.4486	0.0002*

*Significant at the 1% level (P < 0.01)

Table 4. Summary of Goodness of Fit.

Distribution	Kolmogorov Smirnov		Anderson Darling	
	Statistic	Rank	Statistic	Rank
Gen. Pareto	0.04748	1	0.22227	1
Weibull	0.06242	2	0.40493	3
Lognormal	0.07501	3	0.27738	2
Gamma	0.21116	4	3.76200	4
Normal	0.27925	5	7.37670	5

plateau with right endpoint around k = 55. This evidence suggests the estimate k = 55, that is, 30,502 should be used in subsequent processing. The selection of optimal threshold may also be done in an automatic manner by “Xtremes” package and it gives the same results.

We then calculate the parameters in the GPD with the help of the “Xtremes” package. The results are summarized in Table 3. The likelihood ratio (LR) statistics in Table 3 shows that the p-value is equal to 0.0002, which is smaller than the significance level value of 0.05 (Reiss and Thomas, 1997). As the p-value is smaller than the significance levels, the GPD seems be a good choice. We furthermore use extreme value theory to determine the over-threshold peaks of the data and apply the K-S and A-D goodness of fit tests to show that the Generalized Pareto distribution fits the heavy-tailed distribution better than Normal and some common right-skewed distribution (Gamma, Lognormal and Weibull distributions), as shown in Table 4. To further analyze the models we also present the probability difference graph in Figure 3. The probability difference graph is a plot of the difference between the empirical CDF and the theoretical CDF and it can be used to determine how well the theoretical distribution fits the observed data. In our case, it shows that the Generalized Pareto distribution fits best, especially on tail behavior fitting. The heavy-tailed distribution characterizing these data tells us that the associated loss distribution is non-normal. Since, a problem is estimating the PML with extreme values for a particular class of natural disasters, for example rice damaged. It is difficult for government's agricultural organization to know where the actual risk is at all times, be it on landing a typhoon, for example a wind-force. One can then derive the PML for limited risk retention as costs of premium principles due to taking place crop insurance in future. In this approach, we can assess the best-fitted loss distribution and the threshold value of PML of risk retained by the government or to make crop insurance for adjusting the compensation mechanism of agricultural disasters in the government's substantial funds (cash succor, subsidy or/and loan of preferential interest rate) to farmers to compensation for rice loss resulting from natural disaster. We also can undoubtedly allow that there seems to be existing fat tail behavior of crop-damaged (e.g. paddy rice, vegetables, fruit and other special crops), leading to a great impact on budget

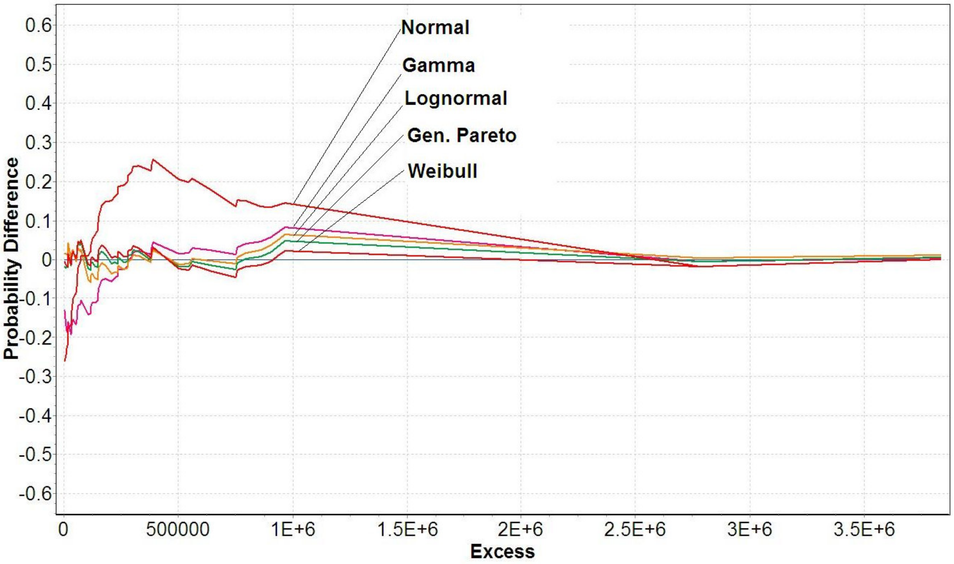

Figure 3. Probability Difference for Normal and some right-skewed distributions

making of national capacities for risk financing of major agricultural disasters (e.g. typhoon, floods, frost, hail, drought, disease, earthquake or wind) and disaster risk management programs (e.g. risk retention-cash succor, subsidy or/and loan of preferential interest rate, or and crop insurance coverage). Throughout the results, the government and the related institutions could command the losses of typhoon in the future and the potential frequency of the events every year, and they should be to grouping general and large losses and to taking advantage of the outcome to develop the efficient mechanism of compensating or to build the crop insurance which is suitable for the environment and the background of agriculture.

Conclusion

In this paper we examine the loss distribution of rice damaged by typhoons in Taiwan. We see that the loss distribution of rice damage due to typhoons is considerably skewed to the right, and is non-normal. We furthermore use extreme value theory to determine the over-threshold peaks of the data and apply the K-S and A-D goodness of fit tests to show that the generalized Pareto distribution fits the heavy-tailed distribution better than the Lognormal, Gamma, Weibull and Normal distributions. It is important to note that our empirical results show that the relation linking of the best-fitted loss distribution, the threshold value and the probable maximum loss (PML). The government's agricultural natural disaster assistance program and crop insurance have been the subject of change or/and considerable research in recent years, but the implication and impact of such programs by government's budgetary outlays for the threshold value of PML resulting from natural disaster have been neglected. On the important consequence of such an appropriate calculation of effective threshold value and probable maximum loss are calculated as the risk retention associated with the systematic risk of major agricultural disasters, offset by fiscal revenues, which is guaranteed by the government supporting and enabling the development of large loss or catastrophe insurance market through public policies, and assistance in building the national institutions of large loss or catastrophe risk management – as an integral function of financial risk management in agricultural under all stakeholder (farmer, governments, insurer and reinsurer) need to cooperation agreement (Larsson, 2005; Muralidharan, 2006). The properties we found are useful in loss assessment of crop loss in the decision making of national capacities for risk financing of major agricultural disasters and disaster risk management programs.

ACKNOWLEDGMENTS

The authors would like to acknowledge the helpful comments of the reviewers.

REFERENCES

Adams RM, Hurd BH, Lenhart S, Leary N (1998). Effects of global climate change on agriculture: an Interpretative review. Climate Res. 11: 19-30.

Aggarwal PK, Kalra N, Chander SH, Pathak H (2006). InfoCrop: a dynamic simulation model for the assessment of crop yields, losses due to pests, and environmental impact of agro-ecosystems in tropical Environments. Agric. Syst. 89: 1-25.

Anderson TW, Darling DA (1952). Asymptotic distribution of certain "goodness-of-fit" criteria based on stochastic processes. Annals of Mathematical Stat. 23**:** 193-212.

Balkema AA**,** Haan L (1974). Residual lifetime at great age. Ann. Probab. 2: 792-804.

Bain LJ, Engelhardt M (1991). Introduction to probability and mathematical statistics. 2nd ed. Boston: PWS-KENT.

Brabson BB, Palutikof JP (2000). Test of the generalized Pareto distribution for predicting extreme wind speeds. Am. Meteorol. Soc. 39: 1627-1640.

Chang CC (2002). The potential impact of climate change on Taiwan's agriculture. Agric. Econ. 27: 51-64.

Embrechts P, Resnick SI, Samorodnitsky G (1999). Extreme value theory as a risk management tool. North Amer. Actuarial J. 3: 30-41.

Hansen LM (2004). Economic damage threshold model for pollen beetles (Meligethes aeneus F.) in spring oilseed rape (Brassica napus L.) crops. Crop Prot. 23: 43-46.

Hosking J, Wallis J (1987). Parameter and quantile estimation for the generalized Pareto distribution. Technol. 29: 339-349.

Ji QZ, Hayakawa S, Yamamoto H, Okada N, Tatano H (2002). Comparisons of agricultural damages by typhoon 9917, 9919 and 9918. Jpn. J. Crop Sci. 71 (2) : 239-249.

Lansigan FP, De los Santos WL, Coladilla JO (2000). Agronomic impacts of climate variability on rice production in the Philippines. Agric. Ecosyst. Environ. 82: 129-137.

Larsson HA (2005). A crop loss model and economic thresholds for the grain aphid, Sitobion avenae (F.), in winter wheat in southern Sweden. Crop Prot. 24: 397–405.

Luceno A (2006). Fitting the generalized Pareto distribution to data using maximum goodness-of-fit estimators. Computational Statistics and Data Analysis. 51: 904 – 917.

Luo Y, Teng PS, Fabellar NG, Tebeest DO (1998). Risk analysis of yield losses caused by rice leaf blast associated with temperature change above and below for five Asian countries. Agric. Ecosyst. Environ. 68: 197-205.

McNeil AJ (1997). Estimating the Tails of Loss Severity Distributions using Extreme Value Theory. ASTIN Bulletin 27: 117-137.

Muralidharan K, Pasalu IC (2006). Assessment of crop losses in rice ecosystems due to stem borer damage (Lepidoptera: Pyralidae). Crop Prot. 25: 409-417.

Pickands J (1975). Statistical Inference using Extreme Order Statistics. Ann. Stat. 3:119-131.

Prudhomme C, Jakob D, Svensson C (2003). Uncertainty and climate change impact on the flood regime of small UK catchments. J. Hydrol. 277:1-23.

Reiss RD, Thomas M (1997). Statistical Analysis of Extreme Values, Basel: Birkhäuser Verlag.

Rosenzweig C, Tubiello FN, Goldberg R (2002). Increased crop damage in the US from excess precipitation under climate change. Global Environ. Change. 12: 197-202.

Silverman BW (1986). Density Estimation for Statistics and Data Analysis. London: Chapman and Hall.