

Análise da probabilidade de perda no rendimento da soja: implicações para o seguro agrícola

Priscila N. Faria¹, Ricardo A. Olinda², Vitor A. Ozaki³ e Rogério C. Campos⁴.

Introdução

No Brasil, historicamente, o seguro agrícola não conseguiu se popularizar, diversos problemas inibiram o desenvolvimento desse mercado, entre eles destacam-se: alto custo do seguro, elevado risco sistêmico, seleção adversa e risco moral. Nesse contexto, segundo uma recente pesquisa feita pelo IBGE, o Paraná recuperou a posição de maior produtor brasileiro de grãos com a safra nacional de grãos 2009/10.

A suposição de normalidade dos dados de produtividade agrícola é usualmente aceita pelo mercado segurador na quantificação e precificação do risco. No entanto, a distribuição normal é incapaz de captar assimetrias e bimodalidades que possam estar presentes nos dados (FERNANDEZ, 2003).

A forma da distribuição é particularmente importante no contexto de estudos de seguros de cultura, porque ele reflete o risco (probabilidade de perda) do produtor. Em outras palavras, quando se trata da modelagem agrícola da produção deve-se olhar a massa concentrada na cauda esquerda da distribuição (Ozaki et al., 2008).

Com base nessas informações, o objetivo do trabalho será utilizar a teoria dos valores extremos para modelar a cauda esquerda da distribuição de probabilidade da produtividade agrícola dos dez maiores municípios produtores de soja do estado do Paraná.

Metodologia

A teoria de valores extremos tem grande aplicação na área de finanças, em análise de risco, que busca estimar perdas futuras através da modelagem do comportamento menos frequente, de perdas associadas a certos eventos (PINTO, 2002).

¹Professor e Coordenador Administrativo do Grupo de Estudos em Risco e Seguro - GESER, Departamento de Economia/ Universidade de São Paulo. Av. Pádua Dias, 11 Agronomia / CEP 13418-900–Piracicaba/SP–o Brasil. E-mail: vaozaki@esalq.usp.br

²Estudante de doutorado em Estatísticas Aplicada e integrante do GESER / Universidade de São Paulo

³ Estudante de doutorado em Estatísticas Aplicada e integrante do GESER / Universidade de São Paulo

⁴Estudante de Pós-doutorado e Coordenador Científico do Risco Grupo de Estudos em Risco e Seguro - GESER, Departamento de Economia/ Universidade de São Paulo.

Os autores agradecem Grupo de Estudos em Risco e Seguro (GESER/USP) e o Centro de Estudos Avançados em Economia Aplicada (CEPEA/USP) pela ajuda deles/delas nesta pesquisa. Este manuscrito conta com apoio financeiro da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

A parte fundamental de um modelo de 'Value at risk' é a escolha da distribuição de probabilidades para os retornos dos fatores de risco (ISHIBASHI et al., 2004). Os modelos probabilísticos são obtidos através da aplicação da Teoria dos Valores Extremos (TVE) e visam prever, a partir de um conjunto de valores mínimos num processo registrado num período relativamente curto, os valores mínimos esperados num período maior de tempo.

Conceitos Básicos

Suponha uma amostra $\mathcal{X}_n = \{X_1, X_2, \dots, X_n\}$. Chama-se de estatísticas de ordem os elementos de \mathcal{X}_n ordenados de forma que $X_1 \leq X_2 \leq \dots \leq X_n$. Denotam-se as observações ordenadas, de agora em diante, de estatísticas de ordem, da forma $X_{\langle 1 \rangle} \leq X_{\langle 2 \rangle} \leq \dots \leq X_{\langle n \rangle}$. Denominando-se o máximo amostral por: $M_1 = X_1$ e $M_n = \max(X_1, X_2, \dots, X_n)$, para $n \geq 2$, supondo \mathcal{X}_n i.i.d. escrever-se a distribuição exata do máximo como $\Pr\{M_n \leq x\} = \Pr\{X_1 \leq x, \dots, X_n \leq x\} = \Pr\{X_1 \leq x\} \Pr\{X_2 \leq x\} \dots \Pr\{X_n \leq x\} = F^n(x)$ onde $F(x) = \Pr\{X \leq x\}$, sugerindo uma lei de potência para a distribuição do máximo. O resultado central da TVE é enunciado pelo Teorema de Fisher-Tippett (também conhecido como Lei de Limite para o Máximo).

Teorema de Fisher-Tippett (Fisher and Tippett, 1928): Suponha (X_n) uma sequência de variáveis aleatórias i.i.d. Se existem constantes de padronização, $c_n > 0$, $d_n \in \mathfrak{R}$, e alguma função de distribuição não-degenerada H tal que $\lim_{n \rightarrow \infty} \Pr\{c_n^{-1}(M_n - d_n) \leq x\} = \lim_{n \rightarrow \infty} F_x^n(c_n x + d_n) \rightarrow H(x)$ então $H(x)$ pertence a algum dos três tipos de funções de distribuição de valores extremos.

$$\text{Fréchet: } \Phi_\alpha(x) = \begin{cases} 0, & \text{se } x \leq 0 \\ \exp(-x^{-\alpha}), & \text{se } x > 0 \end{cases} \quad \alpha > 0$$

$$\text{Weibull: } \Psi_\alpha(x) = \begin{cases} \exp(-x^\alpha), & \text{se } x \leq 0 \\ 1, & \text{se } x > 0 \end{cases} \quad \alpha > 0$$

$$\text{Gumbel: } \Lambda(x) = \exp(-e^{-x}), \quad x \in \mathfrak{R}$$

GEV (*General Extreme Value distribution*)

A pergunta relevante neste ponto deve ser como identificar a correta distribuição do extremo de uma dada amostra, entre as três possíveis. É possível introduzir uma família paramétrica que englobe as três distribuições, e o problema recai simplesmente na estimação do parâmetro de interesse $\xi = \frac{1}{\alpha}$ (índice de caudas ou *tail index*) que dará o formato da distribuição:

$$\text{Definindo a família: } H_{\xi} = \begin{cases} \Phi_{1/\xi}, & \text{se } \xi > 0 \\ \Lambda, & \text{se } \xi = 0 \\ \Psi_{-1/\xi}, & \text{se } \xi < 0 \end{cases}$$

Tem-se que, para qualquer F pertencente ao domínio de atração do máximo de H_{ξ} , vale a aproximação $F^n(c_n x + d_n) \approx H_{\xi}$ para a distribuição do extremo, sendo ξ a cauda da distribuição.

Além disso, é possível substituir-se x na expressão acima pela transformação de escala e locação $\frac{x - \mu}{\sigma}$ com $\mu \in \mathfrak{R}$ e $\sigma > 0$. Dessa forma produz-se a distribuição

$$H_{\xi; \mu, \sigma} = \exp \left\{ - \left(1 + \xi \frac{x - \mu}{\sigma} \right)^{-1/\xi} \right\} \text{ denominada de } GEV \text{ (General Extreme Value}$$

distribution), definida em, $-\infty < x < (\mu - \sigma)/\xi$ para $\xi < 0$, $-\infty < x < +\infty$ para ξ tendendo a zero, $(\mu - \sigma)/\xi < x < +\infty$ para $\xi > 0$, sendo μ , σ e ξ , respectivamente, os parâmetros de locação, escala e de forma, com $\sigma > 0$. A partir da GEV obtém-se a função densidade de probabilidade da distribuição GVE, dada por

$$f(x) = \frac{1}{\sigma} \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-\left(\frac{1+\xi}{\xi}\right)} \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\}, \quad \text{definida em,}$$

$-\infty < x < (\mu - \sigma)/\xi$ para $\xi < 0$, e $(\mu - \sigma)/\xi < x < +\infty$ para $\xi > 0$, em que x é a variável aleatória associada a valores máximos do período.

Resultados e discussão

Como primeira etapa, foi realizada uma análise exploratória da variável aleatória Rendimento médio da produção (em Quilogramas por Hectare), verificando-se que a

mediana é sistematicamente maior do que a média (exceto para os municípios de Cascavel e Toledo), o que sugere que as distribuições sejam assimétricas à esquerda, apresentados na Tabela 1.

Tabela 1. Estatísticas descritivas da variável aleatória Rendimento médio da produção (em kg/ha), nos períodos de 1980 a 2008 em dez municípios do estado do Paraná.

Cidades	Média	Mediana	Desvio padrão	C.A.	C.V.
Cascavel	2.412,667	2398	4.454,629	0,107	18,464
Tibagi	2.611,074	2650	4.501,075	-0,061	17,238
Castro	2.702,741	2800	3.951,379	0,005	14,620
Assis Chateaub.	2.392,741	2479	4.441,380	-0,145	18,562
Toledo	2.504,963	2398	5.831,576	-0,137	23,280
Ponta Grossa	2.618,815	2700	4.526,016	-0,111	17,286
Terra Roxa	2.147,222	2158	5.679,708	-0,057	26,451
Mamborê	2.450,704	2592	4.579,311	-0,360	18,686
Ubiratã	2.531,444	2603	4.014,807	-0,432	15,859
Campo Mourão	2.410,630	2564	4.591,812	-0,098	19,048

A Figura 1 apresenta os gráficos de caixa (box-plot) da variável Rendimento médio da produção, para cada um dos dez municípios em questão.

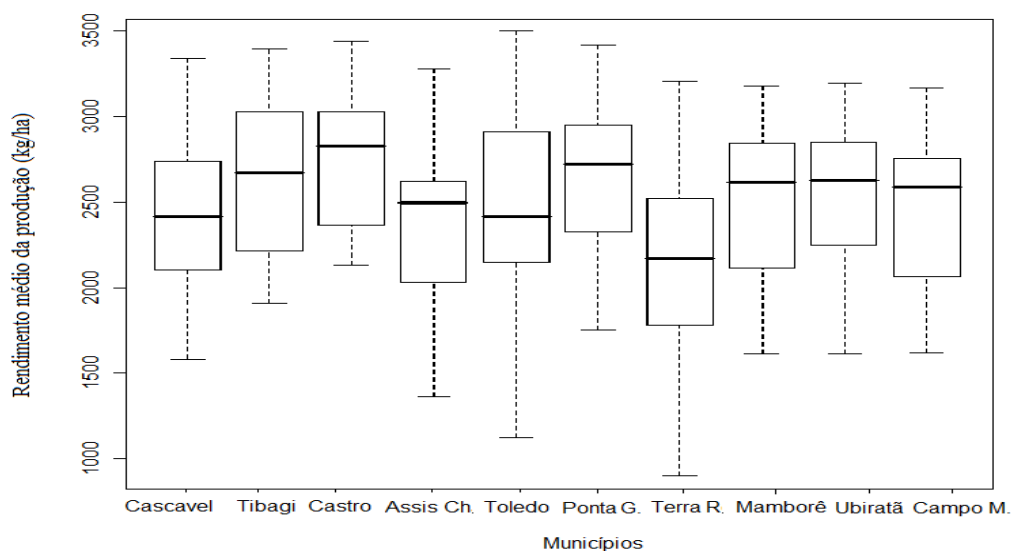


Figura 1. Gráficos de caixa (box-plot) para variável velocidade máxima de vento para cada um dos dez municípios.

Nota-se assim que apesar do município de Terra Roxa apresentar um dos menores valores de média de Rendimento e menor mediana, possui uma das maiores dispersões, que pode ser visualizado na Figura 1 e quantificado por meio do coeficiente de variação (C.V.), apresentado na Tabela 1.

Analisando-se a Tabela 2, observa-se que todas as estimativas pontuais do parâmetro de forma (índice de caudas) apresentaram valores menores do que zero que corresponde à distribuição de Weibull. Esta distribuição, segundo Holmes & Moriarty (1999), é a mais apropriada para representar fenômenos ambientais (estiagem, precipitação máxima ou mínima, entre outros).

Tabela 2. Estimativa dos parâmetros de locação escala e forma.

Cidades	Locação	Escala	Forma
Cascavel	2257	429	-0.2737
Tibagi	2499	473	-0.4820
Castro	2587	400	-0.3943
Assis Chateaubriand	2252	447	-0.3535
Toledo	2345	604	-0.4296
Ponta Grossa	2499	471	-0.4502
Terra Roxa	1972	573	-0.3626
Mamborê	2352	513	-0.6063
Ubiratã	2447	433	-0.5593
Campo Mourão	2274	513	-0.5277

O fato dos dados serem modelados pela distribuição de Weibull foi confirmado com a aplicação do Teste da Razão de Verossimilhança Modificado (TRVM), com o valor tabelado da estatística qui-quadrado com um grau de liberdade e nível de 5% de significância (cujo valor tabelado é 3,84).

A Tabela 3 mostra as probabilidades de perdas dos municípios selecionados usando a distribuição de Weibull. Deve-se notar que o nível de cobertura é um parâmetro escolhido pelo segurado no início do contrato de seguro. Neste contexto, o produtor agrícola pode escolher os níveis de 60-80% do rendimento médio do município i (μ_i), resultando na produção crítica. Se o rendimento final, ao final da colheita, é menor do que a produção crítica, o segurado recebe a indenização.

Tabela 3. Probabilidades de perda dos municípios assumindo a distribuição de Weibull e considerando os níveis de cobertura de 60, 70 e 80%.

Cidades	Nível de convergência (α)		
	60%	70%	80%
Cascavel	0,01	0,045	0,136
Tibagi	0,018	0,053	0,127
Castro	0,004	0,023	0,088
Assis Chateaubriand	0,017	0,055	0,142

Toledo	0,051	0,104	0,191
Terra Roxa	0,067	0,129	0,221
Manbore	0,039	0,08	0,154
Ponta Grossa	0,017	0,051	0,127
Ubirata	0,017	0,047	0,113
Campo Mourao	0,04	0,087	0,169

Conclusão

Na presente investigação foram analisados pressupostos estatísticos alternativos para estimar a probabilidade de perda com dados de rendimento agrícola da SEAB. Com base nos resultados obtidos, os dados de séries de tempo de soja foram ajustados à distribuição de Weibull. Conclui-se, através da Tabela 3, que a taxa de prêmio puro é realmente menor que a taxa de prêmios cobrada.

Referências

FERNANDEZ, Viviana (2003). "Extreme Value Theory e Value at Risk", Revista de Analisis Economico - Economic Analysis Review, Ilades-Georgetown University, Economics Department, vol. 18(1), pages 57-85, June. p. 57-85, jun.

Fisher, R.A., L. H. C. Tippett (1928). Limiting forms of the frequency distribution of the largest and smallest member of a sample, Proc. Cambridge Phil. Soc., 24, 180-190.

Ishibashi, Regina et al. (2004). Estimación dos parâmetros da distribuição extremo aplicadas em finanças. In: XXVII Congresso Nacional de Matemática Aplicada e Computacional, Porto Alegre-RS, 2004, p 239.

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, 2008, v. 2, pp.79-94.

PINTO, Flávia C. Teoria de Valores Extremos - Aplicação ao Mercado Financeiro. 2002. Dissertação (Mestrado em Estatística), Instituto de Matemática e Estatística/USP, São Paulo, 2002.