**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

**Precificação do seguro agrícola: novas abordagens utilizando
distribuições de probabilidades alternativas e o uso de cópulas
bidimensionais e tridimensionais**

Gislaine Vieira Duarte

Tese apresentada para obtenção do título de Doutora
em Ciências. Área de concentração: Estatística e Ex-
perimentação Agronômica

**Piracicaba
2018**

Gislaine Vieira Duarte
Licenciada em Matemática

Precificação do seguro agrícola: novas abordagens utilizando distribuições de probabilidades alternativas e o uso de cópulas bidimensionais e tridimensionais

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:

Prof. Dr. **VITOR AUGUSTO OZAKI**

Tese apresentada para obtenção do título de Doutora em Ciências. Área de concentração: Estatística e Experimentação Agronômica

Piracicaba
2018

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP

Duarte, Gislaine Vieira

Precificação do seguro agrícola: novas abordagens utilizando distribuições de probabilidades alternativas e o uso de cópulas bidimensionais e tridimensionais/ Gislaine Vieira Duarte. -- versão revisada de acordo com a resolução CoPGr 6018 de 2011. -- Piracicaba, 2018 .
89 p.

Tese (Doutorado) -- USP / Escola Superior de Agricultura "Luiz de Queiroz".

1.Cópulas paramétricas bidimensionais e tridimensionais 2.Distribuição bimodal 3.Gerenciamento de riscos agrícolas 4.Seguro de faturamento agrícola 5.Seguro de produtividade agrícola I. Título .

DEDICATÓRIA

Aos meus pais, pelo apoio incondicional em todos os momentos e que, por vezes renunciaram aos seus sonhos, para que eu pudesse realizar os meus. Muito obrigada por tudo.

AGRADECIMENTOS

Agradeço à Deus pelo dom da vida, por permitir e direcionar a realização deste sonho.

Aos meus pais, Valdir e Maria do Carmo, pela dedicação, cuidado, educação, ensinamentos e valores transmitidos a mim e minhas irmãs. Muito obrigado pai e mãe pelas conversas de encorajamento e incentivo em momentos diversos e de forma incondicional, mesmo à distância.

Agradeço ao meu esposo Eduardo pelo incentivo diário, paciência, cumplicidade, amor e dedicação durante todo o curso e compreensão nos momentos de ausência. Agradeço aos meus amados filhos Felipe (7 anos), uma criança que transborda alegria e energia e que vive intensamente as curiosidades da infância, e a minha pequena Sophia (11 meses), minha bonequinha presente de Deus. Obrigada filhos por serem fontes de inspiração e dedicação da mamãe.

As minhas irmãs Lisiane e Natália pelo apoio e incentivo, mesmo à distância. Aos meus sogros Regina e Alcides, por apoiar e cuidar do Felipe em alguns momentos de ausência. Aos meus queridos tios Toninho e Rosa pelo estímulo e hospitalidade em sua casa.

Agradeço ao meu orientador professor Vitor Augusto Ozaki pelos ensinamentos, pelas valiosas contribuições e inúmeras dicas durante todo o período de elaboração da tese.

Agradecimento especial aos professores do departamento de Estatística e Experimentação Agronômica da ESALQ pelos valiosos ensinamentos e contribuições a minha formação técnica e pessoal, em especial àqueles com quem estive em contato em sala de aula e seminários, Carlos Tadeu dos Santos Dias, César Gonçalves de Lima, Clarice Garcia Borges Demétrio, João Gomes Martines Filho, Idemauro A. Rodrigues de Lara, Roseli A. Leandro, Silvio S. Zocchi, Sônia M. De Stefano Piedade e Taciana V. Savian.

Aos professores Pedro Alberto Morettin, Clelia Maria de Castro Toloi, Nikolai Kolev e Gilberto de Paula pelo incentivo aos estudos e conhecimentos transmitidos em disciplinas cursadas no IME-USP.

Aos professores Rodrigo Lanna, Renata A. Sermarini e Cláudio Brisolara por participar da minha banca de qualificação e contribuir com sugestões válidas, críticas construtivas, pertinentes e importantes para meu trabalho.

Agradeço aos funcionários do Departamento de Estatística e Experimentação Agronômica da ESALQ, em especial a Solange Paes Sabadin, por sua paciência, gentileza e prontidão em ajudar em todos os momentos.

Aos meus colegas do grupo de pesquisa GESER, pela amizade e o conhecimento sobre seguro rural compartilhado ao longo destes anos, em especial ao Carlos Andrés Onate, Christopher S. Pádua e Daniel Miquelluti. Pois é galera, no final dá tudo certo!

Aos meus colegas e amigos de pós-graduação, obrigada pela amizade, pelas horas de estudos compartilhadas, momentos de diversão e boas risadas, em especial Altemir, Erasnilson (Eras), Daniel, Douglas, Maria Joseane, Rick, Simone, Patrícia, Valiana e Hiron.

À CAPES–Coordenação de Aperfeiçoamento de Pesquisa de Ensino Superior, pela concessão de bolsa para o desenvolvimento deste trabalho.

Finalmente, a todos que de alguma forma contribuíram direta ou indiretamente para a realização deste trabalho. Muito obrigada!

SUMÁRIO

RESUMO	7
ABSTRACT	8
LISTA DE FIGURAS	9
LISTA DE TABELAS	10
LISTA DE ABREVIATURAS E SIGLAS	12
1 Introdução	13
1.1 Histórico do Seguro Rural no Brasil	14
1.2 Seguro Rural nos Estados Unidos	16
1.3 Tipos de Seguros Agrícolas Comercializados no Brasil	19
Referências	21
2 Modelagem da Produtividade de Soja utilizando Distribuições Simétricas, Assimétricas e Bimodais: Implicações para o Seguro de Produtividade agrícola	23
Resumo	23
Abstract	23
2.1 Introdução	23
2.2 Revisão de Literatura	24
2.3 Seguro de Produtividade	26
2.4 Descrição dos dados	27
2.5 Metodologia	27
2.5.1 Modelo Normal	27
2.5.2 Modelo Beta	28
2.5.3 Modelo Skew-normal	28
2.5.4 Modelo Skew-T Student	28
2.5.5 Modelo Odd log-logística - F	30
2.5.6 Avaliação do modelo	31
2.5.7 Tendência na Série de Produtividade	32
2.6 Resultados	32
2.7 Discussões	33
2.8 Conclusões	37
Referências	38
3 Precificação dos Riscos do Seguro de Faturamento Agrícola utilizando Cópulas Paramétricas	43
Resumo	43
Abstract	43
3.1 Introdução	43
3.2 Revisão de Literatura	44
3.3 Seguro de Faturamento Agrícola	46
3.4 Descrição dos dados	46
3.5 Análise Paramétrica para Distribuição das Marginais	47
3.6 Cópulas	47
3.6.1 Estimadores Paramétricos da Função de Verossimilhança	49
3.6.2 Estimadores de Pseudo Máxima Verossimilhança - Método Semi-Paramétrico	50
3.6.3 Testes de ajuste aos dados (<i>Goodness-of-fit tests</i>)	50
3.6.4 Procedimento de Vuong para Comparação de Modelos Paramétricos de Cópulas	51
3.7 Resultados	51

3.7.1	Inferência semi-paramétrica	53
3.7.2	Inferência Paramétrica	54
3.7.3	Comparação entre modelos	55
3.8	Taxa do Seguro de Faturamento	57
3.9	Conclusões	59
	Referências	59
4	Modelagem da Estrutura de Dependência entre Séries Temporais Multivariadas: Aplicação a Precificação dos Riscos do Seguro de Faturamento Agrícola	63
	Resumo	63
	Abstract	63
4.1	Introdução	63
4.2	Taxa do Prêmio do Seguro de Faturamento	64
4.3	Descrição dos dados	65
4.4	Metodologia	65
4.4.1	Modelagem da Produtividade	66
4.4.2	Modelagem do Preço e Dólar	66
4.4.2.1	Modelos GARCH	67
4.4.3	Cópulas	67
4.4.3.1	Inferência das Cópulas Multivariadas	69
4.4.3.2	Inferência via Pseudo Máxima-Verossimilhança	70
4.4.4	Ferramenta Gráfica para Detectar Dependência - Chi-plot	71
4.4.5	Teste de independência multivariada	71
4.5	Resultados	72
4.5.1	Produtividade	73
4.5.2	Análise Bivariada para o Contrato de Faturamento	74
4.5.3	Análise Tridimensional para o Contrato de Faturamento	76
4.6	Discussão	82
4.6.1	Contrato de Faturamento Bidimensional	82
4.6.2	Contrato de Faturamento Tridimensional	83
4.7	Conclusão	84
	Referências	85
	ANEXO	87
	Apêndice	89

RESUMO

Precificação do seguro agrícola: novas abordagens utilizando distribuições de probabilidades alternativas e o uso de cópulas bidimensionais e tridimensionais

A atividade agrícola é uma das atividades mais importantes para o ser humano, pois é fonte de matéria prima, alimentação e energia. No entanto, os principais grandes riscos desta atividade são os de produção e de mercado. A maneira mais popular de gerenciamento destes riscos são os seguros agrícolas. O seguro de produtividade agrícola gerencia o risco de quebra de produção relacionado às adversidades climáticas. Por sua vez, o seguro de faturamento agrícola é uma maneira de gerenciar os riscos de produção e de mercado (preços) conjuntamente. Portanto, este trabalho tem por objetivo apresentar métodos alternativos para precificação dos riscos de produtividade e faturamento agrícola. Na modelagem da distribuição de produtividade utilizou-se distribuições paramétricas que capturam a simetria, a assimetria e a bimodalidade dos dados, características estas geralmente encontradas em produtividades brasileiras. Além disso, a metodologia de cópulas foi utilizada na análise multivariada entre produtividade e preço. Esta metodologia apresenta um ganho significativo para estudos de problemas multivariados comparados à distribuição normal multivariada, pois pode-se levar em consideração estruturas de dependências linear, não linear e dependência somente nas caudas da distribuição multivariada. Além disso especifica qualquer tipo de distribuição para as distribuições marginais. Ademais, analisou-se a estrutura de dependência entre produtividade, preços futuros de contrato de soja negociados na Chicago Mercantile Exchange (CME) e a cotação do dólar. Esta análise tridimensional do problema do seguro de faturamento é inédita e foi utilizada pelo fato da soja produzida no Brasil ser exportada e os produtores utilizarem como referência internacional de preços a bolsa americana. Além disso, esta análise tridimensional via cópulas considera a variabilidade e a influência do dólar e o preço futuro da *commodity* na modelagem do seguro. Todas as análises foram discutidas e comparadas com as taxas dos seguros aplicadas pelas seguradoras brasileiras. No caso do seguro de produtividade e faturamento (tridimensional), os resultados sugerem que as taxas cobradas pelas seguradoras estão superfaturadas quando comparadas com a metodologia apresentada. A superestimação da taxa dificulta a expressiva venda de seguros no Brasil, além de atrair agricultores com maiores riscos, fortalecendo o problema de seleção adversa. No caso do seguro de faturamento (bidimensional), em que não se leva em consideração a influência do câmbio (dólar) na modelagem, os resultados sugerem que as seguradoras subestimam as taxas do seguro de faturamento, o que pode levar a uma perda grande para a seguradora, pois esta pode estar considerando um risco muito menor do que deveria ser levado em consideração. Na realização das análises foi utilizado o software R.

Palavras-chave: Cópulas paramétricas bidimensionais e tridimensionais; Distribuição bimodal; Gerenciamento de riscos agrícolas; Seguro de faturamento agrícola; Seguro de produtividade agrícola

ABSTRACT

Pricing agricultural risks: new approaches using alternative probability distributions and the use of two-dimensional and three-dimensional copulas

The agricultural activity is one of the most important activities for humans, as it is a source of raw materials, food and energy. However, this activity presents two major risks: production and market. Agricultural insurance is the most efficient way of managing these risks. Insuring agricultural productivity allows to manage the risk of crop failure due to weather adversities. However, revenue insurance is a way to manage both the production and the market (prices) risks simultaneously. This study aims to present alternative pricing methods risks of agricultural productivity and billing. In modeling the distribution of yield, parametric distributions that capture symmetry, asymmetry and bimodality of the data were used, features usually found in yield in Brazil. In addition, the coupling methodology was used in multivariate analysis between yield and prices. This methodology offers an exceptional gain of multivariate problems compared to the Multivariate Normal distribution, as it takes into account structures of linear and non-linear dependencies and dependency only on the tails of multivariate distribution, in addition to specifying any distribution type for marginal distributions. In addition, we analyzed the dependence structure between yield soybean, future prices traded at the Chicago Mercantile Exchange(CME) and, exchange rate (USD/ R \$). The three-dimensional analysis of insurance billing via copulas considers variability and influence of dollar and price of commodities exported and traded at CME. All analyses were discussed and compared with insurance rates applied by insurance companies. All analysis are conducted using the R software.

Keywords: Two-dimensional and three-dimensional parametric copulas; Bimodal distribution; Agricultural risk management; Revenue insurance; Yield insurance

LISTA DE FIGURAS

1.1	Quantidade de apólices vendidas em 2017 (MAPA, 2017).	20
2.1	Produção de soja no Paraná em 2014.	27
2.2	Gráfico da fdp da SN para diferentes parâmetros.	29
2.3	Gráfico da fdp da Skew-t para diferentes valores dos parâmetros.	29
2.4	Gráficos da fdp da OLLN para os diferentes valores dos parâmetros.	31
2.5	Série de Produtividade Original e Corrigida.	34
2.6	Distribuições Ajustadas para as séries corrigidas.	35
3.1	Séries Original e Corrigida da Produtividade e Preços.	52
3.2	Distribuições ajustadas para produtividade corrigida em Toledo e a série de preço SEAB.	55
4.1	Séries Originais de Produtividade e Preços.	73
4.2	Preço Original em reais, Preço Ajustado(R\$), Histograma e QQ-Plot Normal para a série de preço de ajuste em reais.	76
4.3	Gráficos das funções ACF e PACF para a Série de Preço (R\$) ajustada e estas ao quadrado.	77
4.4	Preço, Histograma e QQ-Plot Normal para a série de preço de ajuste diário da CME.	78
4.5	Preço, Histograma e QQ-Plot Normal para a série de cotação do dólar.	78
4.6	Gráficos das funções ACF e PACF para os Preços Ajustados da CME e dolar.	79
4.7	Gráficos das funções ACF e PACF para os Preços Ajustados da CME e dolar ao quadrado.	80
4.8	Diagrama de Dispersão.	80
4.9	Chi-Plot para os pares de variáveis (Produtividade em Toledo, CME)	81
4.10	Dependograma.	81

LISTA DE TABELAS

1.1	Percentuais de subvenção adotados em 2016 para o seguro agrícola, de acordo com MAPA (2017b)	14
1.2	Tipos de seguros mais comuns nos Estados Unidos.	18
1.3	Tipos de seguros agrícolas ofertados no Brasil.	19
2.1	Estatísticas descritivas para as séries corrigidas.	33
2.2	Estatísticas e critério de informação para os modelos selecionados.	33
2.3	Taxa Pura (em %) - Cascavel.	36
2.4	Taxa Comercial (em %) com margem de segurança de 20% para os municípios de Cascavel, Guarapuava, Castro e Palmeira.	37
3.1	Resultados do ajuste do modelo linear para retirada da tendência.	53
3.2	Resultados do teste Ljung-Box para as séries de produtividade e preço e estas ao quadrado, cuja hipótese nula é independência entre as observações ao longo do tempo e a homogeneidade das variâncias.	53
3.3	Estimativas para os parâmetros das cópulas (erro padrão entre parênteses).	54
3.4	Critérios de seleção AIC e BIC para cópulas paramétricas com marginais empíricas (ME).	54
3.5	Estatísticas A^* e W^* e os critérios de informação AICc, BIC para as séries univariadas.	55
3.6	Critérios de seleção AIC, BIC, estatística S_n para cópulas paramétricas com marginais paramétricas (MP)	56
3.7	Modelos selecionados por inferência semi-paramétrica (ISP) e paramétrica (IP), com parâmetro de dependência (δ).	56
3.8	Resultados do teste de ajuste ao modelo para o município de Castro x Preço por ISP e IP	57
3.9	Resumo dos modelos selecionados para os diferentes municípios por SPI e PI.	57
3.10	Taxa Pura para o prêmio do seguro de faturamento calculada pela cópula selecionada e Normal Bivariada para todos os municípios.	58
3.11	Taxas Comerciais Médias (%) do Prêmio de Faturamento usando a cópula, Normal Bivariada e a seguradora A.	58
4.1	Estatísticas Descritivas para as séries originais de preços CME (pcme), cotação do dólar (pdolar) e preços em reais (preais) e produtividades para Cascavel e Toledo.	72
4.2	Estatísticas Descritivas para as séries de produtividades corrigidas para Cascavel e Toledo (n=37 observações).	74
4.3	Parâmetros estimados da simulação da distribuição $OLLN(\mu, \sigma, \alpha)$, com erros padrões entre parênteses,	74
4.4	Estatísticas Descritivas para as séries de produtividades simuladas para Cascavel, Toledo e preços.	75
4.5	Valores dos critérios de AIC e BIC e do máximo da log-verossimilhança para os modelos ajustados à série de preço em reais.	75
4.6	Valores do critérios de AIC e BIC para Seleção de Cópulas.	76
4.7	Estatísticas Descritivas para as séries ajustadas de preços (US\$) e cotação do dólar(R\$).	77
4.8	Resultados dos teste de Ljung-Box para as séries de preços ajustados da CME (rcme) e dólar (rdolar) e estas ao quadrado (rcme ² e rdolar ²).	78
4.9	Valores dos critérios de AIC e BIC e do máximo da log-verossimilhança para os modelos GARCH(m,n) com erros t-Student, às séries da CME e Dolar,	79

4.10 Critérios AIC, BIC e máxima log-verossimilhança para seleção de cópulas.	82
4.11 Modelos selecionados para a precificação do seguro de faturamento com os parâmetros de dependência.	82
4.12 Taxa Pura e Comercial para o prêmio do seguro de faturamento bidimensional para os municípios de Toledo e Cascavel utilizando a cópula Frank ($\delta = -0,3358$).	83
4.13 Taxa Comercial média para o seguro de faturamento usando a cópula e a da seguradora A para Toledo (Cascavel).	83
4.14 Taxas pura e comercial (20%) de carregamento para o seguro de faturamento agrícola utilizando a copula t-Student para (Produtividade Toledo, Preço CME, Dólar)	84
4.15 Taxas médias comerciais para o seguro de faturamento agrícola utilizando a copula t-Student e a seguradora A.	84

LISTA DE ABREVIATURAS E SIGLAS

CME	Chicago Mercantil Exchange
N	Distribuição Normal
SN	Distribuição Skew-normal
OLLN	Distribuição Odd log logística normal
ST	Distribuição Skew-t Student
NC	Nível de Cobertura
TP	Taxa Pura do Seguro
TC	Taxa Comercial do Seguro
AMH	Cópula Ali-Mikhail-Haq
Kg/ha	Quilograma por hectare
IGP-DI	Índice Geral de Preços- Disponibilidade Interna
IPARDES	Instituto Paranaense de Desenvolvimento Econômico e Social
IPEADATA	Base de Dados Macro Econômicos e Regionais do Instituto de Pesquisa Econômica Aplicada (Ipea)
ISU	Iowa State University
MAPA	Ministério da Agricultura, Pecuária e Abastecimento
fdp	Função densidade de probabilidade
fda	Função de distribuição acumulada

1 INTRODUÇÃO

A atividade agrícola é uma das atividades mais importantes para o ser humano, pois é fonte de matéria prima, alimentação e energia. No entanto, esta atividade apresenta basicamente cinco grandes riscos: riscos de produção, mercado, financeiro, legal e humano. Basicamente, os riscos de produção estão relacionados às adversidades climáticas, pragas, incêndio, doenças ou uso inadequado de tecnologias. Por sua vez, os riscos de mercado estão associados às oscilações dos preços de insumos e produtos agrícolas. Os riscos financeiros estão relacionados ao custo financeiro do endividamento do produtor, devido à necessidade do capital de giro para manter a operação agrícola. Por outro lado, o risco legal está relacionado à não observância da legislação quanto aos tributos, normas, restrições ambientais e regulamentações sanitárias. Por fim, o risco humano está relacionado a erros cometidos pelo produtor ou empregados, por exemplo, acidentes de trabalho, exposição a defensivos agrícolas e combustíveis inflamáveis, inadequada operação de máquinas e implementos. No entanto, os riscos de produção e de mercado são os que apresentam maior impacto na renda do produtor, pois eles não estão sob controle dos produtores rurais. Portanto, são os riscos mais interessantes de serem estudados e gerenciados (BRISOLARA, 2013).

Há basicamente dois mecanismos de gerenciamento de riscos na agricultura: público e privado. No gerenciamento público, os governos atuam com serviços de educação, sistemas de gerenciamento de pragas e doenças, construção de infraestrutura, prorrogação ou perdão de dívidas, transferência de recursos diante de catástrofes, entre outros. Por outro lado, o gerenciamento de riscos privado apresenta as opções de seguro rural, contratos a termo de comercialização, contratos futuros e de opções (BRISOLARA, 2013), (BUANAIN and SILVEIRA, 2017).

No Brasil, o seguro rural é destinado à cobertura de riscos peculiares às atividades agrícola, pecuária, aquícola e florestal, cobertura do patrimônio dos produtores, seus produtos, o crédito para comercialização desses produtos, além do seguro de vida dos produtores. No entanto, a modalidade de seguro agrícola é a que mais tem crescido ao longo dos anos.

De acordo com MAPA (2017b), o número de produtores rurais atendidos pelo seguro rural passou de 16 mil para 48 mil entre os anos de 2006 a 2016. Neste mesmo período, a área segurada passou de 1,8 para 5,6 milhões de hectares, um crescimento de 220%.

O seguro agrícola garante ao produtor proteção contra intempéries climáticas, eventos naturais adversos e de mercado, o que o torna um dos principais mecanismos para o crescimento do setor agrícola. Este seguro é um grande avanço no gerenciamento dos principais riscos na agricultura: produção e mercado. No entanto, por apresentar alta exposição a riscos financeiros associados ao clima, doenças e preços, o desenvolvimento do seguro agrícola depende fortemente do apoio do governo, através de subvenções e do resseguro.

O programa de subvenção ao prêmio do seguro rural é um dos pilares da política agrícola brasileira e tem sido fundamental para o crescimento deste setor econômico. Segundo MAPA (2017b), a concessão de subvenção por parte do governo cresceu ao longo dos anos, passando de R\$ 31,16 milhões em 2006 para R\$ 398,58 milhões em 2016.

O Programa de Subvenção Econômica ao Prêmio do Seguro Rural (PSR) almeja estimular a contratação de seguros pelos produtores. Para isso, o governo federal paga uma porcentagem da taxa do prêmio enquanto a outra parte é paga pelo produtor ou subvencionado pelos estados. A porcentagem de subvenção federal depende da cultura, da modalidade do seguro, da região e das condições contratuais. Além disso, esses valores podem ser atualizados a cada ano.

Por exemplo, no ano de 2016, no estado do Paraná foram comercializadas 16.572 apólices (12.227 beneficiários) de seguros para a cultura da soja, com uma arrecadação de prêmio total de R\$ 140.699.840,35, deste valor R\$ 61.005.301,49 foram subsidiados pelo governo. O prêmio pago pelo produtor foi R\$ 79.694.538,84 e a subvenção média por produtor foi de R\$ 4.989,39. Além disso, a taxa média

do prêmio foi 6,4%, entretanto, a taxa efetiva paga pelo produtor rural foi 3,6% (MAPA, 2017b).

A Tabela 1.1 apresenta os percentuais de subvenção para o seguro agrícola adotados em 2016, de acordo com MAPA (2017b). Nota-se que, para trigo com cobertura multirrisco¹, 55% do valor do prêmio do seguro é pago pelo governo federal, já os outros 45% ficam a cargo do produtor e/ou estados (quando existir subvenção estadual) para o produtor que escolher um nível de cobertura maior do que 60%. No caso de grãos, com cobertura multirrisco para nível de cobertura de 70 – 75%, o valor subvencionado é de 40% do prêmio. No caso de grãos com riscos nomeados², independente do nível de cobertura, a porcentagem do prêmio paga pelo governo federal é de 35% e o restante é pago pelo produtor.

Tabela 1.1. Percentuais de subvenção adotados em 2016 para o seguro agrícola, de acordo com MAPA (2017b) .

Grupos de Atividades	Tipos de Coberturas	Nível de Cobertura	Subvenção(%)
Trigo	Multirrisco	> 60%	55%
Grãos	Multirrisco	60 – 65%	45%
	Multirrisco	70 – 75%	40%
	Multirrisco	> 80%	35%
	Riscos Nomeados	-	35%
Frutas, Olerícolas, Cafés e Cana-de-Açúcar	-	-	45%

As culturas que receberam maior aporte de recursos da subvenção no ano de 2016 foram soja, milho 2ª safra e trigo, com respectivamente, 42,1% (R\$ 167,90 milhões), 18,6% (R\$ 74,07 milhões) e 10,8% (R\$ 42,93 milhões). Além disso, a subvenção destinada aos estados de RS, SC, PR e SP para a cultura de milho 2ª safra e soja corresponderam à 51,9% (R\$ 38,45 milhões) e 64,6% (R\$ 108,43 milhões), respectivamente. Observa-se que, para estes estados, há um valor expressivo da subvenção, em virtude de um maior número de apólices vendidas, decorrentes da própria importância agrícola nesses estados e decorrente perdas causadas por eventos climáticos.

O seguro agrícola começou a se desenvolver mais rapidamente no Brasil a partir de meados de 2000, com a estabilização da moeda através do Real, aumento das exportações de commodities e a subvenção do governo à taxa do prêmio do seguro, tornando-o mais acessível aos produtores.

A próxima seção apresenta o histórico do seguro agrícola no Brasil.

1.1 Histórico do Seguro Rural no Brasil

A experiência em seguro agrícola no Brasil começou na década de 30 quando o estado de São Paulo estipulou o seguro obrigatório contra granizo para lavoura algodoeira, através do Decreto nº 10.554/39. Em 1948, foi criada a carteira de seguro agrícola contra granizo para viticultura (Lei nº 11/48) e a carteira de seguro agrícola contra geada para horticultura (Lei nº 8.375/64) pela Secretaria de Agricultura do Estado de São Paulo.

Na década de 50 foi instituído o Fundo de Estabilidade de Seguro Agrário (FESA) pela Lei nº 2.168/54 que tinha por objetivo garantir a estabilidade do mercado, permitir o ajustamento gradual das taxas de prêmios, cobrir riscos de catástrofe, bem como quaisquer outras iniciativas relativas ao aperfeiçoamento e à ampliação do seguro rural. A administração dos recursos do fundo ficava a cargo do Instituto de Resseguros do Brasil (IRB)³ por meio do seu conselho técnico (OZAKI, 2005).

¹O seguro agrícola com cobertura multirrisco depende muito de cada seguradora, no entanto, as condições de coberturas são parecidas. A cobertura multirrisco prevê cobertura para causas não evitáveis de perda de produção para os casos de adversidades climáticas (seca, chuvas excessivas, geadas, incêndio, raio, tromba d'água, ventos fortes).

²Riscos nomeados são aqueles cujos riscos cobertos são apresentados em coberturas distintas, possibilitando ao produtor contratar apenas as coberturas de maior interesse como granizo, geada e incêndio.

³O IRB foi criado a partir do Decreto-Lei nº 1.186/39 e tinha por objetivo regular as operações de resseguro e cosseguro no país. Este decreto obrigou as seguradoras a fazerem as operações de resseguro exclusivamente com IRB.

A mesma lei autorizou o governo a criar uma sociedade de economia mista com sede na cidade do Rio de Janeiro, denominada Companhia Nacional de Seguro Agrícola (CNSA), com o objetivo de desenvolver progressivamente operações de seguros rurais.

No entanto, apesar dos esforços, as ações desenvolvidas pela CNSA não obtiveram saldo positivo, ocasionando sua extinção em 1966 (Decreto-Lei nº 73/66). O insucesso da CNSA foi atribuído a alguns fatores, como a excessiva centralização da administração na cidade do Rio de Janeiro, e a adoção de planos que se aplicavam a todo o País de modo indistinto.

Além disso, a não-permissão de operar com outros seguros mais rentáveis, contribuindo assim para o equilíbrio de sua carteira, e a falta de apoio e colaboração das instituições governamentais e financeiras ajudaram no fracasso da CNSA. Ademais, algumas medidas importantes não foram levadas em consideração, como a falta de composição de planos regionais levando em consideração as peculiaridades do meio ambiente, a limitação da área a ser segurada e a seleção dos produtores. Como consequência, houveram dificuldades de se estabelecer prêmios acessíveis aos produtores e manter o contrato sustentável (OZAKI, 2005).

Na metade do anos 60, o Sistema Nacional de Seguros Privados foi estabelecido através do Decreto-Lei nº 73/66, o qual foi regulamentado pelo Decreto nº 60.459/67. Estes dispositivos legais servem até hoje como arcabouço de leis para atividade dos seguros privados no Brasil. Esse Decreto-Lei também constituiu o Fundo de Estabilidade do Seguro Rural (FESR), com o objetivo de garantir a estabilidade das operações e a cobertura suplementar dos riscos de catástrofes.

Além disso, foi instituído o Sistema Nacional de Seguros Privados, constituído pelo Conselho Nacional de Seguros Privados (CNSP), Superintendência de Seguros Privados (SUSEP), Instituto de Resseguros do Brasil (IRB), sociedades autorizadas a operarem em seguros privados e corretores habilitados. Ademais, as seguradoras autorizadas a operarem em seguros privados estavam obrigadas a obedecerem às normas e instruções da SUSEP e do IRB sobre operações de seguro, cosseguro e resseguro, além de fornecer dados e informações de quaisquer aspectos de suas atividades.

No início da década de 70, uma medida importante foi aprovada pelo governo por meio da Lei nº 5969/73 que institui o Programa de Garantia da Atividade Agropecuária (Proagro). O Proagro foi destinado a exonerar o produtor rural do cumprimento de obrigações financeiras em operações de crédito de custeio, no caso de perdas de receitas motivadas pelas adversidades naturais inerentes à exploração agropecuária, ou seja, o governo federal garante ao produtor o pagamento de sua dívida em caso de sinistro.

Historicamente, o Proagro pode ser dividido em três fases distintas: o Proagro “velho”, que cobre o período entre 1975 até 1991; o Proagro “novo”, que se estende de 1991 até 2004 e o Proagro Mais, a partir de 2004. O Proagro “velho” desde seu início até meados de 1993 acumulou déficits de aproximadamente US\$ 1,6 bilhão, em função de alguns problemas, entre eles, atrasos e não pagamento dos débitos pendentes, desistências técnico-operacionais e ocorrência de fraudes.

Por outro lado, o Proagro “novo” trouxe uma inovação ao programa instituído pela Lei nº 8.171/1991 e regulamentada pelo Decreto nº 175/91 e pela Resolução nº 1855/91. O Proagro passou a cobrir os recursos próprios utilizados pelo produtor na atividade segurada. O programa retirou a vinculação necessária ao crédito de custeio, ampliando a cobertura para atividades não financiadas.

Além disso, foi instituído o programa Garantia Safra (GS) que também é direcionado para agricultores familiares que contratam crédito agrícola, mas em regiões majoritariamente semiáridas e que sofrem perdas sistemáticas de safra por motivo de seca ou excesso de chuvas, direcionados para produtores do nordeste e norte do Brasil e nos estados de Minas Gerais e Espírito Santo, que representa a área de atuação da Superintendência do Desenvolvimento do Nordeste (Sudene).

Em 2009, o Proagro “novo” foi alterado pela Lei nº 12.058/09, passando a constituir o marco legal vigente do Proagro Mais.

O Proagro Mais, conhecido também como Seguro da Agricultura Familiar (SEAF), foi instituído pela Resolução nº 3.234/04 com o objetivo de atender os produtores rurais nas operações de custeio agrícola e que trabalham na agricultura familiar vinculados ao Pronaf ⁴.

Tanto o Proagro quanto o Proagro Mais apresentam mecanismos semelhantes de indenização de perdas comparados ao seguro agrícola de custeio, entretanto, o seu principal foco é evitar a inadimplência das operações de crédito agrícola causada pelos riscos vinculados à atividade. Os produtores rurais que contratam tais programas pagam um prêmio, tendo o Governo Federal como uma espécie de seguradora que arcará com as despesas em caso de sinistro.

Os recursos para financiar o Proagro Mais são provenientes dos cofres da União e administrados pelo Banco Central do Brasil (BACEN). Assim, apesar de ser parecido com o seguro de custeio, o Proagro não é considerado seguro, pois as suas operações não são realizadas por seguradoras e sua regulamentação não é realizada pela SUSEP (CARVALHO *et al.*, 2013).

No início dos anos 2000, foi editada a Lei nº 10.823/03 (regulamentada pelo Decreto Federal nº 5.121/04), a qual dispõe sobre a subvenção econômica ao prêmio do seguro rural e institui o Programa de Subvenção ao Prêmio do Seguro Rural (PSR). O PSR tem como principais objetivos: reduzir o custo de aquisição do seguro (prêmio) pelo produtor, massificar a utilização do seguro rural no país, aumentando o número de lavouras e hectares amparados e estabilizar a renda dos produtores rurais, reduzindo a demanda por renegociação e prorrogação de dívidas (MAPA, 2017b).

Foi a partir de 2007, com a Lei Complementar nº 126/07, que o país começou a equiparar sua legislação securitária com a de outros países que já operam com mercado de resseguros competitivos e globalizados. Esta lei extinguiu o monopólio das operações de resseguro exercido pelo IRB. Além disso, foi a partir de 2006 que a operacionalização do PSR se iniciou.

O percentual subvencionado pelo governo é variável e depende do tipo de cultivo e da modalidade de seguro. Estes percentuais têm sido alterados ocasionalmente pelo Ministério da Agricultura, Pecuária e Abastecimento (MAPA), assim como o valor máximo subvencionável por produtor. No ano de 2016, um percentual de 35% à 55% do valor do prêmio foi pago pelo MAPA, enquanto o limite individual subvencionável, por ano civil, passou para R\$ 72 mil para a modalidade agrícola e para R\$ 24 mil para as demais modalidades (MAPA, 2017b). O maior percentual de cobertura é destinado aos cultivos mais afetados e em regiões prioritárias.

O público-alvo do PSR é formado por agricultores não enquadrados na categoria familiares, pois estes são atendidos pelo Proagro ou pelo Garantia-Safra. A adesão é geralmente de pequenos e médios produtores na forma definida em lei.

Para fins de contratação do seguro rural, o agricultor precisa seguir o calendário de plantio do zoneamento agrícola de risco climático (ZARC) disponibilizado pelo MAPA, pois caso o produtor não tenha observado o ZARC, e em caso de sinistro, a indenização pode ser indeferida (BACEN, 2010).

1.2 Seguro Rural nos Estados Unidos

Uma das primeiras experiências mundiais conhecidas de seguro privado em larga escala foi observada na região de Minneapolis, nos Estados Unidos, em 1899. O seguro de riscos múltiplos para a lavoura de trigo foi oferecido por apenas um ano. Posteriormente, em 1917, o seguro foi ofertado por três seguradoras privadas dos estados de Dakota do Norte, Dakota do Sul e Montana, que sofreram prejuízos devido à seca e à limitada abrangência geográfica. Nas décadas seguintes, outras experiências com seguro agrícola de riscos múltiplos não tiveram sucesso, devido em grande parte à quebra de safra por seca, geada e excesso de umidade (OZAKI, 2005).

⁴Pronaf – Programa Nacional de Fortalecimento da Agricultura Familiar

Em 1938, o congresso americano, por meio de lei, criou a *Federal Crop Insurance Corporation* (FCIC) para a formulação das políticas, administração, pesquisa e coordenação com outras agências. A FCIC iniciou como um experimento e as atividades de colheita foram principalmente limitadas às principais culturas (trigo, linho, algodão e milho) nas principais áreas produtoras.

O seguro de colheita manteve-se como um experimento até a aprovação da Lei Federal do Seguro Agrícola de 1980 (*Federal Crop Insurance Act/1980*). De acordo com esta lei, o Seguro Agrícola de Múltiplo Riscos⁵ (MPCI) seria garantido como a principal forma de proteção contra desastres, estabelecendo ainda o subsídio de, no mínimo, 30% do prêmio para os níveis garantidos de produtividade de 50% e 65%.

Além disso, esta lei ampliou o programa de seguro de colheitas para muitas outras culturas e regiões do país, incentivou a expansão para substituir a cobertura gratuita de desastres (compensação aos agricultores por perdas de plantação preventivas e perdas de produtividade) oferecidas em leis agrícolas, criadas nas décadas de 1960 e 1970, porque a cobertura gratuita competia com o programa experimental de seguro de colheita (RMA, 2018a).

Embora mais agricultores tenham participado do programa após a aprovação da Lei de 1980, este não alcançou o nível de participação que o Congresso esperava. Os produtores acumularam dívidas em virtude de desastres e o governo foi obrigado a criar outros programas para auxiliar os produtores. Então, em 1994, foi promulgada a *Federal Crop Insurance Reform Act/1994*, que tornou "obrigatória" a participação de produtores ao programa de seguro de colheita. Essa obrigação estava vinculada a elegibilidade dos produtores para recebimento de benefícios ou participar de qualquer outro programa de incentivos do *United States Department of Agriculture (USDA)* e para a obtenção de certos tipos de créditos.

Como a participação era obrigatória, a *catastrophic coverage* (CAT) foi criada. A cobertura da CAT compensou os agricultores por perdas superiores a 50% de um rendimento médio pago em 60% do preço estabelecido para a safra naquele ano. O prêmio para a cobertura da CAT foi completamente subsidiado. Os participantes pagaram US\$ 50 por cultivo por município, sujeito a montantes máximos para várias culturas e municípios segurados pelo mesmo indivíduo. Os subsídios para maiores níveis de cobertura foram aumentados.

Em 1996, o congresso americano revogou o requisito de participação obrigatória. No entanto, os agricultores que aceitaram outros benefícios foram obrigados a comprar um seguro de colheita ou renunciar à sua elegibilidade para quaisquer outros benefícios disponíveis para o ano agrícola. Essas disposições ainda estão em vigor.

No mesmo ano, a Agência de Gerenciamento de Riscos (RMA)⁶ foi criada para administrar os programas na FCIC e outros programas de gerenciamento de risco e educação não relacionados ao seguro, que ajudam a apoiar a agricultura nos Estados Unidos. A RMA e a FCIC, juntas, regulam e administram as operações de seguro agrícola nos Estados Unidos até hoje.

Existem diferentes tipos de seguros disponibilizados no mercado americano. Os tipos de seguro diferem em função das unidades de seguro, classificação que leva em conta a cultura produzida, a localização dos talhões⁷, o tipo de lavoura e de propriedade (OZAKI, 2005). Cada parcela do terreno para qual o seguro é calculado é denominado unidade de seguro.⁸

Além disso, é possível receber um pagamento de indenização em uma unidade enquanto outras unidades da fazenda têm safras recordes. Por isso, muitos agricultores gostam de dividir suas terras em

⁵Do termo em inglês: *Multiple Peril Crop Insurance*.

⁶Do termo em inglês: *Risk Management Agency (RMA)*.

⁷Talhão é a unidade mínima de cultivo de uma propriedade que é construído com base em relevo e planejamento de mecanização. Em grãos no sul e sudeste do Brasil, os talhões são criados contornando as curvas de nível e com comprimento ideal para preencher a capacidade de um transbordo.

⁸Existem quatro tipos de unidades: unidade básica, opcional, unidade empresarial e de fazenda inteira, mais detalhes dessas unidades estão descritos em ISU (2017).

tantas unidades quanto forem possíveis, o que pode resultar em prêmios mais elevados em cada unidade básica.

Tendo em vista essas observações sobre os diferentes tipos de unidades disponíveis no mercado americano, passa-se a explicitar os principais tipos de seguros oferecidos nos Estados Unidos. Em 2011, a política de seguro nos Estados Unidos sofreu modificações em que alguns programas foram renomeados e fundiram-se, passando por alguns ajustes nos seus termos e condições RMA (2018b).

A Tabela 1.2 apresenta os diferentes tipos de seguros oferecidos nos Estados Unidos, de acordo com RMA (2018b).

Tabela 1.2. Tipos de seguros mais comuns nos Estados Unidos.

Tipos de seguros	Cobertura
<i>Actual Production History (APH)</i>	Perdas de produtividade em função de causas naturais, como a seca, umidade excessiva, granizo, vento, geada, insetos e doenças (níveis de produtividade baseados no histórico próprio do produtor).
<i>Actual Revenue History (ARH)</i>	Diminuição da receita em função de perdas de produtividade, preços baixos, baixa qualidade ou qualquer combinação desses eventos. (histórico próprio)
<i>Adjusted Gross Revenue (AGR)</i>	Proteção contra baixas receitas devido a desastres naturais e flutuações de mercado inevitáveis. A receita agrícola coberta é composta por rendimentos de produtos agrícolas, incluindo renda de animais e produtos de origem animal e aquicultura criados em um ambiente controlado.
<i>Adjusted Gross Revenue-Lite (AGR-L)</i>	AGR para até US\$ 1.000.000,00 de capital segurado
<i>Area Risk Protection Insurance (ARPI)</i>	Seguro de receita com cálculos de prêmio e indenizações baseados nas estatísticas do distrito(município). ⁹
<i>Commodity Exchange Price Provisions (CEPP)</i>	Condições adicionais para apólices das seguintes commodities: cevada, canola (incluindo colza), milho, algodão, sorgo de grãos, arroz, soja, girassol e trigo.
<i>Contract Price Addendum (CPA)</i>	Condições especiais para apólices de produtos orgânicos que tenham preços de venda pré-estipulados por contrato
<i>Rainfall Index (RI)</i>	Cobre efeitos de redução de produção devido a mudança na precipitação. Dados do NOAA's Climate Prediction Center ¹⁰ .
<i>Vegetation Index (VI)</i>	Para culturas as quais a produtividade acompanha o crescimento da vegetação local. Dados do EROS ¹¹ , derivados de satélites, observando mudanças de longo prazo no verde da vegetação da Terra desde 1989.
<i>Whole-Farm Revenue Protection (WFRP)</i>	Gerenciamento de riscos para todas as commodities na fazenda sob uma apólice de seguro. Este plano de seguro é adaptado para qualquer fazenda com até US\$ 8,5 milhões em receita segurada, incluindo fazendas com especialidades ou commodities orgânicas (culturas e pecuária).
<i>Revenue Protection (RP)</i>	Cobre variação no faturamento devido a variações de produtividades ocasionadas por eventos climáticos e variações de preço. O produtor seleciona a quantidade de produtividade média que deseja assegurar (50 – 75%), o preço projetado e o preço de colheita são 100% dos valores determinados de acordo com as provisões de preço cambial de commodities e são baseados em preços de liquidação diária para certos contratos de futuros. O montante da proteção do seguro baseia-se no maior preço projetado ou no preço de colheita.
<i>Revenue Protection With Harvest Price Exclusion</i>	Proteção de receita (RP) a menos que o valor da proteção do seguro seja baseado apenas no preço projetado (o valor da proteção do seguro não é aumentado se o preço de colheita for maior do que o preço projetado).
<i>Yield Protection (YP)</i>	Assegura a proteção de produtividade da mesma forma que APH, a menos que um preço projetado seja usado para determinar a cobertura do seguro. O preço projetado é determinado de acordo com as provisões de preço de commodities e é baseado em preços de liquidação diária para certos contratos de futuros. O YP foi instituído para culturas que disponham tanto da proteção de produtividade quanto de receita e utilizem o mercado de bolsa de valores para a descoberta de preços.
<i>Catastrophic Risk Protection (CAT)</i>	Catástrofes (50% de garantia da produção) paga 55% do preço da mercadoria estabelecida pela RMA sobre perdas de colheita superiores a 50%. O prêmio sobre a cobertura CAT é pago pelo governo federal; no entanto, os produtores devem pagar uma taxa administrativa de US\$ 300 (a partir da Farm Bill de 2008) para cada cultura segurada em cada município.

Para cada produto apresentado, há uma lista de culturas e regiões elegíveis, sendo que cada cultura possui um documento próprio, em forma de lei, onde são apresentadas as regras para o contrato de apólice, indenização e coberturas.

O mercado de seguro americano é um dos mais antigos no mundo, o que o tornou base para o desenvolvimento de mercados em outros países, como o Brasil. Os produtos de seguros oferecidos no Brasil são em grande parte baseados nos produtos americanos. Na próxima seção apresenta-se os tipos de seguros oferecidos no mercado brasileiro.

1.3 Tipos de Seguros Agrícolas Comercializados no Brasil

Existem vários tipos de seguros privados agrícolas oferecidos no Brasil: seguro de custeio, de produtividade, receita ou faturamento e de índice agrícola. Para cada tipo de seguro, a abrangência de cobertura de sinistro dos produtos é variável entre as seguradoras ofertantes. No entanto, a cobertura básica é semelhante. A Tabela 1.3 apresenta os tipos de seguros e a cobertura básica.

Tabela 1.3. Tipos de seguros agrícolas ofertados no Brasil.

Modalidade de Seguro	Cobertura
Custeio	cobre a despesa de custeio da safra, desde o plantio até a colheita, em casos de perda de produção ocasionados por granizo, geadas, chuva excessiva, ventos fortes e incêndios.
Produtividade	cobre a perda de produção do agricultor em função de eventos climáticos, caso a produtividade esperada não seja alcançada.
Receita ou Faturamento	cobre a perda de receita ou faturamento do agricultor por hectare cultivado em função dos riscos climáticos e de mercado.
Índice	cobre a perda de produtividade, associada a um indicador regional.

O seguro de custeio cobre os custos desde o plantio até a colheita em casos de perda de produção ocasionados por granizo, geada, chuva excessiva, incêndios e ventos fortes. Em caso de sinistro a indenização é calculada pelo custeio da safra por hectare(ha)¹².

No seguro de produtividade a indenização é a diferença entre a produção em quantidade (sacas ou ton./ha) estimada na contratação da apólice e a produção efetiva na colheita¹³.

No caso do seguro de receita ou faturamento, a perda de faturamento é a diferença entre o faturamento esperado e o faturamento realizado com a venda da produção após a colheita. O faturamento esperado depende da produtividade da lavoura (sacas ou toneladas por ha) e também do preço do produto¹⁴.

Por fim, os seguros de índices paramétricos baseiam-se na definição de índices para a ocorrência de eventos naturais. Por exemplo, no caso de chuva, um dos indicadores é a precipitação pluviométrica inferior a um determinado índice acordado entre o segurado e a seguradora. Esse índice pode ser precipitação, vazão de rio, temperaturas extremas, ventos, radiação solar e índices de *El Niño*. O seguro de índice começou a ser vendido no Brasil em 2016 e é customizado para cada cliente, por isso, ainda não existem dados públicos sobre este tipo de seguro.

Segundo MAPA (2017a), em 2017, foram vendidas 22.465 (33%), 39.327 (58%) e 1.965 (2,92%) apólices do seguro de custeio, produtividade e faturamento, respectivamente. Das apólices de custeio, produtividade e faturamento, 88,32%, 65,19% e 97,91% destas foram vendidas para grãos, respectivamente (Figura 1.1).

⁹ARPI substitui o Plano de Risco do Grupo (GRP) e o Plano de Proteção de Renda de Risco do Grupo (GRIP).

¹⁰Do termo em inglês: *National Oceanic and Atmospheric Administration's Climate Prediction Center*.

Figura 1.1. Quantidade de apólices vendidas em 2017 (MAPA, 2017).

Mesmo com um crescimento expressivo do seguro, ainda existem problemas que afetam o mercado de seguro agrícola brasileiro, tais como, problemas relacionados à assimetria de informação, como risco moral¹⁵ e a seleção adversa¹⁶ e o risco sistêmico¹⁷.

Além disso, a falta de dados de produtividade agrícola relativamente longos e por fazendas dificultam o cálculo de maneira precisa da estrutura de riscos dos produtores. Ademais, a escassa opção de metodologias atuariais para precificação dos riscos do seguro agrícola podem resultar em taxas imprecisas e mal calculadas ocasionando graves implicações no problema da seleção adversa. Consequentemente, apenas aqueles que possuem maior risco demandam contratos, tornando a carteira de seguros altamente deficitária. Além disso, o escasso número de metodologia é apontado como um dos principais entraves para o desenvolvimento de um mercado de seguro agrícola no Brasil (OZAKI, 2005).

Portanto, este trabalho tem por objetivo apresentar métodos alternativos para precificação dos riscos do seguro de produtividade e faturamento agrícola, trazendo abordagens alternativas àquelas apresentadas em estudos anteriores, (RAMIREZ *et al.*, 2003), (SHERRICK *et al.*, 2004), (OZAKI, 2005), (GOODWIN and TEJEDA, 2008), (BRISOLARA, 2013).

Para isso, três artigos são apresentados nos capítulos a seguir. No primeiro artigo, a precificação é feita utilizando-se distribuições paramétricas que capturam a simetria, a assimetria e a bimodalidade dos dados de produtividade de soja para municípios selecionados do estado do Paraná, características

¹¹Do termo em inglês: *Geological Survey's Earth Resources Observation and Science*.

¹²Para mais detalhes, veja OZAKI (2005) e CARVALHO *et al.* (2013).

¹³Para mais detalhes, veja capítulo 2, seção 3 e CARVALHO *et al.* (2013).

¹⁴Para mais detalhes, veja capítulo 3, seção 3 e CARVALHO *et al.* (2013).

¹⁵Segundo OZAKI (2005), o problema do risco moral ocorre quando o produtor segurado opta por decisões que afetam a probabilidade de perdas e não podem ser observados pela seguradora acarretando no aumento do volume das indenizações pagas pela seguradora. Por exemplo, quando produtores adquirem seguro agrícola e utilizam menor quantidade de fertilizantes, comparando com os não segurados.

¹⁶A seleção adversa ocorre quando a seguradora não consegue distinguir os diferentes grupos de riscos. Neste caso, as pessoas mais sujeitas a um determinado dano coberto pelo seguro são as mais propensas a contratar o seguro.

¹⁷As eventualidades climáticas podem afetar vários produtores de uma mesma região, extrapolando as fronteiras da fazenda, municípios ou região. Esse efeito contagioso é chamado de risco sistêmico. O risco sistêmico da produtividade agrícola viola uma das condições de segurabilidade da teoria atuarial, que pressupõe a existência de elevado número de unidades expostas, homogêneas e independentes, o que implica na necessidade de uma dispersão geográfica das unidades seguradas.

estas geralmente encontradas em produtividades brasileiras e que não foram levadas em consideração em trabalhos anteriores. As distribuições utilizadas foram Normal, Beta, Skew-Normal, Skew-t e Odd-Log-Logística-Normal, respectivamente. Além disso, calculou-se as taxas de prêmio de seguro de produtividade e comparou-se com a distribuição Normal (geralmente utilizada pelo mercado segurador) e discutiu-se suas implicações para o mercado segurador.

No segundo artigo, utiliza-se a metodologia de cópulas para análise multivariada entre produtividade e preço. Esta metodologia apresenta um ganho significativo para estudos de problemas multivariados comparados à distribuição Normal multivariada, pois pode-se levar em consideração estruturas de dependências linear, não linear e dependência somente nas caudas da distribuição multivariada, além de especificar qualquer tipo de distribuição para as distribuições marginais. Neste artigo, comparou-se e discutiu-se as implicações das taxas calculadas por esta metodologia e as aplicadas pelo mercado segurador.

No terceiro artigo, aplicou-se a metodologia de cópulas considerando o problema tridimensional, no qual supõem-se a produtividade simulada da distribuição paramétrica OLLN (apresentada no capítulo 1), preços futuros de contrato de soja da CME e cotação do dólar. Esta análise é inédita e foi utilizada pelo fato da soja produzida no país ser exportada, e produtores utilizarem como referência internacional de preços a bolsa americana, e portanto esses preços têm grande influência nos contratos de seguros negociados no Brasil.

Referências

- BACEN, 2010 Manual de crédito rural. Acessado em 14 de março de 2018. Disponível em <https://www3.bcb.gov.br/mcr/completo>.
- BRISOLARA, C. S., 2013 *Proposições para o desenvolvimento do seguro de receita agrícola no Brasil: do modelo teórico ao cálculo das taxas de prêmio*. Ph.D. thesis, Escola Superior de Agricultura “Luiz de Queiroz”.
- BUANAIN, A. M. and R. L. F. SILVEIRA, 2017 *Manual de Avaliação de Riscos na Agropecuária - Um guia metodológico*, volume único. Fundação Escola Nacional de Seguros, first edition.
- CARVALHO, A. L., C. A. PAREDES, D. MIQUELLUTI, D. RUIS, E. PASSARELLI, and G. V. DUARTE, 2013 Boletim do seguro rural, edições 1, 2, 3 e 4. Acessado em 23 de junho de 2015. Disponível em <http://www.esalq.usp.br/geser/docs/boletim>.
- GOODWIN, B. and H. TEJEDA, 2008 Modeling crop prices through a burr distribution and analysis of correlation between crop prices and yields using a copula method. In *American Agricultural Economics Association Annual Meeting, Orland-Flórida..*
- ISU, 2017 Proven yields and insurance units for crop insurance. Acessado em 20 de janeiro de 2018. Disponível em <https://www.extension.iastate.edu/agdm/crops/html/a1-55.html>.
- MAPA, 2017a Atlas do seguro rural. indicador das taxas. Acessado em 21 de fevereiro de 2017. Disponibilizado em <http://indicadores.agricultura.gov.br/atlasdoseguro/index.htm>.
- MAPA, 2017b Relatório estatístico geral 2016- seguro rural- ministério da agricultura, pecuária e abastecimento. Acessado em 10 de janeiro de 2018. Disponibilizado em <http://www.agricultura.gov.br/assuntos/riscos-seguro/seguro-rural/documentos-seguro-rural/ResultadoGeral2016.pdf>.
- OZAKI, V. A., 2005 *Métodos atuariais aplicados à determinação da taxa de prêmio de contratos de seguro agrícola: um estudo de caso..* Ph.D. thesis, Universidade de São Paulo.

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

RMA, 2018a History of the crop insurance program. Acessado em 01 de fevereiro de 2018. Disponível em <https://www.rma.usda.gov/aboutrma/what/history.html>.

RMA, 2018b Policies. Acessado em 30 de janeiro de 2018. Disponível em <https://www.rma.usda.gov/policies/>.

SHERICK, B. J., F. C. ZANINI, G. D. SCHNITKEY, and S. H. IRWIN, 2004 Crop insurance valuation under alternative yield distributions. *American Journal of Agricultural Economics* **86**: 406–419.

2 MODELAGEM DA PRODUTIVIDADE DE SOJA UTILIZANDO DISTRIBUIÇÕES SIMÉTRICAS, ASSIMÉTRICAS E BIMODAIS: IMPLICAÇÕES PARA O SEGURO DE PRODUTIVIDADE AGRÍCOLA

Resumo

A maior parte dos trabalhos utilizam distribuições paramétricas para modelar a produtividade agrícola, tais como: Normal, Beta, Log-normal e Skew-normal (SN). Estes modelos são bem definidos matematicamente e computacionalmente, mas não incorporam a bimodalidade. Portanto, na modelagem é necessário estudar distribuições mais flexíveis, uma vez que a maioria dos dados de produtividade no Brasil apresentam evidências de assimetria ou bimodalidade. O objetivo deste estudo foi modelar e prever a produtividade de soja para municípios selecionados do Estado do Paraná, no período de 1980 a 2014, utilizando as distribuições Odd-Log-Logística Normal (OLLN), Beta, SN e Skew-t-Student (ST) para as séries simétricas, assimétricas e bimodais. O modelo OLLN foi o que melhor se ajustou aos dados. Os resultados sugerem que, para o município de Cascavel a taxa do seguro de produtividade estimada pelo modelo Normal está subestimada quando comparadas às do modelo OLLN, com nível de cobertura (NC) entre 55% a 80%. Entretanto, para NC de 83% para 90% o modelo Normal superestima a taxa quando comparada ao modelo OLLN. Para o município de Guarapuava, sugere-se uma superestimação das taxas calculadas pelo modelo Normal quando comparadas às do modelo OLLN. Além disso, para o município de Palmeira a taxa calculada pelo modelo Normal está superestimada para o NC inferior a 80%. Finalmente, para o município de Castro sugere-se uma subestimação da taxa calculada pelo modelo Normal quando comparado ao modelo OLLN com NC inferior a 87%.

Palavras Chaves: Assimetria; Distribuição bimodal; Gerenciamento de riscos agrícolas; Seguro de produtividade.

Abstract

Most studies papers used parametric distributions to model crop yields, such as: normal, 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 shows evidence of asymmetry or bimodality. The aim of this study was to model and forecast soybean yields for municipalities in the State of Paraná, 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 (ST) 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.

Keywords: Asymmetry; Bimodality distribution; Crop insurance; Risk management.

2.1 Introdução

A agricultura é uma das atividades mais antigas e importantes para o ser humano, sendo considerada fonte de matérias-primas, alimentos e energia. Porém, há riscos relacionados à agricultura em termos de produção devido a eventos climáticos extremos. Ao longo dos anos, várias ferramentas de gerenciamento de riscos foram desenvolvidas para minimizar o impacto do clima na produção agrícola, entre elas, o seguro de produtividade (HARWOOD *et al.*, 1999). Além de minimizar os riscos de perda na agricultura, o seguro de produtividade fornece a recuperação da capacidade financeira do produtor em caso de danos causados por eventos naturais extremos.

No entanto, o seguro de produtividade apresenta alguns problemas, tais como, a assimetria

da informação, o que causa a seleção adversa¹, risco moral², e a ausência de séries históricas longas desencorajam a oferta de seguros pelas companhias seguradoras (QUIGGIN *et al.*, 1993), (SMITH and GOODWIN, 1996), (GOODWIN, 2001), (COBLE and KNIGHT, 2002).

Desta forma, o Estado comumente desenvolve ações para compensar essas deficiências, como os subsídios nos prêmios de seguros. No entanto, a ausência de uma metodologia atuarial que calcula precisamente a taxa do prêmio e a falta de continuidade no volume de subsídios do governo são apontados como alguns dos principais problemas para o desenvolvimento do mercado de seguros agrícolas no Brasil.

Portanto, este trabalho tem por objetivo propor métodos alternativos para o cálculo do prêmio do seguro de produtividade com base em distribuições paramétricas que capturam a assimetria ou a bimodalidade dos dados. Considerando os dados de produtividade brasileiros, essas características devem ser levadas em consideração para obter o prêmio atuarialmente justo e mais preciso.

Além disso, comparou-se as taxas calculadas pela metodologia proposta com as taxas comerciais cobradas pelas companhias de seguros brasileiras sendo suas implicações discutidas para o mercado segurador.

O restante do artigo está organizado da seguinte forma: a seção 2.2 apresenta a revisão de literatura. A seção 2.3 apresenta uma introdução ao seguro agrícola e aos formas de cálculo do seguro. As descrições dos dados utilizados na pesquisa estão descritos na seção 2.4. A metodologia utilizada para modelar as séries de produção agrícola é apresentada na seção 2.5. A Seção 2.6 apresenta os resultados. A seção 2.7 apresenta as discussões. Finalmente, a seção 2.8 aborda as observações finais.

2.2 Revisão de Literatura

Para obter o prêmio atuarialmente justo para o seguro de produtividade é necessário a correta modelagem dos dados de produtividade. A literatura apresenta três abordagens para esta questão: a primeira envolve a estimativa dos parâmetros de uma distribuição paramétrica; a segunda está relacionada a uma grande variedade de métodos não paramétricos; e a terceira é composta por métodos semi-paramétricos, que combinam elementos da primeira com a segunda alternativas (KER and COBLE, 2003).

De acordo com GOODWIN and MAHUL (2004), os métodos não paramétricos, apesar de serem mais flexíveis, requerem grandes amostras e também apresentam dificuldades na medição de eventos raros ou catastróficos. Mesmo assim, eles são os mais utilizados em estudos com pequenas amostras. Estudos que fizeram uso de métodos não paramétricos, tais como, GOODWIN and KER (1998), TURVEY and ZHAO (1999), KER and GOODWIN (2000), OZAKI *et al.* (2008b) utilizaram o estimador kernel para a densidade da produtividade e calcularam as taxas do prêmio de seguro.

Considerando os métodos paramétricos, JUST and WENINGER (1999) e BOTTS and BOLES (1958) utilizaram a distribuição Normal (N) para estimar a produtividade.

No trabalho de JUST and WENINGER (1999), os autores identificaram evidências de não-normalidade em dados de produtividade agrícola em Kansas, nos Estados Unidos. Os autores identificaram três problemas metodológicos comuns na análise da distribuição da produtividade: erros de especificação de componentes não aleatórios na distribuição de produtividade, má apresentação e análise da significância estatística e uso de dados agregados para representar distribuições de produtividades a nível de fazenda.

¹ De acordo com QUIGGIN *et al.* (1993), o problema da seleção adversa ocorre quando os produtores que estão mais dispostos a contratar o seguro são os mais propensos a apresentar sinistralidade.

² O risco moral refere-se ao fato de que o segurado pode tomar certas ações que a seguradora é incapaz de monitorar, levando a um aumento no risco de produção. Por exemplo, depois de contratar o seguro o produtor pode usar menos fertilizantes ou pesticidas, fazendo com que a produtividade diminua (QUIGGIN *et al.*, 1993).

Segundo JUST and WENINGER (1999), a presença de um ou mais desses problemas afeta qualquer evidência contra a normalidade. Além disso, o uso da produtividade dos municípios (dados agregados), remove a informação específica das fazendas e enfatiza a variabilidade temporal dos produtores da região estudada, suprimindo a variabilidade da fazenda.

Outros estudos, como DAY (1965), TAYLOR (1990), RAMÍREZ (1997), RAMÍREZ *et al.* (2003) e SHERRICK *et al.* (2004) encontraram evidências contra a normalidade, sugerindo que os dados não são simétricos quando comparados com a média.

Além da distribuição normal, outras distribuições paramétricas foram utilizadas para modelar a produtividade, tais como: a distribuição gama em GALLAGHER (1987), a distribuição hiperbólica inversa em MOSS and SHONKWILER (1993), as distribuições Beta e Weibull em SHERRICK *et al.* (2004) e a distribuição Skew-normal (SN) em OZAKI and SILVA (2009). Entre estes modelos, o mais utilizado é a distribuição Beta. Os trabalhos de ?; BABCOCK and HENNESSY (1996); COBLE *et al.* (1996); HENNESSY *et al.* (1997) encontraram fortes evidências de assimetria ou curtose.

O trabalho de RAMÍREZ *et al.* (2003) confirmou as evidências de não-normalidade e assimetria da produtividade agrícola encontradas em RAMÍREZ (1997). As produtividades do milho e da soja foram modeladas na região do *Corn Belt* e algodão no Texas, Estados Unidos. Os autores também testaram a tendência, a heteroscedasticidade, a curtose e assimetria não linear. Uma expansão da distribuição Johnson SU é usada na modelagem por causa de sua flexibilidade. Os testes de tendência não linear e heteroscedasticidade são conduzidos ao mesmo tempo permitindo a possibilidade de distribuições não normais usando a informação adicional transmitida através de uma matriz de correlação. Conclui-se que diferentes padrões de distribuições não normais podem resultar em diferentes fatores que afetam a produtividade no nível da fazenda e dos municípios, dependendo da cultura, do sistema de cultivo e da região geográfica. A principal recomendação do estudo é que os pesquisadores devem reconhecer e considerar a aplicação de distribuições não normais na modelagem da produtividade em qualquer análise de risco, seja a política, o mercado, a indústria ou fazenda.

No artigo de SHERRICK *et al.* (2004), a produtividade de milho e soja foi investigada em 26 fazendas nos Estados Unidos, usando as distribuições Normal, Logística, Weibull, Beta e Log-normal. Neste estudo, os autores calcularam e compararam os pagamentos aos produtos de seguros de produtividade agrícola para cada uma das distribuições paramétricas dentro de cada fazenda e entre eles. Os autores concluíram que a escolha da distribuição tem um impacto significativo na avaliação de risco da produtividade e na taxa do prêmio do seguro.

Os modelos hierárquicos bayesianos também foram aplicados à modelagem da produtividade agrícola, como mostrado por OZAKI and SILVA (2009). Este estudo propôs uma fórmula alternativa para calcular a taxa do prêmio do seguro, em culturas de milho para os municípios do Estado do Paraná, através de modelos espaciais-temporais hierárquicos bayesianos usando a distribuição SN. Neste modelo, a tendência do tempo foi simultaneamente modelada com a correlação espacial entre os municípios. As taxas de prêmio foram calculadas diretamente da distribuição prevista por simulação de Monte Carlo através de Cadeias de Markov. Os autores concluíram que a metodologia proposta melhora os procedimentos atuariais e estatísticos frequentemente utilizados no cálculo das taxas do prêmio, especialmente quando se trabalha com um número reduzido de observações. Os autores concluíram que as taxas empíricas³, comumente utilizadas pelas seguradoras brasileiras superestimam o risco dos produtores em áreas de baixo risco e subestimam o risco de produtores com alto risco.

Portanto, com base nos estudos citados, séries curtas de produtividade limitam o uso de técnicas não paramétricas, no entanto, elas reduzem os erros associados a premissas incorretas relacionadas às

³De acordo com OZAKI and SILVA (2009), essas taxas são baseadas na relação entre a perda média do segurado e não leva em consideração uma análise estatística robusta.

distribuições. A limitação das séries de produtividade e os resultados de vários estudos sugerem o uso de distribuições paramétricas na modelagem de produtividade.

No Brasil, não existe uma base de dados pública com informações sobre a produtividade agrícola em nível de fazenda, apenas em nível municipal. Além disso, a produtividade agrícola no Brasil tem características específicas, como a assimetria e a bimodalidade. Portanto, para um cálculo mais preciso da taxa do seguro, é importante considerar as distribuições que capturam essas características por meio de uma abordagem paramétrica que ainda não foram consideradas na literatura existente.

No trabalho de BRAGA *et al.* (2016), foi proposto uma nova extensão da distribuição Normal com caudas mais pesadas, chamada modelo OLLN. Esta distribuição pode ser simétrica, platicúrtica ou leptocúrtica, unimodal ou bimodal dependendo dos parâmetros. Os autores apresentam três aplicações para dados reais para comprovar empiricamente a flexibilidade da distribuição OLLN. A primeira aplicação envolve as variáveis temperatura ($^{\circ}\text{C}$) e a radiação diária total ($\text{cm}^{-2}\text{d}^{-1}$) que correspondem a dados diários. O segundo conjunto de dados refere-se ao experimento realizado para avaliar os efeitos das doses de um composto anti-helmíntico (ml) para controlar um parasita (efeitos fixos) usando um design inteiramente casualizado (CRD) com cinco tratamentos e a terceira aplicação refere-se ao ganho de peso de gado. O objetivo desta última aplicação é ajustar um modelo com delineamento em blocos casualizados com cinco tratamentos (RCBD) utilizando a distribuição OLLN. Em estudos recentes, CRUZ *et al.* (2017) obteve o modelo bivarido de regressão logístico-Weibull para qualidade de vida relacionada à saúde bucal. BRAGA *et al.* (2017) obteve a nova distribuição skew-bimodal com aplicações e ALIZADEH *et al.* (2017) obteve a família de distribuição odd log-logística logarítmica com aplicações em diferentes áreas.

2.3 Seguro de Produtividade

O seguro agrícola é um importante mecanismo de transferência de risco, no entanto, também possui outros benefícios, entre eles, a estabilidade econômica dos produtores e o desenvolvimento e adoção de novas tecnologias. Todos esses fatores contribuem para a redução do êxodo rural e da dívida de empréstimo do produtor em caso de perdas (BRISOLARA, 2013).

Neste tipo de contrato, o produtor deve ser indenizado se a produtividade observada Y , após a colheita, ficar abaixo da produtividade garantida Y_g no contrato. A produtividade garantida é definida como $Y_g = \lambda Y_e$, em que λ é o nível de cobertura (NC) escolhido pelo produtor, $0 < \lambda < 1$, Y_e é a produtividade média esperada. Na prática, a produtividade esperada Y_e é a média aritmética dos últimos cinco anos.

Os conceitos teóricos relacionados à determinação das taxas do prêmio do seguro de produtividade foram descritos nos trabalhos de LAWAS (2005) e MIQUELETO (2011). De acordo com OZAKI (2005), o prêmio é considerado atuarialmente justo quando a probabilidade do evento de um acidente é igual ao prêmio por unidade de remuneração ou quando o prêmio é igual ao pagamento esperado de indenização.

Assim, a taxa justa do prêmio é dada pela expectativa de perda ou perda esperada:

$$E(\text{perda}) = E(\max\{(y_g - Y); 0\}) = P(Y < y_g)(\lambda y_e - E(Y|Y < \lambda y_e)). \quad (2.1)$$

De acordo com GOODWIN and KER (1998), GOODWIN and MAHUL (2004) e KER and GOODWIN (2000), a taxa do prêmio justo de um contrato de seguro de produtividade é dada por:

$$\text{Taxa Prêmio} = \frac{F_Y(\lambda y_e)[\lambda y_e - E(Y|Y < \lambda y_e)]}{\lambda y_e}, \quad (2.2)$$

em que $F_Y(\lambda y_e)$ é a distribuição acumulada da variável Y , denominado "probabilidade de perda do produtor". Neste contexto, denotamos a importância de estimar corretamente a distribuição da produtividade, uma vez que inferências imprecisas podem alterar os resultados da taxa de prêmio.

2.4 Descrição dos dados

Para a modelagem da produtividade agrícola, os dados anuais da produtividade de soja (kg/ha) de alguns municípios importantes no Paraná-Brasil foram obtidos do Instituto de Desenvolvimento Econômico e Social do Paraná (IPARDES, 2015). As observações disponíveis estão relacionadas aos anos agrícolas de 1979/1980 a 2013/2014, totalizando 35 observações. A Figura 2.1 mostra a produtividade agrícola da soja em diferentes municípios do Paraná para 2014, de acordo com IPARDES (2015).

Figura 2.1. Produção de soja no Paraná em 2014.

Os municípios selecionados são Cascavel, Guarapuava, Castro e Palmeira. Esses municípios, além de serem grandes produtores de soja, também foram selecionados porque são os líderes estaduais de área plantada de soja, de acordo com o Boletim de Monitoramento Agrícola da Companhia Nacional de Abastecimento (Conab) da safra de temporada de verão 2015/2016 (CONAB, 2016). As taxas comerciais aplicadas pelas seguradoras A, B, C e D aos municípios selecionados foram fornecidas pelo Ministério da Agricultura, Pecuária e Abastecimento (MAPA, 2016).

2.5 Metodologia

Esta seção apresenta os modelos paramétricos utilizados na modelagem de produtividade e as medidas de ajuste para escolha do melhor modelo. O Apêndice A apresenta o método de estimação por máxima verossimilhança aplicado aos modelos utilizados neste estudo. Todas as análises foram realizadas utilizando a função *Optim* e o pacote *AdequacyModel* do software R (R CORE TEAM, 2016). Além disso, o apêndice I apresenta o *script* do software R, no qual se calcula a função densidade (*pdf_oln*) e a distribuição acumulada (*cdf_oln*) para o modelo OLLN.

Os conhecidos modelos Normal e Beta foram utilizados como uma primeira abordagem para modelar a produtividade. Como a produtividade pode apresentar um comportamento assimétrico, os modelos SN (AZZALINI, 1985) e ST também foram considerados.

2.5.1 Modelo Normal

Uma variável aleatória contínua X é normalmente distribuída, (para $x \in \mathbb{R}$), quando sua função densidade de probabilidade (fdp) e distribuição acumulada (fda) são dados por:

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\} = \frac{1}{\sigma} \phi \left(\frac{x - \mu}{\sigma} \right),$$

e

$$F(x; \mu, \sigma) = \Phi\left(\frac{x - \mu}{\sigma}\right),$$

em que $\mu \in \mathbb{R}$ é o parâmetro de locação, $\sigma > 0$ é o parâmetro de escala, $\phi(\cdot)$ e $\Phi(\cdot)$ são fdp e fda da distribuição normal padrão, respectivamente.

Uma vez que a integração explícita não é possível, os métodos de integração numérica são utilizados para o cálculo do fda. O parâmetro de locação μ está relacionado ao deslocamento da função no eixo das abscissas (eixo x). Enquanto o parâmetro de escala σ está relacionado a um aumento ou diminuição da variabilidade da série.

2.5.2 Modelo Beta

A distribuição beta padrão é uma função de distribuição de probabilidade contínua definida no intervalo $[0, 1]$, com dois parâmetros positivos a e b , que definem a forma da distribuição. A função de distribuição de densidade beta padrão é dada por:

$$f(x; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1}, \quad a, b > 0, \quad 0 \leq x \leq 1$$

em que $\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt$ é a função gamma.

No entanto, é possível generalizar a distribuição beta para qualquer variável aleatória Y , restrita a um intervalo finito (c, d) . Para isso, temos que usar a relação $Y = (d-c)X + c$. Então, se $X \sim \text{Beta}(a, b)$, então Y segue a distribuição beta estendida com 4 parâmetros (a, b, c, d) , com a fdp dada por:

$$f(y; a, b, c, d) = \frac{(y-c)^{a-1} (d-y)^{b-1}}{(d-c)^{a+b-1} B(a, b)}, \quad d > c \quad c \leq y \leq d,$$

em que $B(a, b)$ é a função beta (JOHNSON and BALAKRISHNAN, 1995).

2.5.3 Modelo Skew-normal

A função de distribuição de probabilidade da skew-normal (SN) apresenta uma transição contínua da normal para não normal (AZZALINI, 1985). Considerando uma variável aleatória X , então a fdp da SN depende dos parâmetros, $\mu \in \mathbb{R}$ (parâmetro de localização), $\sigma > 0$ (parâmetro de escala) e $\lambda \in \mathbb{R}$ (parâmetro de assimetria) e é dado por:

$$f(x; \mu, \sigma, \lambda) = \frac{2}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right) \Phi\left[\lambda \left(\frac{x - \mu}{\sigma}\right)\right]. \quad (2.3)$$

A fdp (2.3) é válida para $x \in \mathbb{R}$ e é simétrica se $\lambda = 0$.

A forma da distribuição possui características interessantes dependendo dos parâmetros, como pode ser visto na Figura 2.2. Observa-se nas Figuras 2.2(a) e 2.2(b) que, quando os parâmetros μ e σ são fixados, o parâmetro λ caracteriza a assimetria dos dados. Se $\lambda < 0$, as curvas assimétricas são deslocadas para a esquerda e se $\lambda > 0$, as curvas assimétricas são deslocadas para a direita. Observa-se na Figura 2.2(c) que quando σ é fixado, o gráfico é transladado no eixo x (variação de μ) e apresenta a assimetria (variação de λ).

2.5.4 Modelo Skew-T Student

Neste estudo, a produtividade pode conter valores discrepantes ou *outliers* devido a falhas nas colheita (quebra de safra) causadas por eventos climáticos adversos. Quando os dados apresentam comportamento assimétrico e na presença de valores extremos, o uso da distribuição normal pode ser

Figura 2.2. Gráfico da fdp da SN para diferentes parâmetros.

ineficaz. Em tais situações, propõe-se, uma alternativa mais robusta ao modelo padronizado, que é a versão assimétrica da distribuição t-Student (ST), proposta por AZZALINI and CAPITANIO (2003).

Uma variável X é chamada Skew-t Student, com parâmetros de locação ($\mu \in \mathbb{R}$), escala ($\sigma \in \mathbb{R}_+^*$), assimetria ($\nu \in \mathbb{R}$) e curtose ($\tau \in \mathbb{R}_+^*$), se sua fdp é dada por:

$$f(x; \mu, \sigma, \tau, \nu) = \frac{2}{\sigma} t(z) T[\omega],$$

em que $z = \frac{x-\mu}{\sigma}$, $\omega = \nu\lambda^{1/2}z$, $\lambda = \frac{\tau+1}{\tau+z^2}$, t_{Z_1} e T_{Z_1} são respectivamente, a fdp da t-Student padrão com τ graus de liberdade e a fda da t-Student padrão com $\tau + 1$ graus de liberdade. Como notação, adotamos $Z \sim ST(\mu, \sigma, \nu, \tau)$. Note que quando $\nu = 0$, a distribuição é equivalente a simétrica t-Student.

Figura 2.3. Gráfico da fdp da Skew-t para diferentes valores dos parâmetros.

A Figura 2.3 ilustra o comportamento da fdp da ST. Nota-se nos gráficos 2.3(a) e 2.3(b) a contribuição do parâmetro ν na forma da distribuição, com μ , σ e τ fixados. Os valores positivos de ν indicam assimetria positiva e os valores negativos de ν indicam a assimetria negativa. Na Figura 2.3(c), observa-se o efeito do parâmetro de curtose na distribuição quando os outros parâmetros são fixados.

2.5.5 Modelo Odd log-logística - F

Uma família de distribuição contínua recentemente introduzida por CRUZ (2016) é a *odd log-logística-F* (OLLF). Esta classe está bem definida e algumas de suas propriedades matemáticas foram demonstradas, como momentos comuns e incompletos e função quantílica. A inferência e a estimativa pelo método de máxima verossimilhança também foram apresentadas para dados censurados de sobrevivência.

Seja uma fda de qualquer função de linha de base $F(x, \xi)$ com um vetor de parâmetro ξ , a fda da OLLF com um parâmetro de forma $\alpha > 0$, é definido por:

$$G(x, \xi) = \int_0^{\frac{F(x, \xi)}{\bar{F}(x, \xi)}} \frac{\alpha t^{\alpha-1}}{(1+t^\alpha)^2} dt = \frac{F(x, \xi)^\alpha}{F(x, \xi)^\alpha + \bar{F}(x, \xi)^\alpha}, \quad (2.4)$$

em que $\bar{F}(x, \xi) = 1 - F(x, \xi)$ e $\alpha = \log \frac{G(x, \xi)}{G(x, \xi)} / \log \frac{F(x, \xi)}{F(x, \xi)}$.

O parâmetro α representa o quociente do log da razão da função geradora e da função base de distribuições.

Podem ser geradas múltiplas distribuições a partir da equação descrita, como as apresentadas nos trabalhos de ALIZADEH *et al.* (2015), CRUZ *et al.* (2016) e BRAGA *et al.* (2016).

A função densidade de probabilidade (fdp) da nova família é definida como:

$$g(x, \xi) = \frac{\alpha f(x; \xi) \{F(x; \xi)[1 - F(x; \xi)]\}^{\alpha-1}}{\{F(x; \xi)^\alpha + [1 - F(x; \xi)]^\alpha\}^2}. \quad (2.5)$$

A nova família de distribuição OLLF permite uma maior flexibilidade nas caudas da distribuição. Seja $F(x; \xi)/\bar{F}(x; \xi) = \Phi(x; \mu, \sigma)/\bar{\Phi}(x; \mu, \sigma)$, na equação (2.4), a fda do modelo OLLN com um parâmetro de forma adicional $\alpha > 0$ é definido por:

$$F(x; \mu, \sigma, \alpha) = \int_0^{\frac{\Phi(x; \mu, \sigma)}{\bar{\Phi}(x; \mu, \sigma)}} \frac{\alpha t^{\alpha-1}}{(1+t^\alpha)^2} dt = \frac{\Phi^\alpha\left(\frac{x-\mu}{\sigma}\right)}{\Phi^\alpha\left(\frac{x-\mu}{\sigma}\right) + [1 - \Phi\left(\frac{x-\mu}{\sigma}\right)]^\alpha}, \quad (2.6)$$

em que $\bar{\Phi}(x; \mu, \sigma) = 1 - \Phi(x; \mu, \sigma)$. A fdp é dada por:

$$f(x; \mu, \sigma, \alpha) = \frac{\alpha \phi\left(\frac{x-\mu}{\sigma}\right) \Phi\left(\frac{x-\mu}{\sigma}\right) [1 - \Phi\left(\frac{x-\mu}{\sigma}\right)]^{\alpha-1}}{\sigma \{\Phi^\alpha\left(\frac{x-\mu}{\sigma}\right) + [1 - \Phi\left(\frac{x-\mu}{\sigma}\right)]^\alpha\}^2}, \quad (2.7)$$

respectivamente. Note que $\alpha > 0$ é um parâmetro de forma. A partir de agora, assumimos que uma variável aleatória com função de densidade como da equação (2.7) é denotado por $X \sim \text{OLLN}(\alpha, \mu, \sigma)$. Observe que para $\mu = 0$ e $\sigma = 1$, obtém-se a distribuição OLLN padrão. Além disso, a distribuição OLLN com $\alpha = 1$ reduz para a distribuição Normal.

Observe que o novo modelo apresentado pela equação (2.7) possui apenas um parâmetro adicional (α) e dois parâmetros da distribuição Normal (μ, σ). Isso se compara favoravelmente com outras famílias, por exemplo, a família beta generalizada (B-G) proposta por EUGENE *et al.* (2002), que inclui dois parâmetros de forma extras e sua fda depende da função beta incompleta; a família generalizada Kummer Beta (KB-G), PESCI *et al.* (2012), que envolve três parâmetros extras e sua fda depende da função hipergeométrica confluyente; a classe McDonald generalizada (Mc-G), ALEXANDER *et al.* (2012), que inclui três parâmetros extras e sua fda também depende da função beta incompleta.

A Figura 2.4 apresenta a fdp da OLLN para alguns valores dos parâmetros μ, σ e α . Observa-se nos gráficos 2.4(a) e 2.4(b) a contribuição do parâmetro α na unimodalidade e bimodalidade da distribuição, quando μ e σ são fixados. Quando o parâmetro α aproxima-se de zero, a fdp apresenta bimodalidade. Por outro lado, quando o valor de α aumenta, a função apresenta unimodalidade. Além disso, quando μ varia, os gráficos são transladados no eixo das abcissas, independentemente da forma como mostra a Figura 2.4(c).

Figura 2.4. Gráficos da fdp da OLLN para os diferentes valores dos parâmetros.

2.5.6 Avaliação do modelo

Para a seleção do modelo que melhor se ajusta aos dados, são usados alguns critérios ou testes estatísticos. Os critérios mais utilizados para a seleção de modelo na prática são os critérios de informação de Akaike (AIC), proposto por AKAIKE (1974) e o critério de informação de Bayes (BIC), proposto por SCHWARZ *et al.* (1978). Esses critérios são baseados no logaritmo da função de verossimilhança estimada. O AIC penaliza modelos com grande número de parâmetros k , uma vez que se espera que o logaritmo da função de verossimilhança cresça com o número de parâmetros adicionados ao modelo.

Várias correções deste critério foram propostas, a fim de reduzir a probabilidade de escolha de modelos com maior ordem do que o desejado. Entre eles, existe o critério de informação corrigido de Akaike (AICc), proposto por AKAIKE (1998).

Uma abordagem diferente para escolha do melhor modelo é o uso do estudo de estatísticas modificadas, de Anderson-Darling (W^*) e Cramer-von Mises (A^*), proposto por LIN *et al.* (2008) e PAKYARI and BALAKRISHNAN (2012), respectivamente.

Estas estatísticas foram calculadas da seguinte forma: Dado o estimador de máxima verossimilhança (EMV) $\hat{\theta}$, para um vetor n -dimensional θ da amostra censurada direita Tipo-II observada $x_1 < \dots < x_n$. Então, calcula-se $v_i = F(x_i; \theta)$, e converte a amostra censurada v_i para uma amostra completa do tamanho n , $v_1^*, \dots, v_r^*$. Então, calcula-se $y_i = \Phi^{-1}(v_i)$, em que $\Phi^{-1}(\cdot)$ é a fda da inversa da distribuição Normal padrão.

O modelo que apresentar o menor valor dessas estatísticas deve ser o escolhido.

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

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

$$A^* = A^2 (1 + 0.75/n + 2.25/n^2)$$

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

2.5.7 Tendência na Série de Produtividade

No período de estudo (1980 a 2014), houve um grande avanço nas tecnologias empregadas nas culturas, como técnicas de plantio, insumos, maquinários, entre outros, o que sugere uma tendência nos dados. Além disso, pode-se esperar dependência do tempo e variância não constante ao longo do tempo. Assim, antes de se ajustar qualquer modelo probabilístico para a série de produtividade, é necessário utilizar técnicas estatísticas que resultem em uma série sem tendência, independente e homoscedástica.

A correção da tendência é realizada pela mesma abordagem utilizada nos trabalhos de GALLAGHER (1987), GOODWIN and MAHUL (2004) e TEJEDA *et al.* (2008). Este procedimento inicialmente estima um modelo linear de produtividade e tempo, dado por:

$$y_t = \alpha + \beta T + \gamma T^2 + e_t, \quad e_t \sim N(\mu, \sigma), \quad (2.8)$$

em que y_t é o vetor de produtividade, T é o vetor de tempo, α , β e γ os parâmetros de regressão. Utilizando os resíduos $\hat{e}_t$ da regressão, a estimativa da última observação através do modelo ajustado $\hat{y}_{2014}$, a tendência é retirada de acordo com a equação:

$$\tilde{y}_t = \hat{y}_{2014} \left(1 + \frac{\hat{e}_t}{\hat{y}_t} \right). \quad (2.9)$$

Para a verificação da dependência temporal, é aplicado o teste proposto por LJUNG and BOX (1978), em que a hipótese nula é que existe independência na série. Além disso, para verificar a homocedasticidade, foi aplicado o teste BREUSCH and PAGAN (1979), no qual a hipótese nula é que a série é homoscedástica.

As estimativas dos parâmetros para as distribuições paramétricas foram obtidas por máxima verossimilhança. Os detalhes são apresentados no Anexo A.

2.6 Resultados

Aplicando o teste de dependência temporal para os municípios de Cascavel (p-valor = 0,7416), Guarapuava (p-valor = 0,1013), Castro (p-valor = 0,1004) e Palmeira (p-valor = 0,2597), conclui-se que não há dependência temporal nas séries de produtividade.

Observando os p-valores do teste de homocedasticidade para os municípios de Cascavel (p-valor = 0,624), Guarapuava (p-valor = 0,1785), Castro (p-valor = 0,2412) e Palmeira (p-valor = 0,0437), conclui-se que as séries são homocedásticas.

A Tabela 2.1 apresenta algumas estatísticas descritivas para as séries selecionadas corrigidas quanto à tendência, dependência temporal e heterocedasticidade. Note que, para todos os municípios, o coeficiente de assimetria é negativo e a mediana é maior do que a média, com exceção de Palmeira, sugerindo uma assimetria negativa. Além disso, o município de Castro apresenta uma curva leptocúrtica. Os municípios de Guarapuava e Castro apresentam coeficiente de variação menor quando comparados com outros municípios, sugerindo menor risco de produção para esses municípios. Para todos os municípios, a produtividade esperada é superior a 3300,00 kg/ha.

A Figura 2.5 apresenta a série de produtividade original e corrigida (sem tendência). Observa-se que a série sem tendência mostra um comportamento semelhante à série original em termos de variabilidade.

A Figura 2.6 apresenta as densidades ajustadas pelos modelos paramétricos. Para os municípios de Castro e Palmeira, não houve convergência para o modelo Beta, portanto, não está presente nas figuras. Além disso, os modelos SN e ST se sobrepõem ao modelo N para Palmeira. Note que, em séries com bimodalidade, como Guarapuava e Palmeira, a distribuição OLLN captura essa característica. Portanto, há um grande avanço para as distribuições paramétricas estudadas na modelagem da produtividade aplicadas ao seguro de produtividade.

Tabela 2.1. Estatísticas descritivas para as séries corrigidas.

	Municípios			
	Cascavel	Guarapuava	Castro	Palmeira
Média	3268,108	3380,111	3526,370	3244,340
Mediana	3310,269	3411,745	3547,058	3210,203
Desvio Padrão	438,727	282,186	231,310	328,003
Coef. Assimetria	-0,075	-0,262	-0,920	-0,369
Curtose	2,908	2,440	5,431	2,219
Máximo	4277,174	3925,333	4014,058	3701,117
Mínimo	2280,380	2703,796	2742,989	2473,699
Produtividade Esperada	3403,802	3589,257	3589,257	3337,692
Coeficiente de Variação	0,134	0,083	0,066	0,101

Na Tabela 2.2 são apresentados os critérios de informação AIC, BIC e as estatísticas A^* e W^* . Observa-se que, para todos os municípios, o modelo que melhor se ajusta aos dados é o modelo OLLN. Portanto, as taxas ajustadas pelo modelo OLLN são mais precisas do que as taxas ajustadas pelas outras distribuições.

Tabela 2.2. Estatísticas e critério de informação para os modelos selecionados.

Municípios	Modelo	A^*	W^*	AICc	BIC
Cascavel	N	0,2190	0,0354	528,5575	531,2932
	SN	0,2190	0,0354	530,9567	534,8486
	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
Guarapuava	N	0,303	0,050	497,665	500,4015
	SN	0,2459	0,037	499,414	503,3063
	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
Castro	N	0,628	0,0999	483,7492	486,4849
	SN	0,4334	0,0646	483,5584	487,4503
	OLLN	0,338	0,0557	482,6576	486,5494
	ST	0,8002	0,0948	484,1971	488,3496
Palmeira	N	0,5713	0,0873	508,1976	510,9333
	SN	0,5713	0,0873	510,5968	514,4887
	OLLN	0,3481	0,0397	506,3232	510,2150
	ST	0,9497	0,1078	513,8916	518,0440

Depois de escolher a distribuição que melhor se ajusta aos dados, o próximo passo é calcular as taxas do prêmio de seguro de produtividade para cada município, de acordo com a equação (2.2) e comparar os resultados com os estimados pela distribuição N, comumente utilizado pelo mercado de seguros.

2.7 Discussões

A Tabela 2.3 ⁴ apresenta as taxas atuariais justas (taxas puras) do prêmio calculado, utilizando diferentes distribuições para a cidade de Cascavel.

Observa-se que, para níveis de cobertura que variam de 55% a 70%, a distribuição OLLN superestima a taxa do prêmio quando comparada com a distribuição N, que é amplamente utilizada pelo

⁴Apresentamos vários níveis de cobertura para compararmos com o nível de cobertura equivalente em nosso modelo.

Figura 2.5. Série de Produtividade Original e Corrigida.

mercado de seguros. Equivalentemente, o modelo N subestima a taxa de seguro. A implicação para a seguradora, neste caso, é a redução do volume total de prêmio quando se utiliza a distribuição N na precificação.

No entanto, para os níveis de cobertura que variam de 75% e 80%, a distribuição OLLN subestima a taxa prêmio em comparação com a distribuição N. Ou seja, a taxa do prêmio calculada considerando a distribuição N superestima a taxa OLLN. A superestimação pode dificultar a massificação do seguro e atrair produtores com maior perfil de risco, aumentando o problema da seleção adversa.

Esta análise também foi realizada para os outros municípios, mas optamos por não apresentar os resultados, visto que o principal objetivo é comparar as taxas comerciais do modelo OLLN com as demais taxas comerciais dos modelos. As taxas comerciais para todos os municípios são apresentadas na Tabela 2.4.

A taxa pura (TP) representa o risco intrínseco do negócio, sem incluir nenhum custo adicional.

Figura 2.6. Distribuições Ajustadas para as séries corrigidas.

Para compará-lo com as taxas oferecidas pelas seguradoras precisamos incluir parâmetros médios do mercado como a margem técnica, despesas administrativas e de negócios e margem de lucro da seguradora.

De acordo com BRISOLARA (2013), algumas seguradoras brasileiras adicionam uma margem de segurança de 10% a 25% às taxas puras, após a inclusão de despesas administrativas, comerciais e margem de lucro, que correspondem a uma média de 10% e 20%, respectivamente. Assim, neste estudo, optou-se por utilizar um fator de carregamento de 40%, o que significa dividir a taxa pura com a margem de segurança em 0,6. Os parâmetros de mercado adotados são 20% para a margem de segurança, 10% para despesas comerciais, 20% para despesas administrativas e 10% para margem de lucro.

As taxas comerciais (TC) com uma margem de segurança de 20% foram obtidas da seguinte

Tabela 2.3. Taxa Pura (em %) - Cascavel.

NC	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

forma:

$$TC = \frac{TP * 1,2}{[1 - (0,1 + 0,2 + 0,1)]} = \frac{TP * 1,2}{0,6}. \quad (2.10)$$

O procedimento de precificação utilizado por cada seguradora não está disponível publicamente. No entanto, é sabido que a maioria das seguradoras utiliza a distribuição N para modelagem da produtividade. As seguradoras também adicionam uma margem de segurança com base nos dados históricos da relação de perda para cada município.

A Tabela 2.4 apresenta as taxas comerciais com uma margem de segurança de 20% para cada modelo e as taxas comerciais cobradas pelas companhias de seguros A, B, C e D para os municípios de Cascavel, Guarapuava, Castro e Palmeira.

Ao comparar as taxas calculadas neste estudo com as aplicadas pelas seguradoras, nota-se que estas são muito inferiores às praticadas pelo mercado. Como exemplo, a taxa comercial calculada pela distribuição OLLN (Tabela 2.4) para Cascavel, com um nível de cobertura (NC) de 55%, corresponde a 0,25% e 0,22% das taxas comerciais das seguradoras B e C, respectivamente. Com um nível de cobertura de 65%, a taxa comercial calculada pela distribuição OLLN corresponde a 3,58% da taxa comercial da seguradora C. Observe que a taxa de 6,66% oferecida pela seguradora B, com um nível de cobertura de 60%, representa aproximadamente a taxa comercial calculada pelo modelo OLLN, com um NC de 79%. Por isso, os resultados sugerem que as seguradoras oferecem um produto inferior com uma alta taxa, dificultando a massificação do seguro no Brasil.

Para Guarapuava, as taxas calculadas pela distribuição N são excessivas quando comparadas as outras distribuições, exceto com a SN. Além disso, nota-se que a taxa 6,70% oferecida pela seguradora B com NC de 65% representa, aproximadamente, a taxa comercial calculada pelo modelo OLLN com NC de 87%. Em Castro, há também uma superestimação da taxa calculada pela distribuição N, quando comparada à distribuição OLLN com NC de 87%.

Por outro lado, o município de Palmeira oferece maior variabilidade e, consequentemente, maiores riscos e taxas. Para NC de 70% a 80%, a distribuição N superestima a taxa quando comparada à distribuição OLLN. No entanto, para 83% ou mais níveis de cobertura, há uma subestimação da taxa. Além disso, a taxa calculada pela distribuição OLLN no nível de cobertura de 70% equivale a 0,12% da taxa utilizada pela seguradora B. Isso é uma evidência de que as seguradoras brasileiras utilizam altas taxas no seguro de produtividade quando comparadas com as taxas calculadas neste estudo.

Este distanciamento das taxas pode ser explicado pelo fato de que o seguro não é comercializado em massa no Brasil, além de se concentrar principalmente na região sul do país, uma região de maior ocorrência de eventos climáticos atípicos. O que resulta numa concentração de risco para a seguradora que, consequentemente, precisa fazer uma taxa de carregamento adicional à taxa de risco puro, além da taxa de carregamento natural, se o seguro fosse massificado no Brasil.

As diferenças nas taxas de prêmio também podem ser explicadas pelo fato de que supõe-se que todos os produtores deste município tenham contratado o seguro, ou seja, presume-se que o seguro

⁵Para os municípios de Castro e Palmeira não houve convergência do modelo Beta.

Tabela 2.4. Taxa Comercial (em %) com margem de segurança de 20% para os municípios de Cascavel, Guarapuava, Castro e Palmeira.

Cascavel	NC	N	SN	OLLN	Beta	ST	A	B	C	D
	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	NC	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	NC	N	SN	OLLN	Beta ⁵	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	NC	N	SN	OLLN	Beta	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				

seja vendido massivamente na cidade, levando a uma taxa mais baixa. Neste estudo, usamos a média de produtividade municipais, reduzindo a variabilidade individual como mostrado no estudo de JUST and WENINGER (1999).

2.8 Conclusões

Este estudo avaliou distribuições de probabilidade alternativas para modelagem da produtividade de soja e estimou as taxas do prêmio de seguro de produtividade. As séries foram ajustadas considerando as distribuições N, Beta, SN, OLLN e ST para municípios selecionados do Estado do Paraná, Brasil.

Concluimos que, para todas as séries, o modelo OLLN foi o mais adequado. Além disso, a bimodalidade encontrada nos municípios de Guarapuava e Palmeira foi considerada no cálculo da taxa do prêmio do seguro.

Para a cidade de Cascavel, a taxa estimada pelo modelo N é subestimada quando comparadas às do modelo OLLN, com NC que variam entre 55% a 80%. No entanto, para NC de 83% para 90%, o modelo N superestima a taxa quando comparado ao modelo OLLN.

Para a cidade de Guarapuava, há uma superestimação das taxas calculadas pelo modelo N quando comparado ao modelo OLLN. Quanto ao município de Palmeira, a taxa é alta para os NC de 79% para 80% e subestimado para níveis de cobertura superiores a 83%. Finalmente, há uma superestimação da taxa calculada pelo modelo N para o município de Castro com NC de 87% ou mais.

A subestimação da taxa pode levar a sérios danos à seguradora, porque oferece um produto que leva em consideração um risco menor do que deveria ser considerado. Além disso, a superestimação

da taxa dificulta a expressiva venda de seguros no Brasil, além de atrair agricultores com maiores riscos, fortalecendo o problema de seleção adversa.

Grande parte do distanciamento da taxa estimada neste trabalho com aquelas aplicadas pelo mercado segurador pode ser explicada pelo fato do seguro brasileiro apresentar grande concentração na região sul (não é massificado) e pelo atraso dos pagamentos dos subsídios governamentais às seguradoras, o que pode encorajá-las a aumentar as suas taxas.

Historicamente, no mercado brasileiro, o seguro agrícola foi evitado pelos produtores, por apresentarem altas taxas. Por outro lado, as seguradoras se deparavam com uma probabilidade alta de pagar elevados valores de indenizações. Uma forma de contornar esse problema é a correta precificação do risco agrícola resultando em taxas mais precisas e que refletem o risco de cada propriedade agrícola.

Referências

- AKAIKE, H., 1974 A new look at the statistical model identification. *Automatic Control, IEEE Transactions on* **19**: 716–723.
- AKAIKE, H., 1998 Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike*, pp. 199–213, Springer.
- ALEXANDER, C., G. M. CORDEIRO, E. M. ORTEGA, and J. M. SARABIA, 2012 Generalized beta-generated distributions. *Computational Statistics & Data Analysis* **56**: 1880–1897.
- ALIZADEH, M., M. EMADI, M. DOOSTPARAST, G. M. CORDEIRO, E. M. ORTEGA, and R. R. PESCI, 2015 A new family of distributions: the kumaraswamy odd log-logistic, properties and applications. *Hacetatepe Journal of Mathematics and Statistics* **44**.
- ALIZADEH, M., S. MIRMOSTAFEE, E. M. ORTEGA, T. G. RAMIRES, and G. M. CORDEIRO, 2017 The odd log-logistic logarithmic generated family of distributions with applications in different areas. *Journal of Statistical Distributions and Applications* **4**: 6.
- AZZALINI, A., 1985 A class of distributions which includes the normal ones. *Scandinavian journal of statistics* pp. 171–178.
- AZZALINI, A. and A. CAPITANIO, 2003 Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65**: 367–389.
- BABCOCK, B. A. and D. A. HENNESSY, 1996 Input demand under yield and revenue insurance. *American journal of agricultural economics* **78**: 416–427.
- BOTTS, R. R. and J. N. BOLES, 1958 Use of normal-curve theory in crop insurance ratemaking. *Journal of Farm Economics* pp. 733–740.
- BRAGA, A. S., G. M. CORDEIRO, and E. M. ORTEGA, 2017 A new skew-bimodal distribution with applications. *Communications in Statistics-Theory and Methods* pp. 1–19.
- BRAGA, A. S., G. M. CORDEIRO, E. M. ORTEGA, and J. N. DA CRUZ, 2016 The odd log-logistic normal distribution: Theory and applications in analysis of experiments. *Journal of Statistical Theory and Practice* **10**: 311–335.
- BREUSCH, T. S. and A. R. PAGAN, 1979 A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society* pp. 1287–1294.

- BRISOLARA, C. S., 2013 *Proposições para o desenvolvimento do seguro de receita agrícola no Brasil: do modelo teórico ao cálculo das taxas de prêmio*. Ph.D. thesis, Escola Superior de Agricultura “Luiz de Queiroz.
- COBLE, K. H. and T. O. KNIGHT, 2002 Crop insurance as a tool for price and yield risk management. In *A Comprehensive Assessment of the Role of Risk in US Agriculture*, pp. 445–468, Springer.
- COBLE, K. H., T. O. KNIGHT, R. D. POPE, and J. R. WILLIAMS, 1996 Modeling farm-level crop insurance demand with panel data. *American Journal of Agricultural Economics* **78**: 439–447.
- CONAB, 2016 Monitoramento agrícola – cultivos de inverno (safra 2015) e de verão (safra 2015/16). Disponível em http://www.conab.gov.br/OlalaCMS/uploads/arquivos/16_02_21_003026oletim_a16_v5_n01_e02.pdf.
- CRUZ, J. N., 2016 *A nova família de distribuições odd log-logística: teoria e aplicações*. Ph.D. thesis, Escola Superior de Agricultura “Luiz de Queiroz.
- CRUZ, J. N., E. M. ORTEGA, and G. M. CORDEIRO, 2016 The log-odd log-logistic weibull regression model: modelling, estimation, influence diagnostics and residual analysis. *Journal of Statistical Computation and Simulation* **86**: 1516–1538.
- CRUZ, J. N. D., E. M. ORTEGA, G. M. CORDEIRO, A. K. SUZUKI, and F. L. MIALHE, 2017 Bivariate odd-log-logistic-weibull regression model for oral health-related quality of life. *Communications for Statistical Applications and Methods* **24**: 271–290.
- DAY, R. H., 1965 Probability distributions of field crop yields. *Journal of Farm Economics* **47**: 713–741.
- EUGENE, N., C. LEE, and F. FAMOYE, 2002 Beta-normal distribution and its applications. *Communications in Statistics-Theory and methods* **31**: 497–512.
- GALLAGHER, P., 1987 Us soybean yields: estimation and forecasting with nonsymmetric disturbances. *American Journal of Agricultural Economics* **69**: 796–803.
- GOODWIN, B. K., 2001 Problems with market insurance in agriculture. *American Journal of Agricultural Economics* **83**: 643–649.
- GOODWIN, B. K. and A. P. KER, 1998 Nonparametric estimation of crop yield distributions: implications for rating group-risk crop insurance contracts. *American Journal of Agricultural Economics* **80**: 139–153.
- GOODWIN, B. K. and O. MAHUL, 2004 Risk modeling concepts relating to the design and rating of agricultural insurance contracts. World Bank Policy Research Working Paper .
- HARWOOD, J. L., R. HEIFNER, K. COBLE, J. PERRY, A. SOMWARU, *et al.*, 1999 *Managing risk in farming: concepts, research, and analysis*. US Department of Agriculture, Economic Research Service.
- HENNESSY, D. A., B. A. BABCOCK, and D. J. HAYES, 1997 Budgetary and producer welfare effects of revenue insurance. *American Journal of Agricultural Economics* **79**: 1024–1034.
- IPARDES, 2015 Instituto paranaense de desenvolvimento econômico e social. base de dados do estado (bdweb). Acessado em: 02/julho/2015. Disponível em: <http://www.ipardes.pr.gov.br/imp/index.php>.
- JOHNSON, N. L. and N. BALAKRISHNAN, 1995 *Continuous Univariate Distributions*, volume 2. Wiley.
- JUST, R. E. and Q. WENINGER, 1999 Are crop yields normally distributed? *American Journal of Agricultural Economics* **81**: 287–304.

- KER, A. P. and K. COBLE, 2003 Modeling conditional yield densities. *American Journal of Agricultural Economics* **85**: 291–304.
- KER, A. P. and B. K. GOODWIN, 2000 Nonparametric estimation of crop insurance rates revisited. *American Journal of Agricultural Economics* **82**: 463–478.
- LAWAS, C. P., 2005 *Crop insurance premium rate impacts of flexible parametric yield distributions: an evaluation of the Johnson family of distributions*. Ph.D. thesis, Texas Tech University.
- LIN, C.-T., Y.-L. HUANG, and N. BALAKRISHNAN, 2008 A new method for goodness-of-fit testing based on type-II right censored samples. *Reliability, IEEE Transactions on* **57**: 633–642.
- LJUNG, G. M. and G. E. BOX, 1978 On a measure of lack of fit in time series models. *Biometrika* **65**: 297–303.
- MAPA, 2016 Seguro rural- ministério da agricultura, pecuária e abastecimento. Disponível em <http://www.agricultura.gov.br/politica-agricola/seguro-rural>.
- MIQUELETO, G. J., 2011 *Contribuições para o desenvolvimento do seguro agrícola de renda para o Brasil: evidências teóricas e empíricas*. Ph.D. thesis, Escola Superior de Agricultura “Luiz de Queiroz”.
- MOSS, C. B. and J. S. SHONKWILER, 1993 Estimating yield distributions with a stochastic trend and nonnormal errors. *American Journal of Agricultural Economics* **75**: 1056–1062.
- NELSON, C. H. and P. V. PRECKEL, 1989 The conditional beta distribution as a stochastic production function. *American Journal of Agricultural Economics* **71**: 370–378.
- OZAKI, V. A., 2005 *Métodos atuariais aplicados à determinação da taxa de prêmio de contratos de seguro agrícola: um estudo de caso*. Ph.D. thesis, Universidade de São Paulo.
- OZAKI, V. A., B. K. GOODWIN, and R. SHIROTA, 2008b Parametric and nonparametric statistical modelling of crop yield: implications for pricing crop insurance contracts. *Applied Economics* **40**: 1151–1164.
- OZAKI, V. A. and R. S. SILVA, 2009 Bayesian ratemaking procedure of crop insurance contracts with skewed distribution. *Journal of Applied Statistics* **36**: 443–452.
- PAKYARI, R. and N. BALAKRISHNAN, 2012 A general purpose approximate goodness-of-fit test for progressively type-II censored data. *Reliability, IEEE Transactions on* **61**: 238–244.
- PESCIM, R. R., G. M. CORDEIRO, C. G. DEMÉTRIO, E. M. ORTEGA, and S. NADARAJAH, 2012 The new class of kummer beta generalized distributions. *SORT-Statistics and Operations Research Transactions* **36**: 153–180.
- QUIGGIN, J. C., G. KARAGIANNIS, and J. STANTON, 1993 Crop insurance and crop production: an empirical study of moral hazard and adverse selection. *Australian Journal of Agricultural Economics* **37**: 95–113.
- R CORE TEAM, 2016 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- RAMÍREZ, O. A., 1997 Estimation and use of a multivariate parametric model for simulating heteroskedastic, correlated, nonnormal random variables: the case of corn belt corn, soybean, and wheat yields. *American Journal of Agricultural Economics* **79**: 191–205.

- RAMIREZ, O. A., S. MISRA, and J. FIELD, 2003 Crop-yield distributions revisited. *American Journal of Agricultural Economics* **85**: 108–120.
- SCHWARZ, G. *et al.*, 1978 Estimating the dimension of a model. *The annals of statistics* **6**: 461–464.
- SHERRICK, B. J., F. C. ZANINI, G. D. SCHNITKEY, and S. H. IRWIN, 2004 Crop insurance valuation under alternative yield distributions. *American Journal of Agricultural Economics* **86**: 406–419.
- SMITH, V. H. and B. K. GOODWIN, 1996 Crop insurance, moral hazard, and agricultural chemical use. *American Journal of Agricultural Economics* **78**: 428–438.
- TAYLOR, C. R., 1990 Two practical procedures for estimating multivariate nonnormal probability density functions. *American Journal of Agricultural Economics* **72**: 210–217.
- TEJEDA, H. A., B. K. GOODWIN, *et al.*, 2008 Modeling crop prices through a burr distribution and analysis of correlation between crop prices and yields using a copula method. In *annual meeting of the Agricultural and Applied Economics Association, Orlando, FL, USA*, volume 83, pp. 643–649, Citeseer.
- TURVEY, C., J. ZHAO, *et al.*, 1999 *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.

3 PRECIFICAÇÃO DOS RISCOS DO SEGURO DE FATURAMENTO AGRÍCOLA UTILIZANDO CÓPULAS PARAMÉTRICAS

Resumo

O seguro de faturamento agrícola tem sido amplamente discutido, a partir dos anos 2000. Este tipo de seguro tornou-se um mecanismo importante para o gerenciamento dos riscos climáticos (afetam a produtividade) e de mercado (preços das culturas). No entanto, um estudo mais abrangente é necessário para investigar a estrutura de dependência entre as variáveis analisadas e para calcular a taxa do prêmio atuarialmente justa para o seguro de faturamento. Este estudo propõe formas alternativas para calcular as taxas do prêmio do seguro de faturamento utilizando as funções de cópulas paramétricas. Estes métodos foram aplicados aos dados de produtividade de soja para os municípios de Toledo, Cascavel, Castro e Guarapuava no Estado do Paraná e preços nominais recebidos pelos produtores do Paraná (SEAB). Sob o enfoque paramétrico, os modelos que melhor se ajustaram aos dados de produtividade para todos os municípios foi o OLLN e para a série de preços foi o modelo Skew-t. Além disso, a cópula AMH melhor representa a estrutura de dependência entre as variáveis para os municípios de Cascavel e Toledo. Para o município de Guarapuava e Castro a cópula que melhor se ajustou foi a Gaussiana. Além disso, as taxas comerciais médias calculada pela metodologia de cópulas foram superiores àquelas aplicadas pelo mercado segurador.

Palavras Chaves: Cópulas paramétricas; Distribuição OLLN; Gestão de riscos na agricultura; Seguro de faturamento.

Abstract

Crop revenue insurance has been widely discussed recently. It has become an important mechanism for risk management of crop yield and prices. However, a more comprehensive study is needed to investigate the dependence structure between the variables analyzed to calculate the premium rate actuarially fair for revenue insurance. This study proposes alternatives to calculate premium rates for revenue insurance using parametric copula functions. These methods were applied to data on soybean yield in the municipalities of Toledo, Cascavel Castro. Nominal prices received by producers in Parana State were provided by the Secretariat of Agriculture and Supply of Parana (SEAB).

Keywords: Crop risk management; OLLN distribution; Parametric copulas; Revenue insurance.

3.1 Introdução

A agricultura é extremamente importante para o Brasil, tanto do ponto de vista econômico quanto social. No entanto, muitos riscos ameaçam em maior ou menor grau todas as cadeias do agro-negócio, entre eles, os riscos climáticos (que afetam a produtividade) e os riscos de mercado que podem levar a mudanças significativas nos preços das *commodities*.

O desenvolvimento de estratégias de gerenciamento de risco requer a compreensão da natureza do risco, sua origem, sua distribuição de probabilidade, correlação com outros riscos e a capacidade de instrumentos para reduzi-lo (HARDAKER *et al.*, 2004). Várias estratégias podem ser utilizadas pelo setor agrícola para o gerenciamento destes riscos, entre elas, a diversificação de culturas, novas técnicas de produção, derivativos agrícolas e seguros agrícolas.

O seguro agrícola é eficiente para proteger a renda do produtor em condições adversas e é um instrumento importante para transferir o risco dos produtores para outros agentes econômicos. O seguro garante a renda quando ocorre algum evento que cause danos econômicos comparando a sua renda com o cenário sem esse evento, mediante o pagamento de um prêmio e indenização, em caso de danos.

No Brasil, o primeiro seguro de faturamento foi oferecido em 2010. Em 2011, uma segunda empresa também começou a oferecer essa modalidade de seguro. Apesar deste tipo de produto ser relativamente recente no mercado brasileiro, a demanda para essa nova modalidade tem crescido anualmente (CARVALHO *et al.*, 2013).

Por ser um produto recente no mercado brasileiro, não existem trabalhos que avaliam os procedimentos atuariais para precificar este tipo de seguro. Consequentemente, há a necessidade de mais pesquisas sobre os aspectos teóricos dos preços do seguro, que levem em consideração as características específicas das séries de produtividade e preços da soja brasileira, justificando o objetivo deste trabalho.

De acordo com GOODWIN and MAHUL (2004), a principal questão do seguro de faturamento é a modelagem do risco, que consiste em calcular a distribuição conjunta, distribuições marginais e correlação entre variáveis aleatórias de produtividade e preço.

Em problemas multivariados, a distribuição normal multivariada é geralmente a mais utilizada. Esta escolha baseia-se principalmente na sua simplicidade matemática. No entanto, essa suposição de normalidade restringe o tipo de associação entre as distribuições marginais, tornando-a uma relação linear. Além disso, assume-se uma associação simétrica o que, na prática, pode não capturar a idiosincrasia dos dados.

No campo atuarial, os dados tendem a apresentar caudas mais pesadas do que a normal e estruturas de dependência que podem ser lineares, não-lineares e dependência somente nas caudas da distribuição. Portanto, o uso indevido da hipótese de normalidade pode resultar em cálculos equivocados, e consequentemente, levar a perdas financeiras, ou seja, sub ou superestimação das taxas que deveriam ser cobradas.

Este estudo apresenta formas alternativas para calcular e estimar a taxa do prêmio do seguro de faturamento utilizando cópulas paramétricas na modelagem da distribuição conjunta entre preço e produtividade, levando em consideração as marginais aproximadas pelo método empírico (não paramétrico) e o método paramétrico com a distribuição Odd-Log Logística Normal (OLLN) para a produtividade e a distribuição Skew-t-Student para a série de preços. Além disso, as taxas calculadas pela metodologia de cópula foram comparadas com distribuição normal bivariada, geralmente utilizada no mercado de seguros e com taxa comercial aplicada pelas seguradoras.

A próxima seção apresenta uma revisão da literatura. A seção 3.1.2 apresenta as formas de calcular o prêmio do seguro de faturamento agrícola. A seção 3.2 descreve os dados utilizados neste trabalho. A seção 3.3 apresenta a análise paramétrica para as distribuições marginais. A metodologia de cópulas é apresentada na seção 3.4. A seção 3.5 apresenta os resultados, a seção 3.6 apresenta a taxa do seguro de faturamento e, por fim, a seção 3.7 apresenta as conclusões deste trabalho.

3.2 Revisão de Literatura

Uma modelagem precisa das variáveis aleatórias de produtividade e preço é necessária para calcular a taxa do prêmio atuarial justo. Três abordagens são usadas para modelar essas variáveis. A primeira envolve a estimativa de parâmetros da distribuição paramétrica. A segunda requer uma grande variedade de métodos não paramétricos comumente utilizados para modelar distribuições aproximadas. Por fim, os semi-paramétricos são métodos alternativos que combinam elementos dos métodos paramétricos e não paramétricos (KER and COBLE, 2003).

Os métodos não paramétricos são vantajosos, uma vez que não requerem especificação prévia do formato da distribuição, o que significa que os "dados falam por si mesmos". Neste caso, algumas características da distribuição podem ser mostradas, como a assimetria positiva, negativa e a bimodalidade. No entanto, esses métodos requerem uma grande amostra, apesar de apresentarem uma maior flexibilidade para descrever várias formas da densidade (GOODWIN and MAHUL, 2004).

Sob o enfoque paramétrico, JUST and WENINGER (1999) e BOTTS and BOLES (1958) sugeriram a distribuição normal para a produtividade. No entanto, outros autores DAY (1965), RAMIREZ *et al.* (2003) e TAYLOR (1990) encontraram evidências contra a normalidade em suas pesquisas.

Além da distribuição normal, outras distribuições paramétricas foram consideradas na modelagem da produtividade, tais como, a distribuição Gamma, GALLAGHER (1987), Hiperbólica Inversa, MOSS and SHONKWILER (1993), Beta e Weibul, SHERRICK *et al.* (2004), e distribuição Skew-Normal OZAKI and SILVA (2009), SANTOS (2011). No entanto, a distribuição Beta é a distribuição paramétrica mais utilizada, NELSON and PRECKEL (1989), BABCOCK and HENNESSY (1996), COBLE *et al.* (1996) e HENNESSY *et al.* (1997).

DUARTE *et al.* (2017) comparou as distribuições Normal, Beta, Skew-normal, Skew-t Student e a distribuição Odd Log-Logística Normal (OLLN) para o modelo de produtividade de soja em quatro municípios do estado do Paraná. Nesses municípios, a distribuição OLLN melhor se ajustou aos dados. Além disso, os autores compararam as taxas do seguro de produtividade calculadas pela distribuição OLLN com taxas comerciais aplicadas pelas seguradoras, indicando que as taxas calculadas pela distribuição OLLN foram menores que as aplicadas pelas seguradoras.

Considerando a abordagem não-paramétrica, o estimador kernel é sugerido para a função de densidade de probabilidade de produção GOODWIN and KER (1998), TURVEY and ZHAO (1999), KER and GOODWIN (2000) e OZAKI *et al.* (2008).

Na modelagem dos preços, a distribuição log-normal foi utilizada em estudos sobre seguro agrícola, por exemplo por, GOODWIN *et al.* (2000), HERNAN TEJEDA *et al.* (2008). As abordagens semi-paramétricas e não paramétricas são comumente usadas como alternativas à distribuição log-normal.

De acordo com GOODWIN and KER (2002), a principal dificuldade em modelar o faturamento agrícola reside na necessidade de determinar o grau de correlação entre o preço e a produtividade, uma vez que estes são raramente independentes.

Essa dependência entre variáveis pode ser estudada através de funções cópulas. A teoria das cópulas tem sido frequentemente utilizada para analisar séries financeiras e fatores de risco, como mostrado em BOUYÉ *et al.* (2000), EMBRECHTS *et al.* (2003) e SCHÖNBUCHER and SCHUBERT (2001).

No caso do seguro de faturamento, poucos estudos foram elaborados. No trabalho de GOODWIN and TEJEDA (2008), os autores utilizaram o método de cópulas para medir a correlação entre preço e produtividade e avaliar o seu impacto no cálculo das taxas do prêmio do seguro de faturamento. A série de preços foi modelada pela distribuição Burr e log-normal, e as séries de produtividade, pela distribuição Beta. Os resultados mostraram menor probabilidade de pagamento de indenizações quando o método da cópula foi utilizado em comparação com os métodos atuais.

De acordo com VEDENOV (2008), a metodologia de cópula é uma técnica alternativa de modelagem de distribuições conjuntas de variáveis aleatórias, com grande flexibilidade para as marginais e estrutura de dependência. O autor modelou a distribuição conjunta de produtividade de um município e propriedades dentro deste município e concluiu que a metodologia de cópula é eficiente e pode contribuir para a gestão de riscos e desenvolvimento de seguros.

AHMED and SERRA (2015) investigaram os efeitos dos contratos de seguros de receita sobre os custos de seguros em relação aos esquemas de seguro na Espanha. Uma análise empírica foi realizada nos setores de maçã e laranja usando cópulas estáticas para modelar preços e produtividade de culturas. As taxas de seguro de receita foram simuladas pelos métodos de Monte Carlo, o que mostrou que o seguro de receita provavelmente reduz o preço do seguro agrícola na Espanha, resultando em maior aceitação e demanda por programas de seguro agrícolas.

No contexto brasileiro, MIQUELETO (2011) discutiu o uso de cópulas para calcular a taxa do prêmio do seguro de faturamento para os municípios do Estado do Paraná. Ao modelar as marginais, a distribuição não paramétrica foi utilizada com o kernel Gaussiano para a série de preços e a série

de produtividade de milho foi simulada da distribuição Beta. O autor conclui que existem diferenças significativas entre as taxas de prêmio de diferentes municípios e em relação aos diferentes meses de vencimento dos contratos futuros negociados na Bovespa (BM&F), utilizados na análise. Além disso, o pressuposto de que a produtividade de milho brasileira segue uma distribuição Beta foi realizada com base no trabalho de NELSON and PRECKEL (1989), BABCOCK and HENNESSY (1996), COBLE *et al.* (1996) e HENNESSY *et al.* (1997). No entanto, as diferenças nos sistemas de clima, geografia, plantação e produção não permitem afirmar que a produtividade no Brasil apresenta as mesmas características que a produtividade nos Estados Unidos.

3.3 Seguro de Faturamento Agrícola

O seguro de faturamento cobre as perdas de produtores após a colheita, ou seja, a perda de produtividade causada principalmente por eventos climáticos, reduções de preços ou ambos.

O modelo de seguro utilizado como referência neste trabalho foi "revenue protection"(RP), adotado no mercado dos Estados Unidos a partir de 2011, substituindo o antigo "revenue assurance"(RA). Os conceitos teóricos para calcular as taxas do seguro de faturamento foram descritos em MIQUELETO (2011).

Este trabalho considerou o faturamento (F) como uma função de duas variáveis, a produtividade (X) e o preço (Y), cuja expressão é $F = X.Y$. A produtividade garantida é definida por $X_g = \lambda X_e$, em que $0 < \lambda < 1$ é o nível de cobertura (NC) escolhido pelo produtor e X_e é a produtividade esperada, normalmente calculada pela média das últimas cinco safras.

O preço garantido é definido por $Y_g = \lambda Y_e$, em que $0 < \lambda < 1$ é o NC e Y_e é o preço esperado, normalmente calculado pela média dos últimos 15 dias do preço futuro negociado na CME no dia da contratação do seguro. No entanto, neste trabalho, o preço esperado utilizado foi a média dos últimos 15 preços da distribuição ajustada. Neste tipo de seguro, a indenização paga por unidade de área ao produtor é dada por $I = \max\{(x_g y_g - XY); 0\}$.

A probabilidade de perda é dada por $p = \text{Prob}(X < x; Y < y | XY < x_g y_g)$. Esta probabilidade condicional expressa a probabilidade cumulativa conjunta de um evento (perda) dependendo das variáveis aleatórias X e Y , dado que a escolha dos valores de XY é restrita à condição $XY < x_g y_g$.

Portanto, a taxa do prêmio do seguro é dada pela perda esperada, de acordo com a equação:

$$E[\text{Perda}] = E(\max\{(x_g y_g - XY); 0\}) = \text{Prob}(X < x; Y < y | XY < x_g y_g)[x_g y_g - E(XY | XY < x_g y_g)].$$

A taxa ótima do prêmio do seguro de faturamento é dada por:

$$\pi = \frac{\text{Prob}(X < x; Y < y | XY < x_g y_g)[x_g y_g - E(XY | XY < x_g y_g)]}{x_g y_g}. \quad (3.1)$$

O faturamento esperado diz respeito aos riscos nas marginais preço e produtividade. A distribuição marginal de preço e produtividade é deduzida de dados históricos utilizando métodos empíricos e paramétricos. A construção da distribuição conjunta de preço e produtividade, por sua vez, utiliza uma função de cópula paramétrica.

3.4 Descrição dos dados

Para a modelagem da produtividade foram analisadas as séries de produtividade anual de soja (Kg/ha) nos municípios de Toledo, Cascavel, Guarapuava e Castro, no estado do Paraná (Brasil). Estes dados são disponibilizados pelo Instituto de Desenvolvimento Econômico e Social de Paraná (IparDES) IPARDES (2015), disponível de 1980 a 2015.

A série de preços utilizada é o preço mensal nominal recebido pelos produtores do Estado do Paraná em R\$, cuja unidade é a saca de 60 Kg. Estes dados são disponibilizados pela Secretaria de Agricultura e Abastecimento do Estado do Paraná SEAB (2015).

Com o objetivo de igualar a periodicidade das duas séries, utilizou-se os preços nominais médios no período de vendas da safra, correspondentes aos meses de colheita de soja, março, abril e maio de cada ano. Além disso, a série de preços foi deflacionada pelo índice geral de preços - disponibilidade interna (IGP-DI), disponível em IPEADATA (2015).

As taxas comerciais da companhia de seguros A, utilizadas neste estudo, foram fornecidas pelo Ministério da Agricultura, Pecuária e Abastecimento de Alimentos MAPA (2017).

3.5 Análise Paramétrica para Distribuição das Marginais

Na modelagem da produtividade, utilizou-se as distribuições Normal, Skew-Normal (SN), proposta por AZZALINI (1985), as distribuições OLLN, BRAGA *et al.* (2016), e Skew- *t* Student (ST). Na modelagem dos preços utilizou-se as distribuições Log-Normal e Skew- *t* Student, (AZZALINI and CAPI-TANIO, 2003).

Para selecionar o modelo que melhor se ajustou aos dados, foram utilizados os critérios e testes estatísticos, conforme descritos na seção 2.5.6. A retirada da tendência das séries de produtividade e preços segue o procedimento descrito na seção 2.5.7

Para verificação da dependência temporal, foi aplicado o teste proposto por LJUNG and BOX (1978), em que a hipótese nula é que existe independência na série. Além disso, para verificar a homoscedasticidade das variâncias, o teste de LJUNG and BOX (1978) foi aplicado à série ao quadrado.

3.6 Cópuas

A teoria das cópuas é uma ferramenta de modelagem multivariada amplamente utilizada em diferentes ramos da ciência em que o interesse está na dependência multivariada e o uso da normalidade multivariada está em questão. Na ciência financeira e atuarial, as cópuas são usadas na modelagem de eventos correlacionados e de riscos competitivos (MIQUELETO, 2011).

Essa teoria torna-se atraente, pois as cópuas cobrem uma ampla gama de estruturas de dependência e são capazes de modelar completamente a estrutura de dependência dos dados. Uma função cópula é definida da seguinte forma:

Definição 1. Uma cópula é uma função $C : I^2 \rightarrow I$, com $I = [0, 1]$, que satisfaz para $0 < u_1 \leq u_2 < 1, 0 < v_1 \leq v_2 < 1$, (u_1, v_1) e (u_2, v_2) em I^2 , as seguintes condições:

1. $C(u, 0) = C(0, v) = 0$, e $C(u, 1) = u, C(1, v) = v$;
2. $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$.

A relação 1 mostra a uniformidade das distribuições das marginais e a relação 2 significa que, se (X, Y) tiver uma função de distribuição C , então $P(u_1 \leq X \leq u_2, v_1 \leq Y \leq v_2) \geq 0$. Simplificando, uma cópula é uma função que liga as distribuições das marginais a distribuição conjunta MORETTIN (2008).

Além disso, as cópuas permitem relacionar uma distribuição conjunta $H(x, y)$ com suas distribuições marginais $F(x)$ e $G(y)$. O teorema de Sklar, proposto por SKLAR (1959), garante as condições de representação da singularidade da função de distribuição $H(x, y)$ por meio de cópuas, descrito abaixo.

Teorema 1. Seja H uma função de distribuição bivariada com marginais F e G . Assim, existe uma cópula bivariada C tal que $\forall (x, y) \in \mathbb{R}^2$,

$$H(x, y) = C(F(x), G(y)). \quad (3.2)$$

Por outro lado, se C é uma cópula bivariada e F, G são funções de distribuição univariadas, a função H definida na equação (3.2) é uma função de distribuição bivariada com marginais F e G . Além disso, se as distribuições marginais forem contínuas, C é única. Caso contrário, C é dado exclusivamente em $\text{Dom}(F) \times \text{Dom}(G)$.

A demonstração deste teorema pode ser encontrada na página 19 de NELSEN (2006).

De acordo com KOLEV *et al.* (2006), a função cópula é uma das ferramentas mais úteis para lidar com distribuições multivariadas dadas ou conhecidas pelas marginais unidimensionais. Dada uma função de distribuição conjunta H com as marginais contínuas F e G , como no teorema Sklar, pode-se construir a cópula correspondente:

$$C(u, v) = H(F^{-1}(u), G^{-1}(v)), \quad (3.3)$$

em que F^{-1} e G^{-1} significa, respectivamente, a inversa generalizada de F e G , ou seja, $F^{-1}(u) = \sup_z \{F(z) \leq u\}$ e $G^{-1}(v) = \sup_z \{G(z) \leq v\}$.

Se X e Y forem variáveis aleatórias contínuas com a função de distribuição definida anteriormente, então C é uma função de distribuição conjunta com variáveis aleatórias $U = F(X)$ e $V = G(Y)$, que é obtida pela *probability integral transform*, dos quais U, V apresentam distribuição uniforme $U(0, 1)$.

A função de densidade da cópula c bivariada pode ser definida por:

$$c(F(x), G(y)) = \frac{\partial}{\partial x \partial y} C(F(x), G(y)), \quad (3.4)$$

se a derivada existir.

Do teorema de Sklar, a densidade conjunta de X e Y é dada por:

$$f(x, y) = c(F(x), G(y))f(x)g(y), \quad (3.5)$$

em que f e g são as funções densidade de probabilidade de F e G , respectivamente.

Portanto, qualquer função de distribuição conjunta que atenda aos requisitos do teorema tem uma representação de cópula. Existem vários exemplos de cópulas, como a cópula normal ou gaussiana, cópula t e as cópulas Arquimedianas. A maioria delas depende de um ou mais parâmetros, chamado δ , que caracteriza a dependência entre as variáveis (CHERUBINI *et al.*, 2004).

Por exemplo:

- **Cópula Gaussiana:** é a cópula de uma distribuição normal bivariada com parâmetro de correlação δ , dada por:

$$C_N(u, v; \delta) = \Phi_\delta(\Phi^{-1}(u), \Phi^{-1}(v)) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\delta^2}} e^{-\frac{x^2 - 2xy\delta + y^2}{2(1-\delta^2)}} dy dx, \quad (3.6)$$

em que Φ_δ é a distribuição normal bivariada com coeficiente de correlação δ .

- **Cópula t:** a cópula t bivariada é definida como:

$$C_{\nu, \Sigma}(u, v) = t_{\nu, \mathbf{0}, \Sigma}(t_\nu^{-1}(u), t_\nu^{-1}(v)) \quad (3.7)$$

em que $t_{\nu, \mathbf{0}, \Sigma}$ é a função de distribuição t-Student de média $\mathbf{0}$ e matriz de correlação Σ , e t_ν é função de distribuição univariada t-Student, com ν graus de liberdade. Temos, pois, que: $\delta = (\nu, \Sigma)$.

- **Cópulas Arquimedianas:** escritas na forma:

$$C(u, v) = \phi^{-1}[\phi(u) + \phi(v)] \quad (3.8)$$

para uma função $\phi : [0, 1] \rightarrow \mathbb{R}^+$, contínua, estritamente decrescente, de modo que $\phi(1) = 0$. A função $\phi(\cdot)$ é chamada função de geração ou função geradora da cópula $C(u, v)$.

As cópulas Arquimedianas são usadas principalmente em estudos financeiros, pois abrangem múltiplas estruturas de dependência (simétricas, assimétricas com dependência de caudas). As mais utilizadas são as cópulas de Gumbel, Clayton e Frank, definidas a seguir (PIRES, 2004):

1. **Gumbel:** $C(u, v) = \exp\{-(\ln u^\delta + (-\ln v)^\delta)^{1/\delta}\}$, em que $\delta \in [1, \infty)$ e a função geradora da cópula é $\phi_\delta(t) = (-\ln t)^\delta$.
2. **Clayton:** $C(u, v) = \max[(u^\delta + v^\delta - 1)^{1/\delta}, 0]$, em que $\delta \in [-1, 0) \cup (0, \infty)$ e $\phi_\delta(t) = (t^{-\delta} - 1)/\delta$.
3. **Frank:** $C(u, v) = \ln(1 + \frac{(\exp(-\delta u) - 1)(\exp(-\delta v) - 1)}{\exp(-\delta) - 1}(1/\delta))$, em que $\delta \in (-\infty, 0) \cup (0, \infty)$ e $\phi_\delta(t) = -\ln \frac{\exp(-\delta t) - 1}{\exp(-\delta) - 1}$.
4. **Ali-Mikhail-Haq(AMH)** (KUMAR, 2010): $C(u, v) = \frac{uv}{1 - \delta(1-u)(1-v)}$, em que $\delta \in [-1, 1]$ e $\phi_\delta(t) = \ln \frac{1 - \delta(1-t)}{t}$.

De acordo com MORETTIN (2008), estimadores paramétricos, não paramétricos e semi-paramétricos podem ser usados para estimar a cópula C . No primeiro caso, os estimadores de máxima verossimilhança são usados. No segundo caso, são utilizadas cópulas empíricas (com base em postos) e estimadores suavizados (via kernel e wavelets). No terceiro caso, são utilizados os estimadores de Pseudo máxima verossimilhança.

3.6.1 Estimadores Paramétricos da Função de Verossimilhança

Dado uma amostra $(X_i, Y_i), i = 1, \dots, n$ de H bivariado com distribuições marginais F e G , da Equação (3.5) a densidade da distribuição conjunta dada por:

$$f(x_i, y_i, \eta) = c(F(x_i, \alpha_1), G(y_i, \alpha_2))f(x_i, \alpha_1)g(y_i, \alpha_2), \quad (3.9)$$

em que α_1 contém os parâmetros de F , α_2 contém os parâmetros de G e θ contém parâmetros para c . Dado $\eta = (\alpha_1, \alpha_2, \theta)$ a log-verossimilhança é dada por:

$$l(x, y; \eta) = \sum_{i=1}^n \ln\{c(F(x_i, \alpha_1), G(y_i, \alpha_2); \theta)\} + \sum_{i=1}^n \ln\{f_1(x_i, \alpha_1)\} + \sum_{i=1}^n \ln\{f_2(y_i, \alpha_2)\} \quad (3.10)$$

Os estimadores de máxima verossimilhança (EMV) são obtidos ao maximizar esta função. Geralmente, esta pode ser uma tarefa difícil quando há muitos parâmetros, isto é, em casos multivariados $n > 5$ e quando as marginais são mais complexas. Então, a literatura recomenda um procedimento de dois estágios, denominado *inference function for margins (IFM)*, na qual a estimativa é dividida em duas partes.

- No passo 1, os estimadores dos parâmetros das marginais são obtidos,
 $\hat{\alpha}_i = \arg \max \sum_{i=1}^n \ln\{f_i(x_i, \alpha_i)\}, i = 1, 2$
- No passo 2, os estimadores da cópula são obtidos,
 $\hat{\theta} = \arg \max \sum_{i=1}^n \ln\{c(F(x_i, \hat{\alpha}_1), G(y_i, \hat{\alpha}_2); \theta)\}.$

De acordo com JOE and XU (1996), este procedimento leva a estimadores consistentes e assintoticamente normais.

3.6.2 Estimadores de Pseudo Máxima Verossimilhança - Método Semi-Paramétrico

Neste caso, F e G são estimados usando modelos não paramétricos, função de distribuição empírica (f.d.e.) ou uma combinação de f.d.e e uma distribuição de valores extremos para as caudas das distribuições. Para encontrar os estimadores de pseudo máxima verossimilhança, propostos por GENEST *et al.* (1995), seguem-se as seguintes etapas:

- Obter pseudo amostras para cópula $(\hat{u}_i, \hat{v}_i) = (\hat{F}(x_i), \hat{G}(y_i))$, em que $\hat{F}(x_i) = n^{-1}[\sum_{j=1}^n I(x_{ij} \leq x) - 0, 5]$ para $i = 1, \dots, n$.
- Encontre log-verossimilhança e maximize em relação a θ por métodos numéricos.

$$l(\theta, \hat{u}, \hat{v}) = \sum_{i=1}^n \ln\{c(\hat{u}_i, \hat{v}_i; \theta)\} \quad (3.11)$$

Este método também é conhecido como máxima verossimilhança canônica.

3.6.3 Testes de ajuste aos dados (*Goodness-of-fit tests*)

Os procedimentos *Goodness-of-fit* são utilizados para a seleção da cópula que melhor se ajusta aos dados. O procedimento de ajuste de dados descrito nesta seção baseia-se no trabalho de GENEST *et al.* (2009), BERG (2009) e KOJADINOVIC *et al.* (2011).

Dado um conjunto de variáveis aleatórias e independentes $X = (X_1, X_2, \dots, X_n)$, assume-se que as variáveis foram geradas de uma distribuição conjunta $H(x) = C[F_1(x_1), F_2(x_2), \dots, F_n(x_n)]$, $\forall x \in \mathbb{R}^n$, em que $C[\cdot]$ descreve uma cópula paramétrica desconhecida.

Por exemplo, está implícito que um dado conjunto de observações foi gerado por uma cópula gaussiana, descrita da seguinte forma:

$$H_0 : C \in C_\theta,$$

$$H_a : C \notin C_\theta,$$

em que C_θ representa a cópula gaussiana, cujo parâmetro é θ .

De acordo com JOE (2014), os procedimentos de ajuste de dados são baseados na transformação de Rosenblatt, estatísticas de Cramer-von Mises ou Kolmogorov-Smirnov e outras funções de densidade empírica, a fim de envolver uma medida da distância total entre o modelo suposto e a densidade empírica.

Entre os vários métodos de realização de testes relacionados a este tipo de hipótese, o teste de Cramer-von Mises se destaca, pois de acordo GENEST *et al.* (2009), BERG (2009) e KOJADINOVIC *et al.* (2011) é o mais robusto com mais evidências de eficácia.

Portanto, o processo empírico é dado por: $\mathbb{C}_n(u) = \sqrt{n}[C_n(u) - C_{\theta_n}(u)]$, $u \in [0, 1]^d$, em que C_n é a cópula empírica na amostra e C_{θ_n} é cópula estimada sob H_0 . Assim, a estatística de Cramer-von Mises é dada por:

$$S_n = \int_{[0,1]^d} \mathbb{C}_n(u)^2 dC_n(u) = \sum_{i=1}^n [C_n(U_i) - C_{\theta_n}(U_i)]^2.$$

Uma abordagem do p-valor para S_n pode ser obtida usando métodos paramétricos de bootstrap. No entanto, esta abordagem é computacionalmente pesada, pois para cada iteração requer a geração de números aleatórios da cópula e a estimativa dos parâmetros dessa cópula. Portanto, a medida que o tamanho da amostra aumenta, a aplicação do teste paramétrico baseado em bootstrap torna-se proibitiva. Uma abordagem baseada no teorema do limite central é proposta para reduzir o alto custo computacional.

3.6.4 Procedimento de Vuong para Comparação de Modelos Paramétricos de Cópulas

Para comparar dois modelos paramétrico não aninhados da cópula o procedimento de Vuong proposto por VUONG (1989) é descrito. Este procedimento usa o critério de informação Kullback-Leibler para medir a proximidade do modelo à verdade. E as estatísticas com base na razão de verossimilhança são usadas para testar a hipótese nula de que os modelos concorrentes também estão próximos do processo gerador de dados reais contra a hipótese alternativa de que o modelo está mais próximo.

Os testes são direcionais e são sucessivamente derivados para os casos em que os modelos concorrentes não são aninhados, sobrepostos ou aninhados e se ambos, um ou nenhum deles está incorretamente especificado. Como pré-requisito, a distribuição assintótica da estatística da razão de verossimilhança é completamente caracterizada em condições mais gerais. É uma soma ponderada de Qui-quadrado ou uma distribuição normal, dependendo se as distribuições em modelos competitivos mais próximos da verdade são observados de forma idêntica. Este teste também é proposto para esta última condição.

De acordo com JOE (2014), o procedimento de Vuong pode ser descrito como uma versão amostral para calcular a divergência de Kullback-Leibler e o tamanho de amostra para diferenciar dois modelos que podem ser não aninhados.

Considere uma amostra bivariada (u_i, v_i) de tamanho n . Seja $f_1 = c_1(u, v; \theta)$ e $f_2 = c_2(u, v; \theta)$ densidades para duas cópulas bivariadas $C_1(\Theta_1)$ e $C_2(\Theta_2)$, com os respectivos parâmetros Θ_1 e Θ_2 .

A diferença da Divergência de Kullback-Leibler de duas cópulas da densidade verdadeira da cópula pode ser medida como :

$$\hat{D}_{12} = n^{-1} \sum_i D_i = n^{-1}(RLV),$$

em que RLV = razão log-verossimilhança é dada por:

$$D_i = \log \left\{ \frac{f_2(y_i; \hat{\theta}_2)}{f_1(y_i; \hat{\theta}_1)} \right\}. \quad (3.12)$$

Para modelos não aninhados ou aninhados onde $f_1 = c_1(u, v; \theta)$ e $f_2 = c_2(u, v; \theta)$ não são densidades iguais, um intervalo de confiança 95% para o parâmetro $\hat{D}_{12}$ é:

$$\hat{D}_{12} \pm n^{-1/2} \hat{\sigma}_{12}^2, \quad (3.13)$$

em que a variância de D_{12} é dada por: $\hat{\sigma}_{12}^2 = (n-1)^{-1} \sum_{i=1}^n (D_i - \hat{D}_{12})^2$.

Existem também versões com RLV ajustadas pelos critérios de informação Akaike (AIC) ou com base no critério de informação de Bayes (BIC) e são respectivamente, VUONG (1989):

$$\hat{D}_{12} - n^{-1}[\dim(\Theta_2) - \dim(\Theta_1)] \pm 1, 96n^{-1/2} \hat{\sigma}_{12}^2, \quad (3.14)$$

$$\hat{D}_{12} - \frac{1}{2}n^{-1} \log(n)[\dim(\Theta_2) - \dim(\Theta_1)] \pm 1, 96n^{-1/2} \hat{\sigma}_{12}^2. \quad (3.15)$$

Se os intervalos descritos nas Equações (3.13), (3.14) ou (3.15) contiverem o zero, os modelos M_1 e M_2 não são considerados significativamente diferentes. Se o intervalo não contiver o zero, então, o modelo M_1 ou M_2 é o melhor ajuste, dependendo se o intervalo está completamente abaixo de 0 ou acima de 0, respectivamente.

3.7 Resultados

Para retirada da tendência ao longo do tempo nas séries de produtividade e preço foram utilizados o modelo determinístico linear, descrito na seção 2.5.7. A Tabela 3.1 apresenta as estimativas do modelo linear e a significância dos parâmetros para todas as séries. Considerando um nível de 1% de significância, observa-se a presença de tendência linear de ordem 1 para todas as séries de produtividade

(a) Produtividade de soja em Toledo

(b) Produtividade de soja em Guarapuava

(c) Produtividade de soja em Cascavel

(d) Produtividade de soja em Castro

(e) Preço da soja no Paraná

Figura 3.1. Séries Original e Corrigida da Produtividade e Preços.

Tabela 3.1. Resultados do ajuste do modelo linear para retirada da tendência.

Séries	Coefficientes	Estimativas	p-valor	R^2
Produtividade Toledo	α	-76362,08	$5,83 \times 10^{-9}$	0,6629
	β	39,53	$2,86 \times 10^{-9}$	
Produtividade Guarapuava	α	-100300	$2,68 \times 10^{-15}$	0,8549
	β	51,48	$1,33 \times 10^{-15}$	
Produtividade Cascavel	α	-79534,999	$4,79 \times 10^{-8}$	0,6033
	β	41,112	$2,51 \times 10^{-8}$	
Produtividade Castro	α	-81924,200	$4,41 \times 10^{-14}$	0,8306
	β	42,429	$1,74 \times 10^{-14}$	
Preço Deflacionado	α	-81763,886	$3,53 \times 10^{-13}$	0,8172
	β	42,348	$1,46 \times 10^{-13}$	
	γ	42,348	$1,46 \times 10^{-13}$	

Tabela 3.2. Resultados do teste Ljung-Box para as séries de produtividade e preço e estas ao quadrado, cuja hipótese nula é independência entre as observações ao longo do tempo e a homogeneidade das variâncias.

Série	p-valor para Serie	p-valor para Serie ²
Preço	0,0763	0,0842
Produtividade em Toledo	0,2619	0,1273
Produtividade em Guarapuava	0,0509	0,0437
Produtividade em Cascavel	0,3746	0,2975
Produtividade em Castro	0,0936	0,0866

e ordem 2 para a série de preços. A Figura 3.1 apresenta a série de produtividade original e corrigida dos quatro municípios e a série de preços.

Para verificação da dependência temporal e a heterogeneidade de variâncias, a Tabela 3.2 apresenta os p-valores do teste Ljung Box para as séries corrigidas e estas ao quadrado. Com um nível de significância de 1%, a série não possui dependência temporal e heterogeneidade de variâncias ao longo do tempo.

As estimativas dos modelos paramétricos da cópula foram calculadas usando o software R, versão 3.0.3, juntamente com o pacote cópula (YAN, 2007). Além disso, a função mvtnorm do software R foi utilizada para estimar a função Normal multivariada.

3.7.1 Inferência semi-paramétrica

Para estimar os parâmetros das cópulas sob a abordagem semi-paramétrica, foram utilizadas distribuições empíricas para as séries de produtividade para os municípios selecionados e preço e, em seguida, as cópulas paramétricas foram estimadas.

A Tabela 3.3 apresenta as estimativas para os parâmetros de dependência das cópulas e entre parênteses, o erro padrão das estimativas. Para todas as cópulas, o parâmetro de dependência δ é negativo, mostrando uma dependência negativa entre as variáveis preço e produtividade. Na prática, esta relação de dependência negativa é esperada, de acordo com a lei da oferta e demanda, pois quando a produtividade aumenta e o estoque de soja cresce, os preços gerais do produto se reduzem e vice-versa.

A Tabela 3.4 apresenta os Critérios de Informação AIC e BIC, o valor de log-verossimilhança (LLV), bem como as estatísticas de Cramer-von Mises (S_n) e o p-valor do teste de ajuste do modelo descrito na seção 3.6.3. Para todos os municípios e em todas as cópulas consideradas, com nível de significância de 0,01 os testes não rejeitaram nenhuma cópula. Portanto, a escolha da cópula que melhor representa a estrutura de dependência entre as variáveis foi escolhido utilizando pela análise dos menores valores dos critérios AIC e BIC. Observe que, a cópula t só obteve convergência para a produti-

Tabela 3.3. Estimativas para os parâmetros das cópulas (erro padrão entre parênteses).

Copulas	Toledo x Preço δ (erro padrão)	Guarapuava x Preço δ (erro padrão)	Cascavel x Preço δ (erro padrão)	Castro x Preço δ (erro padrão)
Clayton	-0,0889 (0,2006)	-	-0,0243(0,2254)	-0,0585(0,1718)
Frank	-1,0141 (1,0215)	-1,7589(1,0390)	-0,6834(1,0198)	-0,4189(0,9830)
AMH	-0,7185 (0,5505)	-	-0,4077(0,5709)	-0,2753(0,5408)
Gaussiana	-0,1569 (0,1685)	-0,3218(0,1685)	-0,0768(0,1863)	-0,0808(0,1755)
t_ν	-	-0,3087(0,1662)		
		$\nu = 10,8119(29,6171)$		

Tabela 3.4. Critérios de seleção AIC e BIC para cópulas paramétricas com marginais empíricas (ME).

	Critério	Clayton	Frank	AMH	Gaussiana	t
Toledo x Preço	AIC	1,7429	0,9995	0,6766	1,1871	
	BIC	3,2983	2,5549	2,232	2,7425	
	LLV	0,1285	0,5002	0,6616	0,4064	
	Sn	0,0164	0,03224	-	0,03338	
	p-valor	0,7188	0,5609	-	0,6978	-
Cascavel x Preço	AIC	1,9797	1,5366	1,4797	1,8071	
	BIC	3,535	3,0919	3,035	3,3625	
	LLV	0,1016	0,2317	0,2601	0,0964	
	Sn	0,0244	0,0246	0,0262	0,02238	
	p-valor	0,6188	0,8936	0,2672	0,9446	
Castro x Preço	AIC	1,9017	1,8356	1,7945	-1,7865	
	BIC	3,4571	3,3909	3,3498	3,3418	
	LLV	0,0491	0,0822	0,1027	0,1067	
	Sn	0,0490	0,0566	0,05981	0,0588	
	p-valor	0,2592	0,02048	0,0449	0,05145	
Guarapuava x Preço	AIC		-0,8457		-1,5938	0,267
	BIC		0,7097		-0,0384	3,377
	LLV		1,4228		1,7968	1,8665
	Sn		0,02713		0,0324	0,0414
	p-valor		0,8323		0,7904	0,6618

vidade Guarapuava $\times$ preço. Neste mesmo conjunto de dados as cópulas Clayton e AMH não obtiveram convergência.

Para a produtividade de Toledo $\times$ preço e produtividade de Cascavel $\times$ preço, a cópula que melhor se ajusta aos dados é a cópula AMH. Por outro lado, a produtividade para Castro $\times$ preço e Guarapuava $\times$ preço foram melhor ajustadas pela cópula gaussiana.

A próxima seção apresenta a estimativa dos parâmetros da cópula sob a abordagem paramétrica, neste caso, especificando uma distribuição paramétrica para preço e produtividade separadamente e, em seguida, ajustando a cópula às variáveis simuladas dessas distribuições.

3.7.2 Inferência Paramétrica

A Figura 3.2 apresenta os gráficos do preço ajustado e das densidades da produtividade em Toledo. As distribuições Normal, SN e OLLN foram utilizadas para a produtividade. A série de preços foi modelada pelas distribuições Log-Normal e ST.

A Tabela 3.5 apresenta os critérios utilizados para a escolha dos modelos das marginais produtividade em Toledo e preço. Para a produtividade, o modelo escolhido foi a distribuição OLLN, que apresentou o menor valor para as estatísticas A^* e W^* . Além disso, de acordo com o mesmo critério, o modelo escolhido para a série de preços foi a distribuição Skew-t. A mesma análise foi realizada para os municípios de Guarapuava, Cascavel e Castro, em que o modelo OLLN paramétrico melhor se ajusta aos

dados para todos os municípios.

Figura 3.2. Distribuições ajustadas para produtividade corrigida em Toledo e a série de preço SEAB.

Tabela 3.5. Estatísticas A^* e W^* e os critérios de informação AICc, BIC para as séries univariadas.

Variável	Modelo	AICc	BIC	A^*	W^*
Produtividade Toledo	Normal	554,4212	557,1569	0,7902	0,1208
	SN	556,8204	560,7120	0,7902	0,1208
	OLLN	555,9902	559,8864	0,6605	0,1063
Preço	Log Normal	281,5256	284,2630	0,8458	0,1518
	ST	280,2308	286,4522	0,8215	0,0987

O próximo passo é selecionar a cópula que melhor representa a estrutura de dependência dos dados. Para isso, foram utilizados os critérios de informação AIC e BIC. A Tabela 3.6 apresenta o Critério de Informação AIC, BIC, as estatísticas de Cramer-von Mises (Sn) e o p-valor do teste descrito na seção 3.6.3 para cópulas paramétricas com marginais paramétricas (MP).

De forma semelhante ao caso da inferência semi-paramétrica para todos os municípios e em todas as cópulas ajustadas, o teste ajustado não rejeitou nenhuma cópula, com um nível de significância de 0,01.

Portanto, para a escolha da cópula que melhor representa a dependência entre as variáveis, foi escolhido aquela que apresentou os menores valores para AIC e BIC. A cópula AMH caracteriza melhor a estrutura de dependência entre as variáveis de produtividade em Toledo, Cascavel e Castro com preço. A Cópula AMH apresenta maior precisão no cálculo das taxas do prêmio de faturamento para esses municípios. Para a produtividade em Guarapuava e o preço, a cópula gaussiana foi a selecionada.

3.7.3 Comparação entre modelos

A Tabela 3.7 apresenta os modelos escolhidos para ambos os procedimentos de inferência (com ME e MP). Para os municípios de Toledo, Cascavel e Guarapuava, em ambos os tipos de inferências, os

Tabela 3.6. Critérios de seleção AIC, BIC, estatística Sn para cópulas paramétricas com marginais paramétricas (MP)

	Critério	Clayton	Frank	AMH	Gaussiana	t
Toledo x Preço	AIC	1,4722	0,1345	-0,2317	0,6865	
	BIC	3,0276	1,6898	1,3236	2,2419	
	Sn	0,0164	0,03224		0,03338	
	p-valor	0,7228	0,5929		0,7058	
Cascavel x Preço	AIC	1,8011	0,8069	0,7161	1,3959	
	BIC	3,3564	2,3623	2,2715	2,9513	
	Sn	0,0244	0,0246	0,0262	0,02238	
	p-valor	0,6386	0,8886	0,2393	0,9466	
Castro x Preço	AIC	1,9338	1,8133	1,78	1,7859	
	BIC	3,4892	3,3686	3,3353	3,3412	
	Sn	0,0490	0,0566	0,05981	0,0588	
	p-valor	0,2552	0,0244	0,0148	0,0624	
Guarapuava x Preço	AIC		-1,1173		-1,5327	0,3152
	BIC		0,4338		-0,0227	3,4259
	Sn		0,02713		0,0324	0,0414
	p-valor		0,8323		0,7887	0,6317

Tabela 3.7. Modelos selecionados por inferência semi-paramétrica (ISP) e paramétrica (IP), com parâmetro de dependência (δ).

	Cópula	δ
Toledo x Preço	AMH (ISP)	-1
	AMH (IP)	-0,92021
Cascavel x Preço	AMH (ISP)	-0,4811
	AMH (IP)	-0,6452
Castro x Preço	Gaussiana (ISP)	-0,0695
	AMH (IP)	-0,4260
Guarapuava x Preço	Gaussiana (ISP)	-0,3218
	Gaussiana (IP)	-0,3058

modelos escolhidos foram os mesmos, a cópula AMH, AMH e a gaussiana, respectivamente.

Nestes casos, como não é possível aplicar o procedimento proposto por VUONG (1989), os modelos com os menores valores de AIC e BIC foram escolhidos, de acordo com as Tabelas 3.4 e 3.6. Portanto, para os municípios de Cascavel e Toledo, a cópula AMH foi escolhida pela inferência paramétrica (com MP). Para o município de Guarapuava, o modelo de cópula gaussiana foi selecionado sob enfoque da inferência semi-paramétrica (com ME).

Além disso, a tabela 3.7 mostra que, para o município de Castro, existem modelos diferentes em ambos os procedimentos de inferência (ISP e IP). Nesse caso, o procedimento de Vuong foi utilizado. O intervalo de confiança de 95% para $\hat{D}_{12}$, de acordo com a Equação (3.13) é (-0,5352385; 0,8787669) com $\hat{D}_{12} = 0,1717642$. Como o intervalo contém zero, os modelos não podem ser considerados significativamente diferentes.

Por outro lado, aplicando o teste de ajuste aos dados, descrito na seção 3.6.3, observa-se na Tabela 3.8 que o menor valor da estatística de Cramer Von Mises (Sn) é para o modelo ajustado pela inferência semi-paramétrica (ISP). Além disso, com nível de significância de 0,05, pode-se considerar a não rejeição da hipótese nula no caso de inferência semi-paramétrica. Portanto, o modelo de cópula gaussiana se adapta melhor aos dados.

A Tabela 3.9 resume os resultados dos modelos selecionados para os diferentes municípios.

Tabela 3.8. Resultados do teste de ajuste ao modelo para o município de Castro x Preço por ISP e IP

Inferência	Sn	δ	p-valor
Semi-paramétrica (ME)	0,058886	-0,0767	0,05445
Paramétrica (MP)	0,059819	-0,24817	0,03247

Tabela 3.9. Resumo dos modelos selecionados para os diferentes municípios por SPI e PI.

Município x Preço	Modelo de Cópula
Castro	Gaussiana (ISP)
Guarapuava	Gaussiana (ISP)
Toledo	AMH (IP)
Cascavel	AMH (IP)

O próximo passo é calcular as taxas do prêmio do seguro de faturamento utilizando os modelos selecionados e comparar os resultados com a distribuição normal bivariada, geralmente utilizada pelo mercado segurador.

3.8 Taxa do Seguro de Faturamento

No início do contrato de seguro, o produtor rural escolhe o NC. Supondo que o produtor pode escolher o NC entre 60 a 80 % o faturamento garantido $F_g = X_g Y_g$, se o faturamento obtido for inferior ao faturamento garantido no contrato, o segurado recebe uma indenização.

A Tabela 3.10 apresenta taxas puras (TP) calculadas de acordo com a Equação (3.1), utilizando as cópulas escolhidas e a distribuição normal bivariada para todos os municípios.

Para os municípios de Castro e Cascavel, a taxa do seguro de faturamento calculada pela distribuição normal bivariada subestima as taxas quando comparadas às calculadas pelas cópulas, para todos os NC.

Além disso, para Guarapuava, a taxa calculada pela normal bivariada também é subestimada quando comparada à cópula para NC de até 75%. Esta subestimação da taxa do prêmio pode levar a graves perdas para as seguradoras, por considerar um risco menor do que deveria ser levado em consideração.

Para o município de Toledo, a taxa do seguro de faturamento calculada pela distribuição normal bivariada é subestimada para os NC igual ou inferior a 65% e superestima para NC superior à 65% quando comparada com a cópula AMH.

Para os municípios de Guarapuava e Toledo com um NC de 80%, a distribuição Normal Bivariada superestima a taxa do prêmio do seguro em relação à cópula. A superestimação da taxa pode dificultar o uso generalizado do seguro de faturamento no território brasileiro, além de atrair produtores com perfil de maior risco, aumentando o problema de seleção adversa.

De acordo com BRISOLARA (2013), a taxa pura representa o risco intrínseco do negócio, sem incluir qualquer carga (ou custos adicionais). Para compará-la com as taxas oferecidas pelas seguradoras brasileiras, é necessário carregá-la com parâmetros médios de mercado relativos à margem técnica (20%), despesas administrativas (20%), despesas comerciais (10%) e margem de lucro da seguradora (10%).

Portanto, a taxa comercial (TC) é calculada da seguinte forma: $TC = \frac{TP * 1,2}{[1 - (0,1 + 0,2 + 0,1)]}$.

As taxas médias comerciais do prêmio do seguro de faturamento calculadas usando a cópula, normal bivariada e as taxas das companhias de seguros para NC de 60 – 69% e 70 – 79% são apresentados na Tabela 3.11.

As taxas calculadas neste estudo são superiores àquelas aplicadas no mercado de seguros. Por exemplo, considerando o município de Toledo com NC 60 – 69% e 70 – 79%, a taxa da companhia de seguros A é equivalente a 50,64% e 29,84% da taxa calculada pela cópula AMH, respectivamente.

Tabela 3.10. Taxa Pura para o prêmio do seguro de faturamento calculada pela cópula selecionada e Normal Bivariada para todos os municípios.

Toledo $\times$ preço			Cascavel $\times$ preço	
NC	Cópulas AMH (MP)	Normal Bivariada	Cópula AMH (MP)	Normal Bivariada
60%	3,43	2,03	6,10	0,48
65%	4,36	4,03	7,73	1,52
70%	5,41	7,16	9,54	3,26
75%	6,56	11,21	11,54	5,89
80%	7,83	15,71	13,75	10,11

Guarapuava $\times$ preço			Castro $\times$ preço	
NC	Cópula Gaussiana (ME)	Normal Bivariada	Cópula Gaussiana (ME)	Normal Bivariada
60%	2,52	0,020	4,06	$6,3 \times (10)^{-6}$
65%	3,32	0,135	5,10	$1,2 \times (10)^{-4}$
70%	4,23	0,710	6,22	0,0048
75%	5,25	2,713	7,45	0,1070
80%	6,40	8,589	8,78	1,0915

Tabela 3.11. Taxas Comerciais Médias (%) do Prêmio de Faturamento usando a cópula, Normal Bivariada e a seguradora A.

Municípios	Modelo	Nível de Cobertura	
		60 – 69%	70 – 79%
Toledo $\times$ preço	Cópula AMH (MP)	12,351	19,032
	Normal Bivariada	6,164	19,747
	Seguradora A	6,26	5,68
Cascavel $\times$ preço	Cópula AMH (MP)	13,838	21,088
	Normal Bivariada	2,004	9,157
	Seguradora A		5,52
Guarapuava $\times$ preço	Cópula AMH (ME)	5,772	9,491
	Normal Bivariada	0,039	1,621
	Seguradora A		4,85
Castro $\times$ preço	Cópula Gaussiana (ME)	9,166	13,684
	Normal Bivariada	$1,268 \times (10)^{-4}$	0,1117
	Seguradora A	5,20	5,24

No caso do município de Castro, a diferença é ainda maior. As taxas comerciais médias da seguradora A são equivalentes a 56,73% e 38,29% da taxa calculada pela cópula Gaussiana, respectivamente.

Portanto, os resultados sugerem que há uma subestimação da taxa do prêmio por parte da seguradora, o que pode resultar em grandes perdas, pois considera-se um risco menor do que deveria ser levado em consideração.

Por outro lado, de acordo com MAPA (2017), 93,30% das apólices de seguros de faturamento para a cultura de soja no Brasil são vendidos pela companhia de seguros A, que oferece quase exclusivamente este tipo de seguro e cobre todas as regiões do Brasil (espalhando os riscos). Isso permite que a empresa diminua a taxa de seguro em regiões com maior risco, como é o caso da região sul do Brasil.

Essa diferença entre as taxas aplicadas no mercado e aquelas calculadas neste estudo pode ser explicada pelo fato de que esse tipo de seguro é relativamente recente no mercado brasileiro. Isso significa que há uma falta de estudos sobre quais procedimentos atuariais devem ser utilizados para precificar esse tipo de risco, o que pode ter levado a seguradora a adotar um procedimento subjetivo nas taxas dos preços.

3.9 Conclusões

Neste trabalho, foram propostas abordagens alternativas para calcular as taxas do prêmio de seguro de faturamento utilizando cópulas paramétricas com distribuições marginais paramétricas e marginais empíricas. Estes métodos foram aplicados aos dados de produtividade de soja dos municípios de Toledo, Cascavel, Guarapuava e Castro do Estado do Paraná e os preços mensais nominais recebidos pelos produtores do Estado do Paraná. Concluiu-se que os modelos paramétricos que melhor se ajustam às séries de produtividade para todos os municípios e preços foram os modelos OLLN e Skew-t, respectivamente.

Além disso, a cópula AMH melhor representa a estrutura de dependência entre as variáveis para os municípios de Cascavel e Toledo. Para o município de Guarapuava e Castro, a cópula gaussiana foi a selecionada com marginais empíricas.

Para os municípios de Castro e Cascavel, considerando todos os níveis de cobertura, a taxa do seguro de faturamento calculada pela distribuição Normal Bivariada é subestimada quando comparadas às cópulas. Para o município de Guarapuava, a taxa da distribuição Normal Bivariada também é subestimada quando comparada às cópulas para níveis de cobertura de até 75%. Para o município de Toledo, a taxa do seguro de faturamento calculada pela distribuição normal subestima até 65% do NC e superestima 70% de CL quando comparado com a cópula AMH.

A subestimação da taxa do prêmio pode levar a graves perdas às seguradoras, uma vez que considera um risco menor do que deveria ser levado em consideração. A superestimação da taxa, por sua vez, pode dificultar o uso generalizado deste tipo de seguro no território brasileiro, atraindo produtores com perfil de maior risco, o que aumenta o problema de seleção adversa.

Além disso, as taxas comerciais médias calculadas neste estudo foram muito superiores às aquelas aplicadas pela companhia de seguros A. Esta lacuna pode ser explicada por causa da superestimação da taxa de carregamento, que é uma informação confidencial. Em outras palavras, ao considerar a margem técnica, as despesas comerciais e administrativas e a margem de lucro da seguradora utilizada neste trabalho, podemos ter superavaliado a taxa de carregamento, levando a uma taxa mais elevada.

Além disso, este tipo de seguro é relativamente recente no Brasil e a escassez de estudos atuariais e procedimentos definidos para sua precificação podem ter levado a seguradora a adotar um procedimento subjetivo para a precificação de suas taxas.

O próximo capítulo apresenta um estudo mais detalhado da estrutura de dependência tridimensional, entre a produtividade, os preços de contratos futuros negociados na "Chicago Mercantile Exchange Group" (CME) e as taxas de câmbio, visto que grande parte da safra de soja brasileira é exportada e comercializada na CME. Além disso, compara-se as taxas do prêmio do seguro de faturamento calculado sob o enfoque bidimensional (produtividade e preços em reais) e sob o enfoque tridimensional (produtividade, preço em dólar e taxa do dólar) e, estas com as taxas das seguradoras.

Referências

- AHMED, O. and T. SERRA, 2015 Economic analysis of the introduction of agricultural revenue insurance contracts in Spain using statistical copulas. *Agricultural Economics* **46**: 69–79.
- AZZALINI, A., 1985 A class of distributions which includes the normal ones. *Scandinavian journal of statistics* pp. 171–178.
- AZZALINI, A. and A. CAPITANIO, 2003 Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65**: 367–389.

- BABCOCK, B. A. and D. A. HENNESSY, 1996 Input demand under yield and revenue insurance. *American Journal of Agricultural Economics* **78**: 416–427.
- BERG, D., 2009 Copula goodness-of-fit testing: an overview and power comparison. *The European Journal of Finance* **15**: 675–701.
- BOTTS, R. R. and J. N. BOLES, 1958 Use of normal-curve theory in crop insurance ratemaking. *Journal of Farm Economics* **40**: 733–740.
- BOUYÉ, E., V. DURRLEMAN, A. NIKEGBALI, G. RIBOULET, and T. RONCALLI, 2000 Copulas for finance-a reading guide and some applications. Available at SSRN 1032533 .
- BRAGA, A. S., G. M. CORDEIRO, E. M. ORTEGA, and J. N. DA CRUZ, 2016 The odd log-logistic normal distribution: Theory and applications in analysis of experiments. *Journal of Statistical Theory and Practice* **10**: 311–335.
- BRISOLARA, C. S., 2013 *Proposições para o desenvolvimento do seguro de receita agrícola no Brasil: do modelo teórico ao cálculo das taxas de prêmio*. Ph.D. thesis, Universidade de São Paulo.
- CARVALHO, A. L., C. A. PAREDES, D. MIQUELLUTI, D. RUIS, E. PASSARELLI, and G. V. DUARTE, 2013 Aspectos gerais do seguro de faturamento. Acessado em: 23 de junho de 2015.
- CHERUBINI, U., E. LUCIANO, and W. VECCHIATO, 2004 *Copula methods in finance*. John Wiley & Sons.
- COBLE, K. H., T. O. KNIGHT, R. D. POPE, and J. R. WILLIAMS, 1996 Modeling farm-level crop insurance demand with panel data. *American Journal of Agricultural Economics* **78**: 439–447.
- DAY, R. H., 1965 Probability distributions of field crop yields. *Journal of Farm Economics* **47**: 713–741.
- DUARTE, G. V., A. BRAGA, D. L. MIQUELLUTI, and V. A. OZAKI, 2017 Modeling of soybean yield using symmetric, asymmetric and bimodal distributions: implications for crop insurance. *Journal of Applied Statistics* .
- EMBRECHTS, P., F. LINDSKOG, and A. MCNEIL, 2003 Modelling dependence with copulas and applications to risk management. *Handbook of heavy tailed distributions in finance* **8**: 329–384.
- GALLAGHER, P., 1987 Us soybean yields: estimation and forecasting with nonsymmetric disturbances. *American Journal of Agricultural Economics* **69**: 796–803.
- GENEST, C., K. GHOUDI, and L.-P. RIVEST, 1995 A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* **82**: 543–552.
- GENEST, C., B. RÉMILLARD, and D. BEAUDOIN, 2009 Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and economics* **44**: 199–213.
- GOODWIN, B. and H. TEJEDA, 2008 Modeling crop prices through a burr distribution and analysis of correlation between crop prices and yields using a copula method. In *American Agricultural Economics Association Annual Meeting. Orland-Flórida*.
- GOODWIN, B. K. and A. P. KER, 1998 Nonparametric estimation of crop yield distributions: implications for rating group-risk crop insurance contracts. *American Journal of Agricultural Economics* **80**: 139–153.
- GOODWIN, B. K. and A. P. KER, 2002 Modeling price and yield risk. In *A comprehensive assessment of the role of risk in US agriculture*, pp. 289–323, Springer.

- GOODWIN, B. K. and O. MAHUL, 2004 Risk modeling concepts relating to the design and rating of agricultural insurance contracts. Acessado em : 25 junho de 2014. Disponível em <http://elibrary.worldbank.org/doi/book/10.1596/1813-9450-3392>.
- GOODWIN, B. K., M. C. ROBERTS, and K. H. COBLE, 2000 Measurement of price risk in revenue insurance: implications of distributional assumptions. *Journal of Agricultural and Resource Economics* pp. 195–214.
- HARDAKER, J. B., R. B. HUIRNE, J. R. ANDERSON, G. LIEN, *et al.*, 2004 *Coping with risk in agriculture..* Number Ed. 2, CABI publishing.
- HENNESSY, D. A., B. A. BABCOCK, and D. J. HAYES, 1997 Budgetary and producer welfare effects of revenue insurance. *American Journal of Agricultural Economics* **79**: 1024–1034.
- HERNAN TEJEDA, A., B. K. GOODWIN, *et al.*, 2008 Modeling crop prices through a burr distribution and analysis of correlation between crop prices and yields using a copula method. In *annual meeting of the Agricultural and Applied Economics Association, Orlando, FL, USA*, volume 83, pp. 643–649, Citeseer.
- IPARDES, 2015 Instituto paranaense de desenvolvimento econômico e social. base de dados do estado (bdeweb). Acessado em: 02/julho/2015. Disponível em: <http://www.ipardes.pr.gov.br/imp/index.php>.
- IPEADATA, 2015 Índice geral de preços, disponibilidade interna (igp-di), fundação getulio vargas, conjuntura econômica. Acessado em 19/11/2015. Disponível em <http://www.ipeadata.gov.br/>.
- JOE, H., 2014 *Dependence modeling with copulas*. CRC Press.
- JOE, H. and J. J. XU, 1996 The estimation method of inference functions for margins for multivariate models. Technical report, Technical report.
- JUST, R. E. and Q. WENINGER, 1999 Are crop yields normally distributed? *American Journal of Agricultural Economics* **81**: 287–304.
- KER, A. P. and K. COBLE, 2003 Modeling conditional yield densities. *American Journal of Agricultural Economics* **85**: 291–304.
- KER, A. P. and B. K. GOODWIN, 2000 Nonparametric estimation of crop insurance rates revisited. *American Journal of Agricultural Economics* **82**: 463–478.
- KOJADINOVIC, I., J. YAN, and M. HOLMES, 2011 Fast large-sample goodness-of-fit tests for copulas. *Statistica Sinica* **21**: 841.
- KOLEV, N., U. DOS ANJOS, and B. V. DE M MENDES, 2006 Copulas: a review and recent developments. *Stochastic Models* **22**: 617–660.
- KUMAR, P., 2010 Probability distributions and estimation of ali-mikhail-haq copula. *Applied Mathematical Sciences* **4**: 657–666.
- LJUNG, G. M. and G. E. BOX, 1978 On a measure of lack of fit in time series models. *Biometrika* **65**: 297–303.
- MAPA, 2017 Atlas do seguro rural. indicador das taxas. Acessado em 21 de fevereiro de 2017. Disponível em <http://indicadores.agricultura.gov.br/atlasdoseguro/index.htm>.

- MIQUELETO, G. J., 2011 *Contribuições para o desenvolvimento do seguro agrícola de renda para o Brasil: evidências teóricas e empíricas..* Ph.D. thesis, Escola Superior de Agricultura "Luiz de Queiroz -Universidade de São Paulo.
- MORETTIN, P. A., 2008 *Econometria financeira: um curso em séries temporais financeiras*. Edgard Blücher.
- MOSS, C. B. and J. S. SHONKWILER, 1993 Estimating yield distributions with a stochastic trend and nonnormal errors. *American Journal of Agricultural Economics* **75**: 1056–1062.
- NELSEN, R. B., 2006 *An Introduction to Copulas*. Springer Series in Statistics.
- NELSON, C. H. and P. V. PRECKEL, 1989 The conditional beta distribution as a stochastic production function. *American Journal of Agricultural Economics* **71**: 370–378.
- OZAKI, V. A., B. K. GOODWIN, and R. SHIROTA, 2008 Parametric and nonparametric statistical modelling of crop yield: implications for pricing crop insurance contracts. *Applied Economics* **40**: 1151–1164.
- OZAKI, V. A. and R. S. SILVA, 2009 Bayesian ratemaking procedure of crop insurance contracts with skewed distribution. *Journal of Applied Statistics* **36**: 443–452.
- PIRES, R., 2004 *Modelando a Dependência nas Caudas das distribuições: O uso de cópulas em finanças..* Master's thesis, Faculdade de Economia, Administração e Contabilidade- USP.
- RAMIREZ, O. A., S. MISRA, and J. FIELD, 2003 Crop-yield distributions revisited. *American Journal of Agricultural Economics* **85**: 108–120.
- SANTOS, C. O., 2011 A distribuição skew-normal como modelo para a produtividade de milho aplicada ao seguro agrícola. Dissertação de Mestrado-Universidade Federal de Lavras .
- SCHÖNBUCHER, P. and D. SCHUBERT, 2001 Copula-dependent defaults in intensity models. Available at SSRN 301968 .
- SEAB, 2015 Secretaria da agricultura e do abastecimento do paran. departamento de economia rural (deral). Acessado em 30 de setembro de 2015. Disponvel em: <http://www.agricultura.pr.gov.br/modules/conteudo/conteudo.php?conteudo=195>.
- SHERRICK, B. J., F. C. ZANINI, G. D. SCHNITKEY, and S. H. IRWIN, 2004 Crop insurance valuation under alternative yield distributions. *American Journal of Agricultural Economics* **86**: 406–419.
- SKLAR, M., 1959 *Fonctions de rpartition n dimensions et leurs marges*. Universit Paris 8.
- TAYLOR, C. R., 1990 Two practical procedures for estimating multivariate nonnormal probability density functions. *American Journal of Agricultural Economics* **72**: 210–217.
- TURVEY, C. and J. ZHAO, 1999 *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.
- VEDENOV, D., 2008 Application of copulas to estimation of joint crop yield distributions. In *Proceedings of the American Agricultural Economics Association Annual Meeting, Orlando, FL*.
- VUONG, Q. H., 1989 Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society* pp. 307–333.
- YAN, J., 2007 Enjoy the joy of copulas: with a package copula. *Journal of Statistical Software* **21**: 1–21.

4 MODELAGEM DA ESTRUTURA DE DEPENDÊNCIA ENTRE SÉRIES TEMPORAIS MULTIVARIADAS: APLICAÇÃO A PRECIFICAÇÃO DOS RISCOS DO SEGURO DE FATURAMENTO AGRÍCOLA

Resumo

Grande parte da soja produzida no Brasil é exportada e, conseqüentemente, o preço da soja local (R\$) é muito influenciado pelo preço negociado na "Chicago Mercantile Exchange Group" (CME Group) (US\$). Por isso, para modelar a estrutura de dependência entre produtividade de soja e preço, a taxa do câmbio deve ser incorporada na modelagem. Portanto, esse trabalho tem por objetivo modelar a estrutura de dependência entre estas três variáveis utilizando a metodologia de cópulas e calcular as taxas do prêmio de seguro de faturamento, além de, comparar com as taxas ofertadas no mercado segurador. Observou-se que as taxas aplicadas pelo mercado segurador brasileiro estão superfaturadas quando comparadas pela metodologia apresentada neste trabalho com a incorporação da taxa do dólar na modelagem, o que pode aumentar o problema da seleção adversa e dificultar a massificação do seguro no território brasileiro.

Palavras Chaves: Cópulas Paramétricas Tridimensionais; Gerenciamento de Riscos Agrícolas; Seguro de Faturamento.

Abstract

Most of the soybean produced in Brazil is exported; thus, soybean prices (R\$) are greatly influenced by futures prices at the Chicago Mercantile Exchange Group (CME Group) (US\$). Therefore, to model the dependence structure between soybean yield and price, the exchange rate must be incorporated into the modeling. This work aims to model the dependence structure between these three variables using the methodology of copulas to calculate the insurance premium rate and compare the rates that are offered on the insurance market. The rates applied by the Brazilian insurance market are overpriced when compared by the methodology presented in this work with the incorporation of the dollar rate into the modeling, which can increase the problem of adverse selection and hinder the widespread use of agriculture insurance in the Brazilian territory.

Keywords: Three-dimensional parametric copulas; Agricultural risk management; Revenue insurance.

4.1 Introdução

No capítulo anterior, analisou-se a estrutura de dependência entre a produtividade e os preços médios nominais mensais (em reais) recebidos pelos produtores de soja no estado do Paraná. Esse preço (preço à vista) é coletado junto aos núcleos regionais e ponderado pela relação entre a produção do núcleo e a produção do estado e, posteriormente, é utilizada a média aritmética simples para calcular o preço médio mensal.

O objetivo de um negócio no mercado à vista (ou físico) é efetuar uma compra e/ou venda imediata. Assim, a entrega do produto e seu pagamento acontecem no mesmo instante. Esse tipo de mercado é tipicamente esporádico e apresenta alto grau de incertezas no que se refere ao comportamento dos preços, regularidade de suprimentos e qualidade de produtos. Portanto, uma empresa que adota esse mecanismo assume um grande risco, podendo levá-la ao insucesso nas suas operações, por consequência das incertezas dos aspectos relacionados à demanda e à oferta do produto.

Uma forma de reduzir os riscos do mercado físico é a comercialização no mercado futuro, conhecida como *hedge*, que é uma operação de "travamento" do preço para uma data futura. Outra característica do mercado futuro é que uma minoria dos contratos resulta em entrega física da mercadoria. Na grande parte dos contratos, ocorre a liquidação financeira antes da data da entrega, não necessitando

a entrega física no local acordado. Por exemplo, a soja produzida no Brasil normalmente é comercializada na CBOT/CME, mas a entrega física da *commodity* é na China ou União Européia.

As cooperativas ou *tradings* que comercializam as sojas brasileiras utilizam o mercado futuro, como forma de fixar o preço da *commodity* recebida no futuro. Assim, como grande parte da soja produzida no Paraná é exportada, a variável dólar tem grande influência no valor do preço recebido pelos produtores. Por isso, é importante a incorporação da variável "câmbio" no estudo da dependência entre as variáveis estudadas no seguro de faturamento agrícola.

No Brasil existem duas seguradoras que ofertam o seguro de faturamento aos produtores rurais. A companhia de seguros A utiliza em sua precificação a série de produtividade anual para cada município e a série de preços físicos diários (indicador CEPEA/ESALQ). Este indicador representa o valor médio de comercialização da soja brasileira a granel, tipo exportação, referenciado em Paranaguá-PR. Uma abordagem similar a empregada por esta seguradora é apresentada no capítulo anterior com o uso da série de preços físicos da SEAB.

Por outro lado, a seguradora B difere da anterior por utilizar a série de preços futuros da CME transformada em reais. Neste caso, não se leva em consideração a correlação existente entre preço futuro e a taxa do dólar, podendo levar a diminuição do impacto da variação dos preços. Esta abordagem é similar à apresentada neste capítulo com o uso da metodologia de cópulas, sob enfoque bidimensional.

No entanto, com o intuito de compreender a variação e dependência da taxa do câmbio na modelagem do faturamento, optou-se por incorporar essa variável ao modelo. Portanto, neste trabalho apresenta-se uma terceira abordagem com a incorporação da taxa do câmbio do dólar ao modelo de cópulas (enfoque tridimensional). O estudo do relacionando destas três variáveis não foi encontrado na literatura, no contexto de risco agrícola.

Portanto, o objetivo deste trabalho é modelar a estrutura de dependência entre produtividade de soja, preços CME-Group e cotação do dólar por meio de funções cópulas. Esta análise foi realizada sob dois enfoques: o enfoque bidimensional e o tridimensional. Além disso, em ambas as abordagens é realizado o cálculo e estimação das taxas do prêmio de seguro de faturamento e sua comparação com as taxas aplicadas pelo mercado segurador brasileiro.

Na próxima seção, é apresentada a forma do cálculo do prêmio do seguro de faturamento, sob o enfoque tridimensional. Na seção 4.3, apresenta-se a descrição dos dados utilizados na pesquisa. Na seção 4.4, apresenta-se a metodologia de cópulas para o cálculo da distribuição tridimensional e os modelos utilizados na modelagem das distribuições marginais. Na seção 4.5, os resultados da pesquisa são apresentados. Na seção 4.6 é realizada a discussão dos resultados, sob ambos os enfoques. Por fim, as conclusões são apresentadas na seção 4.7.

4.2 Taxa do Prêmio do Seguro de Faturamento

Neste trabalho considera-se o faturamento F como uma função de três variáveis, produtividade de soja (X), preço em dólares (Y) e cotação do dólar (Z), cuja expressão é $F = XYZ$. A produtividade garantida é definida por $X_g = \lambda X_e$, em que $0 < \lambda < 1$ é o nível de cobertura (NC) escolhido pelo produtor e X_e é a produtividade esperada, normalmente calculada pela média das últimas cinco safras. O preço garantido é definido por $Y_g = \lambda Y_e$, em que $0 < \lambda < 1$ é o NC e Y_e é o preço esperado, normalmente calculado pela média dos últimos 15 dias do preço futuro negociado na CME do dia da contratação do seguro. De forma análoga, definimos o dólar garantido ($Z_g = \lambda Z_e$).

Neste tipo de seguro, supomos que a indenização paga por unidade de área ao produtor rural seja dada por: $I = \max\{(x_g y_g z_g - xyz); 0\}$.

A probabilidade de perda é dada por $p = \text{Prob}(X < x; Y < y; Z < z | xyz < x_g y_g z_g)$. Essa probabilidade condicional expressa a probabilidade cumulativa conjunta de um evento (perda) em função

das variáveis aleatórias X , Y e Z , dado que a escolha dos valores de xyz está restrita à condição $xyz < x_g y_g z_g$. Assim, a taxa do prêmio é dada pela perda esperada, de acordo com a equação:

$$\begin{aligned} E(Perda) &= E(\max\{(x_g y_g z_g - xyz); 0\}) = \\ &= Prob(X < x; Y < y; Z < z | xyz < x_g y_g z_g) [x_g y_g z_g - E(XYZ | xyz < x_g y_g z_g)]. \end{aligned}$$

Portanto, a taxa ótima do prêmio é dada por:

$$\pi = \frac{Prob(X < x; Y < y; Z < z | xyz < x_g y_g z_g) [x_g y_g z_g - E(XYZ | xyz < x_g y_g z_g)]}{x_g y_g z_g}. \quad (4.1)$$

Observe que o valor esperado do faturamento, está relacionado com os riscos nas distribuições marginais de preço, produtividade e câmbio. A distribuição marginal de produtividade e preço são inferidas a partir dos dados históricos da produtividade e de preços utilizando métodos empíricos e paramétricos. Por sua vez a construção da distribuição conjunta utiliza-se uma função cópula paramétrica. Os conceitos teóricos relacionados à determinação da taxa de prêmio do seguro de faturamento bidimensional foram descritos no trabalho de MIQUELETO (2011) e BRISOLARA (2013).

4.3 Descrição dos dados

Para a modelagem da produtividade, foram analisadas as séries de produtividades anuais de soja em grãos (kg/ha) dos municípios de Cascavel e Toledo no estado do Paraná no período de 1980 à 2016, totalizando 36 observações, e disponibilizadas pelo IPARDES (IPARDES, 2017).

Para a modelagem do preço, utilizou-se a série de preços diários do contrato futuro de soja negociados na Chicago Board of Trade (CBOT) da CME-Group, com vencimento em março de 2017 (código ZSH17). Estes dados estão disponíveis em BARCHART (2017).

Os preços dos contratos futuros, Y_{CME} , são disponibilizados em cents/bushes, por isso, utilizou-se o fator de transformação $FT = (\frac{Y_{CME}}{27,216} * 60)$ para transformá-los em dólares por sacas de 60 Kg. O período de observação foi de 05/04/2015 à 14/03/2017, totalizando 455 observações.

A série diária de cotação do dólar comercial para o mesmo período acima foi retirada da Base de Dados Macro Econômicos e Regionais do Instituto de Pesquisa Econômica Aplicada IPEA (2017), no dia 05/02/2017.

Para a obtenção da mesma periodicidade das séries de preços e dólar utilizou-se a produtividade simulada da distribuição OLLN com 455 observações. Assim, é possível utilizar a metodologia de cópulas para as três variáveis com a mesma periodicidade (455 observações cada). Além disso, uma segunda análise foi realizada simulando 20.000 observações das três séries provenientes da cópula ajustada.

Por fim, as taxas comerciais do seguro de faturamento aplicado pela seguradora A estão disponíveis em MAPA (2017).

4.4 Metodologia

As etapas metodológicas deste trabalho consistem em três passos descritos a seguir:

- **Passo 1)** Modelagem da Produtividade.

A produtividade agrícola ao longo dos anos apresenta tendência crescente, em virtude do grande avanço nas tecnologias empregadas nas lavouras, como as técnicas de plantio, o uso de insumos, maquinários, entre outros. Por isso, utiliza-se o procedimento determinístico sugerido por GALLAGHER (1987), GOODWIN and MAHUL (2004) e TEJEDA *et al.* (2008) para retirada da tendência, conforme descrito na seção 2.5.7.

- **Passo 2)** Modelagem do Preço CME e do câmbio.

O preço da soja apresenta um comportamento menos previsível do que a produtividade, por conta do seu amplo mercado internacional. A série de preços e dólar apresentam uma dinâmica muito mais complexa, pois são influenciadas por uma série de fatores externos, indo muito além das questões de oferta e demanda domésticas.

Duas das características destas séries são a presença de tendência e variância não constante ao longo do tempo. Por isso, optou-se por trabalhar com a série diferenciada, transformando-a em série de preços ajustados ao final do dia de negociação, retirando assim, a tendência destas séries. Além disso, para extrair o que é importante para o risco do preço, a modelagem da volatilidade dos preços das séries univariadas marginais é incorporada ao modelo final de cópulas.

- **Passo 3)** Modelagem das Cópulas.

Para o estudo da estrutura de dependência das variáveis em questão, utiliza-se as funções cópulas para a construção da função multivariada. Neste trabalho, utilizou-se duas abordagens: bidimensional e a tridimensional.

Na abordagem bidimensional, a série de preços diárias CME e dólar foram multiplicadas e transformadas numa única série de preços em reais, e então, realizou-se toda a análise e modelagem das marginais e da cópula. Neste caso, considerou-se correlação nula entre estas variáveis, esta abordagem é similar àquela utilizada pela seguradora B na precificação do seguro de faturamento agrícola.

Por sua vez, o enfoque tridimensional leva em consideração a correlação existente entre as variáveis preços CME e dólar. Para o entendimento do risco do preço da soja brasileira, influenciado pelas exportações e, conseqüentemente, preços internacionais, optou-se por incorporar a variável exógena câmbio ao modelo da precificação do seguro de faturamento.

4.4.1 Modelagem da Produtividade

Conforme descrito anteriormente, com o passar dos anos há um aumento na produtividade de soja, atribuído principalmente à incorporação da tecnologia empregada no campo.

Por isso, com intuito de estudar o risco de produtividade retirou-se essa tendência das séries através do procedimento determinístico sugerido por GALLAGHER (1987), GOODWIN and MAHUL (2004) e TEJEDA *et al.* (2008), conforme descrito na seção 2.5.7.

Para a modelagem da produtividade de soja em Cascavel e Toledo utilizou-se as distribuições Normal, SN, OLLN e ST, conforme apresentado no capítulo 2 e DUARTE *et al.* (2017).

4.4.2 Modelagem do Preço e Dólar

Um dos objetivos em finanças é a avaliação de riscos de uma carteira de ativos financeiros. Neste trabalho, o objetivo é a avaliação dos riscos na negociação da *commodity* de soja na bolsa CME. O risco é frequentemente medido em termos de variações de preços dos ativos.

Suponha Y_t o preço da *commodity* no instante t , normalmente um dia de negócio na CME ou o preço da cotação do dólar (Z_t). A variação de preços entre os instantes t e $t - 1$ é dado pela diferença $\Delta Y_t = Y_t - Y_{t-1} = d_t$, tornando-a em ajustes de preços diários.

Em finanças utiliza-se muito a variação relativa de preços, chamado de retorno líquido simples ($R_t = \frac{\Delta P_t}{P_{t-1}}$) ou log-retorno $r_t = \log \frac{P_t}{P_{t-1}}$. Segundo MORETTIN (2008) é preferível trabalhar com retornos, pois estes são livres de escalas e tem propriedades estatísticas mais interessantes (como estacionariedade e ergodicidade).

Os modelos da família ARMA, ARCH e GARCH são utilizados para modelar retornos. Os retornos normalmente apresentam algumas características interessantes: os retornos são auto correlacionados, os quadrados dos retornos são auto correlacionados, apresentam agrupamentos de volatilidades ao longo do tempo e a distribuição dos retornos apresentam caudas mais pesadas do que a distribuição normal. Além disso, a distribuição, embora aproximadamente simétrica, é, em geral, leptocúrtica. Estes fatos são conhecidos como fatos estilizados relativos aos retornos (MORETTIN, 2008).

Portanto, a volatilidade determinística pode ser calculada pelos modelos da família ARCH, que foram introduzidos por ENGLE (1982), com o propósito de estimar a variância da inflação. A ideia é que os retornos são serialmente não correlacionados, mas a variância condicional (volatilidade) depende de retornos passados através de uma função quadrática (MORETTIN, 2008).

Um modelo ARCH de ordem m , denotado por ARCH(m) é definido por:

$$r_t = \sqrt{h_t} \epsilon_t$$

$$h_t = \alpha_0 + \alpha_1 r_{t-1}^2 + \dots \alpha_m r_{t-m}^2$$

em que ϵ_t são independentes e identicamente distribuídos (iid) com média zero, $\alpha_0 > 0$, $\alpha_m > 0$ e $\alpha_i \geq 0$, $i = 1, \dots, m-1$.

Uma desvantagem desse modelo é que os retornos positivos e negativos são tratados de forma semelhante, pois o quadrado dos retornos está presente na fórmula de volatilidade (MORETTIN, 2008).

4.4.2.1 Modelos GARCH

Uma generalização dos modelos ARCH foi proposto por BOLLERSLEV (1986), chamado modelo Generalized ARCH (GARCH). Estes modelos podem ser mais parcimoniosos do que os modelos ARCH. Um modelo GARCH(m,n) é definido por:

$$r_t = \sqrt{h_t} \epsilon_t$$

$$h_t = \alpha_0 + \sum_{i=1}^m \alpha_i r_{t-i}^2 + \sum_{j=1}^n \beta_j h_{t-j}$$

em que ϵ_t são iid com média zero, $\alpha_0 > 0$, $\alpha_m > 0$, $\alpha_i \geq 0$, $i = 1, \dots, m-1$, $\beta_j \geq 0$, $j = 1, \dots, n-1$, $\beta_n > 0$, $\sum_{i=1}^q (\alpha_i + \beta_i) < 1$, $q = \max(m, n)$.

Geralmente assume-se que os erros ϵ_t são normalmente distribuídos ou seguem uma distribuição t-Student ou uma "generalized error distribution"(ged), descritos em NELSON (1991).

De acordo com MORETTIN (2008), na prática é recomendado definir modelos de ordem baixa, por exemplo, GARCH(1,1), (1,2) ou (2,1).

4.4.3 Cópias

Nesta seção apresenta-se uma generalização da metodologia de cópias descrita na seção 3.6.

De acordo com JOE (2014), a cópia é uma distribuição multivariada em que todas as distribuições unidimensionais marginais são uniformes $U(0,1)$. Consequentemente, se C é uma cópia, então a distribuição de um vetor de variáveis aleatórias dependentes é $U(0,1)$.

O resultado a seguir é uma extensão do teorema de Sklar, descrito em SKLAR (1959).

Teorema 2. Para uma distribuição d -dimensional $F \in \mathfrak{F}(F_1, F_2, \dots, F_d)$, em que F_j é a j -ésima distribuição univariada marginal, a cópia associada a F é uma função de distribuição $C : [0, 1]^d \rightarrow [0, 1]$,

$$F(\mathbf{y}) = C(F_1(y_1), F_2(y_2), \dots, F_d(y_d)), \mathbf{y} \in \mathbb{R}^d, \quad (4.2)$$

com marginais $U(0,1)$ que satisfazem:

- (a) Se F é uma distribuição d -dimensional contínua com marginais uniformes $F_1, \dots, F_d$ e funções quantílicas $F_1^{-1}, \dots, F_d^{-1}$, então:

$$C(\mathbf{u}) = F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)), \mathbf{u} \in [0, 1]^d$$

é única.

- (b) Se F é uma distribuição d -dimensional de variáveis aleatórias discretas, então a cópula é única somente no conjunto $\text{Dom}(F_1) \times \dots \times \text{Dom}(F_d)$

A demonstração deste teorema pode ser encontrado na página 8 de JOE (2014).

A cópula C é parametrizada por um vetor de parâmetros δ , chamado de parâmetro de dependência multivariado.

As famílias de cópulas são geralmente dadas como função de distribuição acumulada (para a inferência via verossimilhança) e a cópula densidade é obtida pela diferenciação. Se $C(\mathbf{u})$ é uma distribuição acumulada absolutamente contínua então sua função densidade é dada por:

$$c(\mathbf{u}) = c(u_1, \dots, u_d) = \frac{\partial^d C(\mathbf{u})}{\partial u_1 \dots \partial u_d}, \mathbf{u} \in [0, 1]^d. \quad (4.3)$$

Existem diferentes cópulas multidimensionais. Apresentam-se a seguir as cópulas Gaussiana e t-Student multivariadas.

- **Cópula Gaussiana multivariada:** A cópula Gaussiana ou Normal é obtida da distribuição multivariada Gaussiana com média zero, variâncias 1 e matriz de correlação $\mathbf{R}_{d \times d}$, definida como:

$$C(\mathbf{u}; \mathbf{R}) = \Phi_d(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d); \mathbf{R}) = \quad (4.4)$$

$$= \int_{-\infty}^{\Phi^{-1}(u_1)} \dots \int_{-\infty}^{\Phi^{-1}(u_d)} \frac{1}{(2\pi)^{d/2} |\mathbf{R}|^{1/2}} \exp\left(-\frac{1}{2} \mathbf{y}^T \mathbf{R}^{-1} \mathbf{y}\right) dy_1 \dots dy_d, \quad (4.5)$$

em que $\mathbf{u} \in [0, 1]^d$, $\Phi_d(\cdot; \mathbf{R})$ é a d -ésima distribuição acumulada Normal padrão, Φ é a distribuição Normal padrão e Φ^{-1} é a função distribuição Normal inversa ou distribuição quantílica.

A densidade da cópula Normal multivariada é dada por:

$$c(\mathbf{u}; \mathbf{R}) = \frac{\phi_d(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d); \mathbf{R})}{\prod_{j=1}^d \phi(\Phi^{-1}(u_j))}, \mathbf{u} \in (0, 1)^d,$$

em que $\mathbf{R}$ é positiva definida.

- **Cópula t_ν multivariada:** Seja $T_{1,\nu}$ ou T_ν a distribuição acumulada univariada t-Student com parâmetro $\nu > 0$ e $T_{d,\nu}(\cdot; \Sigma)$ a d -ésima distribuição t-Student com parâmetros $\nu > 0$ e matriz de covariância Σ . Similarmente, seja $t_{1,\nu}$ (ou t_ν) e $t_{d,\nu}(\cdot; \Sigma)$ a correspondente para as densidades.

Se $\Sigma = \mathbf{R}$ é a matriz de correlação então a marginal univariada de $T_{d,\nu}(\cdot; \Sigma)$ são iguais para $T_{1,\nu}$.

A d -ésima função densidade t_ν com matriz de correlação $\mathbf{R}$ é:

$$t_{d,\nu}(\mathbf{y}; \mathbf{R}) = |\mathbf{R}|^{-1/2} \frac{T((\nu + d)/2)}{T(\nu/2)(\pi\nu)^{d/2}} (1 + \mathbf{y}^T \mathbf{R}^{-1} \mathbf{y}/\nu)^{-(\nu+d)/2}, \mathbf{y} \in \mathbb{R}^d.$$

Pela transformada da probabilidade integral, a distribuição acumulada da cópula t-Student é:

$$C_{\mathbf{u}; \mathbf{R}, \nu} = T_{d,\nu}(T_{1,\nu}^{-1}(u_1), \dots, T_{1,\nu}^{-1}(u_d); \mathbf{R}), \mathbf{u} \in [0, 1]^d.$$

A densidade multivariada da cópula t_ν é:

$$C_{u; \mathbf{R}, \nu} = \frac{T_{d,\nu}(T_{1,\nu}^{-1}(u_1), \dots, T_{1,\nu}^{-1}(u_d))}{\prod_{j=1}^d t_{1,\nu}(T_{1,\nu}^{-1}(u_j))}, \mathbf{u} \in [0, 1]^d. \quad (4.6)$$

- A descrição das cópulas multidimensionais da família Arquimediana estão em CHERUBINI *et al.* (2004).

4.4.3.1 Inferência das Cópulas Multivariadas

Para estimar os parâmetros das cópulas, os métodos mais comuns são baseados na verossimilhança. Suponha uma amostra iid, $y_{i1}, \dots, y_{id}, i = 1, \dots, d$, então a distribuição conjunta é:

$$F(\cdot; \eta_1, \dots, \eta_d, \delta) = C(F_1(\cdot; \eta_1), \dots, F_d(\cdot; \eta_d); \delta). \quad (4.7)$$

Se todas as variáveis forem contínuas e $F_1, \dots, F_d, C$ forem absolutamente contínuas com relação as densidades $f_1, \dots, f_d, c$, então a log-verossimilhança é dada por:

$$l(\theta) = \sum_{i=1}^n \{ \log c(F_1(y_{i1}; \eta_1), \dots, F_d(y_{id}; \eta_d); \delta) + \sum_{j=1}^d \log f_j(y_{ij}; \eta_j) \} \quad (4.8)$$

em que $\theta = (\eta_1, \dots, \eta_d, \delta)^t$ é o vetor de todos os parâmetros das marginais e da cópula. Assim, o estimador de máxima verossimilhança (EMV) é obtido maximizando a expressão (4.8), $\hat{\theta}_{EMV} = \max_{\theta} l(\theta)$.

Sob condições usuais de regularidade, o estimador de máxima verossimilhança existe e é consistente e assintoticamente eficiente. Além disso, verifica-se a propriedade de assintoticamente normal:

$$\sqrt{d}(\hat{\theta}_{EMV} - \theta_0) \rightarrow N(0, \mathfrak{F}^{-1}(\theta_0)),$$

em que $\mathfrak{F}$ é a matriz de informação de Fisher e θ_0 o valor verdadeiro usual.

A matriz de covariância de $\hat{\theta}_{EMV}$ (matriz de informação de Fisher) pode ser estimada pelo inverso da matriz Hessiana negativa da função de verossimilhança.

O método de máxima verossimilhança pode ser computacionalmente dispendioso ou exigente, especialmente no caso de alta dimensão, pois é necessário estimar conjuntamente os parâmetros das distribuições marginais e os parâmetros da estrutura de dependência representado pela cópula.

No entanto, a função de log-verossimilhança é composta por dois termos positivos: um termo envolvendo a densidade da cópula e seus parâmetros, e outro termo envolvendo as marginais e todos os seus parâmetros.

Por essa razão, JOE and XU (1996) propuseram que esses conjuntos de parâmetros podem ser estimados em duas etapas:

1. No primeiro passo, estima-se os parâmetros das marginais $\theta_1 = (\eta_1, \dots, \eta_d)$ ao realizar as estimativas das marginais $\hat{\theta}_1 = \max \sum_{i=1}^n \sum_{j=1}^d \log f_j(y_{ij}; \eta_j)$.
2. No segundo passo, dado $\hat{\theta}_1$ estima-se os parâmetros da cópula θ_2 ,

$$\hat{\theta}_2 = \max \sum_{i=1}^n \{ \log c(F_1(y_{i1}), \dots, F_d(y_{id}); \theta_2; \hat{\theta}_1) \}.$$

Este método é chamado de inferência por marginais ("inference for margins-IFM). O estimador IFM é definido como o vetor:

$$\hat{\theta}_{IFM} = (\hat{\theta}_1, \hat{\theta}_2)^t.$$

Seja l toda a função de verossimilhança, l_j a log-verossimilhança da j -ésima marginal e l_c a log-verossimilhança da própria cópula. Assim, o estimador IFM é a solução de:

$$\left(\frac{\partial l_1}{\partial \eta_1}, \frac{\partial l_2}{\partial \eta_2}, \dots, \frac{\partial l_d}{\partial \eta_d}, \frac{\partial l_c}{\partial \theta_2} \right)^t = \mathbf{0}^t. \quad (4.9)$$

Por outro lado, o estimador de máxima verossimilhança (EMV) obtém-se da resolução de:

$$\left(\frac{\partial l}{\partial \eta_1}, \frac{\partial l}{\partial \eta_2}, \dots, \frac{\partial l}{\partial \eta_d}, \frac{\partial l}{\partial \theta_2} \right)^t = \mathbf{0}^t. \quad (4.10)$$

De acordo com CHERUBINI *et al.* (2004), o estimador IFM fornece um bom ponto de partida para obter um estimador EMV exato, pois é computacionalmente mais fácil obter o estimador IFM do que o EMV. Assim, é interessante abordar a questão sobre a eficiência assintótica do IFM em comparação com o EMV.

Portanto, é preciso comparar a matriz de covariância assintótica dos dois estimadores. A teoria IFM é um caso especial de usar um conjunto de equações de inferência apropriadas para estimar um vetor de parâmetros. Neste caso, cada equação é uma função escore (isto é, o lado esquerdo é a derivada parcial da log-verossimilhança de cada densidade marginal).

De acordo com JOE (1997) assim como o EMV, sob condições normais o estimador IFM verifica a propriedade da normalidade assintótica:

$$\sqrt{d}(\hat{\theta}_{IFM} - \theta_0) \rightarrow N(0, G^{-1}(\theta_0)),$$

em que G^{-1} é a matriz de informação de Godambe.

Assim, definimos a função escore como:

$$s(\theta) = \left(\frac{\partial l_1}{\partial \eta_1}, \frac{\partial l_2}{\partial \eta_2}, \dots, \frac{\partial l_d}{\partial \eta_d}, \frac{\partial l_c}{\partial \theta_2} \right)^t = \mathbf{0}^t, \quad (4.11)$$

dividindo a log-verossimilhança em duas partes $l_1, l_2, \dots, l_n$ para cada marginal e l_c para a cópula, a matriz de informação Godambe assume a seguinte forma:

$$G^{-1}(\theta_0) = D^{-1}V(D^{-1})^t,$$

em que $D = E\left[\frac{\partial s(\theta)}{\partial \theta}\right]$ e $V = E(s(\theta), s(\theta)^t)$.

JOE (1997) destaca que o método IFM é altamente eficiente em comparação com o método de máxima verossimilhança.

4.4.3.2 Inferência via Pseudo Máxima-Verossimilhança

No estudo da dependência com cópulas, existem várias possibilidades para a escolha da distribuição marginal e da família de cópulas. Para variáveis contínuas, o parâmetro estimado da cópula pode ser comparado usando a aproximação por pseudo máxima-verossimilhança, também conhecido como máxima verossimilhança canônica, proposta por GENEST *et al.* (1995).

Este é um método semi-paramétrico usando as distribuições empíricas $\hat{F}_1, \dots, \hat{F}_d$ e maximizando:

$$L_{pseudo}(\delta) = \sum_{i=1}^n \{\log c(\hat{F}_1(y_{i1}), \dots, \hat{F}_d(y_{id}); \delta)\},$$

em que δ é o parâmetro da cópula e $\hat{F}_j(y) = n^{-1} \{\sum_{i=1}^n I(y_{ij} \leq y) - 0, 5\}$.

Segundo JOE (2014), a pseudo máxima-verossimilhança com distribuição empírica para as marginais somente pode ser implementada se as variáveis forem contínuas e não existirem covariáveis ou observações censuradas.

Se o modelo de cópula é especificado corretamente, então GENEST *et al.* (1995) mostra que δ_{pseudo} é consistente e assintoticamente normal. Além disso, CHEN and FAN (2006) provou outras propriedades assintóticas quando se utiliza a inferência semi-paramétrica no contexto de séries temporais e se desenvolve métodos de estimação para as funções cópulas e marginais unidimensionais.

A seleção da cópula, quando aplicável a abordagem de pseudo-verossimilhança é comumente realizada pela comparação de AIC (AKAIKE, 1974) ou BIC (SCHWARZ *et al.*, 1978), baseado nas pseudo logs-verossimilhança.

Neste trabalho, optou-se por adotar este método para a estimação dos parâmetros da cópula.

4.4.4 Ferramenta Gráfica para Detectar Dependência - Chi-plot

Além do gráfico de dispersão, o Chi-plot é outra ferramenta gráfica usada na literatura para detectar dependência entre duas variáveis, proposto por FISHER and SWITZER (2001).

A descrição desta seção é baseada nos trabalhos de GENEST and FAVRE (2007) e TURSUNALIEVA and SILVAPULLE (2007).

O gráfico de Chi-plot é baseado nos ranks dos dados. O Chi-plot é um gráfico de dispersão entre os pares (λ_i, χ_i) que são transformações dos n pares (X_i, Y_i) , $i = 1, \dots, n$.

Seja H uma função de distribuição bivariada contínua e F_n e G_n as distribuições marginais de X e Y , respectivamente. Estas podem ser estimadas para cada par (X_i, Y_i) como se segue:

$$H_i = \frac{1}{n-1} \sum_{j \neq i} I(X_j \leq x_i, Y_j \leq y_i) \quad (4.12)$$

$$F_i = \frac{1}{n-1} \sum_{j \neq i} I(X_j \leq x_i) \quad (4.13)$$

$$G_i = \frac{1}{n-1} \sum_{j \neq i} I(Y_j \leq y_i), \quad (4.14)$$

$$\text{em que } I(X_j \leq x_i) = \begin{cases} 1, & \text{se } X_j \leq x_i \\ 0, & \text{caso contrário.} \end{cases}$$

FISHER and SWITZER (2001) propõe plotar o gráfico dos pares (λ_i, χ_i) , que são definidos por:

$$\chi_i = \frac{H_i F_i G_i}{\sqrt{F_i(1-F_i)G_i(1-G_i)}} \quad (4.15)$$

$$\lambda_i = 4S_i \max\{(F_i - 0,5)^2, (G_i - 0,5)^2\} \quad (4.16)$$

em que $S_i = \text{sign}\{(F_i - 0,5)^2, (G_i - 0,5)^2\}$.

O gráfico do Chi-plot apresenta bandas de confiança (retas tracejadas no gráfico) que são desenhadas em $\pm \frac{c_p}{\sqrt{n}}$ (o valor aproximado para c_p com nível de significância de 95% é 1,78).

Os pares (λ_i, χ_i) de marginais independentes e contínuas tendem a se localizarem dentro das bandas.

Para marginais negativamente correlacionadas, os pares de (λ_i, χ_i) tendem a espalhar abaixo da faixa inferior. Por outro lado, se as marginais forem positivamente correlacionadas, os pares de (λ_i, χ_i) tendem a espalhar acima da faixa superior. Por fim, pares próximos de zero $\chi_i = 0$ e entre as faixas indicam independência entre as variáveis.

4.4.5 Teste de independência multivariada

Para a construção das cópulas, as variáveis estudadas $Y_1, Y_2, \dots, Y_n$ devem ser uma amostra aleatória proveniente da função de distribuição conjunta $H(Y_1, Y_2, \dots, Y_n)$. Segundo MIQUELETO (2011), para garantir que a amostra seja aleatória, as variáveis devem ser mutuamente independentes. Em outras palavras, se a hipótese nula de que as variáveis são independentes for rejeitada, isso garante evidências de que $Y_1, Y_2, \dots, Y_n$ não foram geradas aleatoriamente. A descrição desta seção é baseada nos trabalhos de KOJADINOVIC and YAN (2011) e GENEST and RÉMILLARD (2004).

Sabe-se que a estrutura de dependência entre as variáveis é completamente resumida pela cópula C e a independência mútua entre elas ocorre se, e somente se, $C(u_1, u_2, \dots, u_n) = \prod_{i=1}^n u_i$, então um teste da independência mútua dos componentes de $Y_i, i = 1, \dots, n$, pode ser a estatística I_n como proposto por GENEST and RÉMILLARD (2004):

$$I_n = \int_{[0,1]^d} n(C_n(u) - \Pi_{i=1}^n u_i)^2 du.$$

Sob a hipótese de independência $Y_1, Y_2, \dots, Y_n$ o processo empírico $\sqrt{n}(C_n - \Pi_{i=1}^n u_i)$ pode ser decomposto usando a transformada de Mobius, em uma coleção de $2^d - d - 1$ subprocessos $\sqrt{n}M_A(C_n)$, em que $A \subseteq \{1, \dots, d\}$, tal que $|A| > 1$, que converge conjuntamente para processo gaussiano mutuamente independente.

GENEST and RÉMILLARD (2004) propõe que $Y_1, Y_2, \dots, Y_n$ mutualmente independente é equivalente a ter $M_A(C(u)) = 0, \forall u \in [0, 1]^d$, $A \subseteq \{1, \dots, d\}$ tal que $|A| > 1$. Ao invés de simples estatística I_n , os autores sugerem considerar $2^d - d - 1$ estatísticas do teste na forma:

$$M_{A,n} = \int_{[0,1]^d} n(M_A(C_n(u)))^d du,$$

que são assintoticamente independentes sob a hipótese de independência.

Para visualizar os resultados dos testes, quando baseados em todas as $2^d - d - 1$ estatísticas $M_{A,n}$, uma representação gráfica, chamada dependograma pode ser usada.

Para cada subconjunto $A \subseteq \{1, \dots, d\}, |A| > 1$, uma barra vertical será desenhada cuja altura é proporcional ao valor de $M_{A,n}$. Os valores críticos de $M_{A,n}$ são representados por bolas pretas. Subconjuntos para os quais a barra excede o valor crítico pode ser considerado como sendo variáveis dependentes.

4.5 Resultados

Nesta seção apresenta-se a modelagem das séries de produtividade de soja, preço futuro CME e cotação do dólar. Além disso, apresenta-se as estimativas das cópulas sob a abordagem bidimensional e tridimensional.

A Tabela 4.1 apresenta as estatísticas descritivas para as séries originais (sem nenhuma correção) de preços CME (pcme), cotação do dólar (pdolar), a série de preços em reais (preais) e a série de produtividades de Cascavel e Toledo.

Tabela 4.1. Estatísticas Descritivas para as séries originais de preços CME (pcme), cotação do dólar (pdolar) e preços em reais (preais) e produtividades para Cascavel e Toledo.

Série	Média	Mediana	Desvio Padrão	Assimetria	Curtose	Valor Máximo	Valor Mínimo	CV(%)
pcme	21,254	21,147	1,361	0,343	2,038	24,652	19,190	6,404
pdolar	3,471	3,393	0,314	0,457	1,909	4,194	2,988	9,053
preais	73,504	73,783	5,0037	-0,127	2,254	84,909	62,801	6,807
Cascavel	2622,946	2500	574,760	0,248	2,167	3655	1583	21,912
Toledo	2638,297	2687	651,752	-0,298	2,232	3583	1131	24,703

Observa-se um preço médio de US\$ 21,254 para a saca de 60 Kg de soja e coeficiente de variação de 6,40%. Além disso, nota-se que o preço médio em reais é R\$ 73,504 e coeficiente de variação 6,80%. Já a cotação média do dólar no período analisado é R\$ 3,471 e coeficiente de variação 9,05%.

Além disso, nota-se uma produtividade média 2622,946 Kg/ha e desvio padrão 574,76 em Cascavel. Ademais, o coeficiente de variação em Toledo (24,703 %) indica um risco de produtividade mais elevado do que em Cascavel (21,912 %).

A Figura 4.1(a) apresenta as séries originais de produtividade de Cascavel (em preto) e Toledo (em verde). Nota-se, em ambas as séries, uma tendência crescente ao longo dos anos. Além disso, observa-se uma variabilidade maior na série de produtividade em Toledo. A Figura 4.1(b) apresenta as séries de preços futuros (US\$ por saca de 60 Kg) negociados na CBOT/CME (eixo vertical à esquerda) e a cotação do dólar (R\$) no mesmo período (eixo vertical à direita), entre os dias 05/04/2015 à 14/03/2017,

totalizando 455 observações. Observa-se no ano de 2015 uma alta na cotação do dólar seguido pela sua diminuição a partir de fevereiro de 2016. Além disso, observa-se um aumento dos preços da soja a partir de abril de 2016, atingindo o pico em 06/10/2016. Além disso, a série de preços não é estacionária, por isso, optou-se por diferenciá-la, tornando-a em ajustes diários.

A Figura 4.1(c) apresenta a série de preços em reais, nota-se, tendência e variabilidade não constante ao longo do tempo.

Figura 4.1. Séries Originais de Produtividade e Preços.

4.5.1 Produtividade

A Tabela 4.2 apresenta as estatísticas descritivas para as séries de produtividades de Cascavel e Toledo corrigidas quanto à tendência. Observa-se uma produtividade média 3405,587 Kg/ha, desvio

padrão 329,91 e medida de assimetria -0,329 para o município de Cascavel. Indicando uma pequena assimetria à esquerda (da distribuição normal). Além disso, o coeficiente de variação em Toledo (15,82%) indica um risco de produtividade mais elevado do que em Cascavel (9,687%).

Tabela 4.2. Estatísticas Descritivas para as séries de produtividades corrigidas para Cascavel e Toledo (n=37 observações).

Série	Média	Mediana	Desvio Padrão	Assimetria	Curtose	Valor Máximo	Valor Mínimo	CV(%)
Cascavel	3405,587	3447,442	329,9146	-0,329	-0,676	4021,803	2707,002	9,68
Toledo	3290,044	3442,873	520,7424	-1,257	1,890	4005,122	1505,833	15,82

As séries de produtividade anuais para os municípios de Cascavel e Toledo foram modeladas utilizando a distribuição OLLN, conforme descrito em DUARTE *et al.* (2017).

A Tabela 4.3 apresenta os parâmetros estimados relativos à distribuição $OLLN(\mu, \sigma, \alpha)$ para a série de produtividade com erros padrões entre parênteses. Observe que os parâmetros são muito parecidos para os dois municípios, indicando que estes apresentam características de produtividades parecidas.

O resultado dessas estimativas permite simular o número necessário de observações e suas respectivas probabilidades acumulativas que sejam úteis para o pareamento com as probabilidades referentes aos preços e dólar. Assim, com as séries com o mesmo número de observações pode-se obter as estimativas das cópulas.

Tabela 4.3. Parâmetros estimados da simulação da distribuição $OLLN(\mu, \sigma, \alpha)$, com erros padrões entre parênteses,

Municípios	μ	σ	α
Toledo	3289,649 (117,051)	637,864 (338,379)	1,042 (0,679)
Cascavel	3415,344 (72,626)	680,888 (797,095)	1,632 (2,086)

4.5.2 Análise Bivariada para o Contrato de Faturamento

Considere X a produtividade simulada para os municípios de Toledo ou Cascavel (Kg/ha), Y o preço da CBOT/CME (US\$ por saca de 60 Kg) e Z a cotação do dólar.

Para o cálculo do faturamento, vamos considerar o preço $P = Y \times Z$, neste caso, não estamos levando em consideração a variabilidade do câmbio. O faturamento F será considerado como uma função de duas variáveis $F = X \times P$. Este procedimento não considera a estrutura de dependência entre as variáveis preço e o dólar. Equivale a assumir que a correlação entre as variáveis é inexistente. Neste caso, estamos assumindo apenas existência de uma estrutura de dependência entre a produtividade e o preço em R\$.

A Tabela 4.4 apresenta as estatísticas descritivas para as séries de produtividades simuladas (20.000 observações) da distribuição OLLN para Cascavel, Toledo e os preços em reais. Observa-se uma produtividade média de 3402,604 Kg/ha, desvio padrão 425,01 e medida de assimetria 0,0015 para o município de Cascavel. Além disso, as estatísticas são próximas entre si, sugerindo uma taxa de prêmio muito parecida para os dois municípios. Ademais, as estatísticas apresentadas na Tabela 4.4 estão próximas dos originais, apresentadas na Tabela 4.2, demonstrando que as simulações foram adequadamente implementadas.

Além disso, a média dos preços em reais é 73,504 por saca de 60 Kg, com coeficiente de variação 6,80%.

A Figura 4.2(a) apresenta o Preço em Reais do contrato futuro de soja entre os dias 05/04/2015 à 14/03/2017, totalizando 455 observações. Há uma tendência apresentada pelo teste de Dickey-Fuller (p-valor=0,4537), por isso, optou-se por diferenciá-la, tornando-a em ajustes diários. A diferenciação da

Tabela 4.4. Estatísticas Descritivas para as séries de produtividades simuladas para Cascavel, Toledo e preços.

Série	Média	Mediana	Desvio Padrão	Assimetria	Curtose	Valor Máximo	Valor Mínimo	CV(%)
Cascavel	3402,604	3406,697	425,0114	0,0015	-0,201	4752,302	2085,194	12,49
Toledo	3279,669	3285,384	593,0511	0,0013	-0,192	5167,45	1437,07	18,08
Preço(R\$)	73,504	73,783	5,003	-0,127	-0,755	84,909	62,801	6,80

série não impede a análise de um contrato de seguro de faturamento, pois a diferença do preço garantido p_g e o preço ocorrido será o ajuste.

Os gráficos da Figura 4.2(b), 4.2(c) e 4.2(d) apresentam o Preço Ajustado (diferença do preço original), Histograma e QQ-Plot Normal para a série de preço de ajuste em R\$, respectivamente. Observa-se, que a série de preços ajustada oscila ao redor do zero, aparentando estacionariedade, comprovado pelo teste de Dickey-Fuller ($p\text{-valor} < 0,01$, rejeitando a hipótese de existência de raiz unitária), mas com variabilidade que depende do tempo (volatilidade).

Além disso, pelo gráfico qq-plot normal, a série ajustada apresenta cauda mais pesada do que a normal. A preocupação central na área atuarial é com a cauda esquerda da distribuição, pois é a região que se observa dependência entre valores extremos e é utilizada para calcular o risco de redução de produtividade (probabilidade de perda). A subestimação da probabilidade de perda implica em taxas de prêmio menores, causando a inviabilização do plano de seguro, tornando-o deficitário para a seguradora.

A Figura 4.3 apresenta as funções ACF e PACF para a série de preços ajustada em reais e esta ao quadrado. Além disso, aplicou-se o teste de Ljung-Box para a série de preços ajustada ($p\text{-valor} = 0,03963$) e esta ao quadrado ($p\text{-valor} = 0,00864$). Observa-se a ausência de autocorrelação serial e presença de heterocedasticidade condicional a nível de significância de 0,01. Portanto, realizou-se a modelagem da heterocedasticidade utilizando modelos da família GARCH(m,n).

A Tabela 4.5 apresenta os valores dos critérios de AIC e BIC e máximo da log-verossimilhança para os modelos da família GARCH(m,n) com erros t-Student para a série de preços ajustada. Pelos critérios AIC e BIC o modelo escolhido foi GARCH(1,1) com erros t-Student.

Aplicou-se o teste de Ljung-Box para os resíduos do modelo escolhido, com $p\text{-valor} 0,626727$. Considerando nível de significância 0,01, o teste indica ausência de autocorrelação. Por outro lado, o resultado do teste de Ljung-Box para os resíduos ao quadrado do modelo escolhido apresentou $p\text{-valor} 0,04475$, o que indica ausência de heterocedasticidade condicional, considerando nível de significância 0,01.

Tabela 4.5. Valores dos critérios de AIC e BIC e do máximo da log-verossimilhança para os modelos ajustados à série de preço em reais.

	Critério	GARCH(1,1)	GARCH(2,1)	GARCH(2,2)
Preço (R\$)	AIC	2,851265	2,856418	2,857883
	BIC	2,896618	2,910905	2,921377
	LLV	642,2371	642,4212	641,7393

Portanto, com a série de produtividade simulada da distribuição OLLN para cada município selecionado e o resíduo do modelo GARCH(1,1) para o preço ajustado, ajustou-se o modelo bidimensional de cópulas.

A Tabela 4.6 apresenta os valores de AIC e BIC para as cópulas ajustadas para os municípios selecionados. Para ambos os municípios, a cópula que melhor se ajusta a estrutura dos dados é a cópula Frank com parâmetro $\delta = -0,3358$ e com erro padrão (0,289).

Portanto, a taxa do prêmio para Cascavel e Toledo será a mesma, pois estes municípios apresentam a mesma estrutura de dependência entre a produtividade e preço. A discussão das taxas do prêmio de seguro de faturamento utilizando a cópula selecionada será apresentada na subseção 4.6.1.

Figura 4.2. Preço Original em reais, Preço Ajustado(R\$), Histograma e QQ-Plot Normal para a série de preço de ajuste em reais.

Tabela 4.6. Valores do critérios de AIC e BIC para Seleção de Cópulas.

	Critério	Frank	AMH	Joe	Gaussiana	t	Gumbel
Cascavel	AIC	0,5473	0,657	2	1,1088	3,1089	2
	BIC	4,6654	4,7751	6,1181	5,2269	11,3451	6,1181
Toledo	AIC	0,5473	0,657	2	1,1088	3,1089	2
	BIC	4,6654	4,7751	6,1181	5,2269	11,3451	6,1181

4.5.3 Análise Tridimensional para o Contrato de Faturamento

Nesta seção, apresenta-se o enfoque tridimensional da análise do problema do seguro de faturamento. Então, considera-se a estrutura de dependência entre produtividade, preços (US\$) e a cotação do dólar, ou seja, leva-se em consideração a variabilidade do câmbio na modelagem.

A Tabela 4.7 apresenta as estatísticas descritivas para as séries de preços ajustadas (US\$) e a taxa do dólar.

Observe que o preço ajustado apresenta média de 21,254 (US\$ por saca de 60Kg), desvio padrão

Figura 4.3. Gráficos das funções ACF e PACF para a Série de Preço (R\$) ajustada e estas ao quadrado.

Tabela 4.7. Estatísticas Descritivas para as séries ajustadas de preços (US\$) e cotação do dólar(R\$).

Série	Média	Mediana	Desvio Padrão	Assimetria	Curtose	Valor Máximo	Valor Mínimo	CV(%)
Preço CME	21,254	21,147	1,361	0,343	2,038	24,652	19,190	6,404
Dólar	3,471	3,393	0,314	0,457	1,909	4,194	2,988	9,053

1,361 e coeficiente de variação 6,40%. Por outro lado, o preço médio do câmbio é R\$ 3,47 e desvio padrão 0,314. Além disso, o coeficiente de variação do câmbio é maior do que o do preço, respectivamente, com 9,05% e 6,40%. Indicando que o risco do câmbio é maior do que o risco do preço, o que é esperado, visto que o câmbio é influenciado por muitas outras variáveis não apenas preço de *commodities*.

As Figuras 4.4 e 4.5 apresentam o preço ajustado (preço diferenciado) para CME e o dólar, seus histogramas e o gráfico QxQ-plot normal, respectivamente. Ambas as séries de preços oscilam ao redor de zero, apresentando estacionariedade, mas apresentam uma variabilidade que depende do tempo (volatilidade). Além disso, o histograma para ambas as séries apresenta a parte central mais alta do que uma normal e há presença de valores bastantes afastados das posições central dos dados. Esses fatos são características de retornos financeiros e são descritos pela medida de curtose (5,168 e 6,689 para CME e Dólar, respectivamente). Ademais, pelo gráfico qq-plot normal, nota-se que as séries ajustadas apresentam caudas mais pesadas que a normal, portanto, sugere-se que estas sejam leptocúrticas.

A Figura 4.6 apresenta os gráficos da função de autocorrelação (ACF) e função de autocorrelação parcial (PACF) para as séries de preços ajustados da CME e dólar. Observa-se (Figura 4.6(c) e 4.6(d))

Figura 4.4. Preço, Histograma e QQ-Plot Normal para a série de preço de ajuste diário da CME.

Figura 4.5. Preço, Histograma e QQ-Plot Normal para a série de cotação do dólar.

que a série de preços do dólar é ruído branco, ou seja, não apresenta autocorrelação temporal. Por outro lado, a série de preços da CME apresenta autocorrelação temporal significativa para o lag 4 (Figura 4.6(a) e 4.6(b)). O que não foi comprovado pelo teste de Ljung-Box a nível de significância 0,05, cujos resultados são apresentados na Tabela 4.8. Por isso, adota-se ambas as séries de preços não autocorrelacionadas no tempo.

A Figura 4.7 apresenta as funções ACF e PACF para as séries ao quadrado da CME e dólar. A existência de heterocedasticidade condicional foi comprovada pelo testes Ljung-Box a nível de significância 0,05, para ambas as séries ao quadrado, conforme a Tabela 4.8.

Tabela 4.8. Resultados dos teste de Ljung-Box para as séries de preços ajustados da CME (rcme) e dólar (rdolar) e estas ao quadrado (rcme² e rdolar²).

Séries	rcme	rdolar	rcme ²	rdolar ²
p-valor	0,5418	0,3736	0,04513	0,0001308

Para modelagem da heterocedasticidade condicional, utilizou-se os modelos GARCH(m,n) com erros *t*-Student. A Tabela 4.9 apresenta os valores de AIC, BIC e o máximo da log-verossimilhança (LLV) para os modelos ajustados aos preços da CME e dolar. Pelo critério BIC, o menor valor para ambas as séries é para GARCH(1,1) com erros *t*-Student.

Os testes de Ljung-Box para os resíduos do modelo GARCH(1,1) com erros *t*-Student para CME e dólar, apresentaram p-valor=0,5505 e p-valor=0,7138, respectivamente. Os resultados do teste de Ljung-Box para os resíduos ao quadrado do modelo apresentaram p-valor= 0,4886 e p-valor=0,2777 para CME

Figura 4.6. Gráficos das funções ACF e PACF para os Preços Ajustados da CME e dólar.

Tabela 4.9. Valores dos critérios de AIC e BIC e do máximo da log-verossimilhança para os modelos GARCH(m,n) com erros t-Student, às séries da CME e Dólar,

	Critério	GARCH(1,1)	GARCH(2,1)	GARCH(2,2)
CME	AIC	-0,1781	-0,1806	-0,1804
	BIC	-0,1327	-0,1261	-0,1169
	LLV	-45,4174	-46,9884	-47,9477
Dólar	AIC	-3,7367	-3,7319	-3,7396
	BIC	-3,6914	-3,6774	-3,6761
	LLV	-853,2376	-853,1311	-855,8955

e dólar, respectivamente. Esses resultados indicam que as suposições do modelo estão verificadas, ou seja, ausência de autocorrelação temporal e heterocedasticidade condicional nos resíduos, respectivamente.

Portanto, os modelos GARCH(1,1) com erros *t*-Student foram os selecionados para os preços ajustados da CME e dólar.

A Figura 4.8(a) e 4.8(b) apresenta o diagrama de dispersão entre Produtividade simulada de Toledo e Cascavel, resíduo do modelo GARCH(1,1) com erros t-Student para preço ajustado CME e dólar.

A Figura 4.9 apresenta o Chi-Plot para os pares de variáveis (Produtividade Toledo, CME), (Produtividade Toledo, dólar) e (CME, dólar). Os gráficos sugerem uma dependência negativa entre o par de variáveis (CME, Dólar) e independência entre os pares (Produtividade, CME) e (Produtividade,

Figura 4.7. Gráficos das funções ACF e PACF para os Preços Ajustados da CME e dolar ao quadrado.

Figura 4.8. Diagrama de Dispersão.

Dólar). Assim, a variável preço CME pode estar negativamente correlacionada com a cotação do dólar, o que pode reduzir o impacto do risco do preço na precificação do seguro.

Figura 4.9. Chi-Plot para os pares de variáveis (Produtividade em Toledo, CME)

A Figura 4.10 apresenta o resultado do teste de independência de variáveis aleatórias via cópulas empírica proposto por GENEST and RÉMILLARD (2004). As alturas das barras representam os valores das estatísticas do teste por subconjunto de variáveis e a bola preta representa o valor crítico do teste. Além disso, a primeira barra representa o resultado do teste para o conjunto de variáveis 1,2, em que 1 representa a variável produtividade, 2 o preço cme e 3 a série do dólar. Observa-se que, com nível de significância 0,05, para o subconjunto variáveis (1,2) e (1,3) podem ser considerados independentes. Por outro lado o subconjunto de variáveis (2,3) é considerado dependente e as variáveis 1,2 e 3, entre si são consideradas mutualmente independentes. Portanto, as variáveis produtividade, preço e dólar foram geradas aleatoriamente e são considerados independentes.

Figura 4.10. Dependograma.

A Tabela 4.10 apresenta os valores dos critério de AIC e BIC e a máxima log-verossimilhança (LLV) para as diferentes cópulas para as variáveis produtividade Toledo e Cascavel, preços ajustados CME e dólar. De acordo com todos os critérios de seleção, a cópula t-Student é a que melhor representa a estrutura de dependência dos dados.

A Tabela 4.11 apresenta os modelos selecionados com o vetor de parâmetros selecionados do trio

Tabela 4.10. Critérios AIC, BIC e máxima log-verossimilhança para seleção de cópulas.

	Critério	Clayton	Frank	Gumbel	Joe	Gaussiana	t
Toledo	AIC	-116,3977	-115,6631	-101,3985	-67,118	-106,273	-450,5504
	BIC	-112,2796	-111,545	-97,2804	-62,9999	-102,1549	-438,1961
	LLV	59,1988	58,8316	51,6993	34,559	54,1365	228,2752
Cascavel	AIC	-116,3977	-115,6631	-101,3985	-67,118	-106,273	-450,5504
	BIC	-112,2796	-111,545	-97,2804	-62,9999	-102,1549	-438,1961
	LLV	59,1988	58,8316	51,6993	34,559	54,1365	228,2752

de variáveis. Observe que para ambos os municípios a cópula selecionada foi a t-Student com parâmetros $(\rho_1, \rho_2, df) = (0, 5458; 0, 0896; 1, 84)$. A inferência foi baseada na pseudo máxima verossimilhança.

Portanto, a taxa do prêmio do seguro de faturamento para os dois municípios será a mesma, pois os parâmetros das cópulas são os mesmos. Isso era esperado, pois estamos trabalhando com a mesma série de preços e dólar, alterando apenas a produtividade simulada (com parâmetros muito parecidos entre os dois municípios).

O fato das séries de preço e dólar serem a mesma nos dois municípios faz com que a produtividade seja importante na diferenciação das taxas dos municípios. Observe que a taxa do risco não pode ser vista apenas como uma proporcionalidade direta do risco. Por exemplo, o desvio padrão da produtividade simulada em Toledo é aproximadamente 39% superior àquela de Cascavel, o que não se pode afirmar que a taxa em Toledo seja 39% maior do que a de Cascavel.

Tabela 4.11. Modelos selecionados para a precificação do seguro de faturamento com os parâmetros de dependência.

Dados	Cópula	Parâmetros
(Toledo, CME, Dólar)	t-Student	$(\rho_1, \rho_2, df) = (0, 5458; 0, 0896; 1, 84)$
(Cascavel, CME, Dólar)	t-Student	$(\rho_1, \rho_2, df) = (0, 5458; 0, 0896; 1, 84)$

4.6 Discussão

Considere que o produtor destes municípios contratou o seguro no dia 08/08/2016, no período em que solicitou o empréstimo bancário para custeio da safra que foi plantada entre os meses de setembro e dezembro de 2016 (21/09 à 31/12), de acordo com o Zoneamento Agrícola de Risco (ZARC) (MAPA, 2018).

Assim, o preço esperado em reais será a média aritmética dos últimos 15 dias a partir da data de contratação do seguro.

4.6.1 Contrato de Faturamento Bidimensional

A Tabela 4.12 apresenta a taxa pura e a taxa comercial (com carregamento de 20%) para o caso bivariado para a cópula de Frank ($\delta = -0,3358$), com N = número de observações.

Para $N = 454$, utilizou-se a série de produtividade simulada da distribuição OLLN com $N = 454$ observações e a série ajustada de preços em reais, em que (Y_e) foi calculado pela média aritmética dos últimos 15 dias a partir de 08/08/2016 (data da contratação do seguro).

Por outro lado, para $N = 20.000$, considerou-se 20.000 observações simuladas da cópula ajustada das séries ajustadas. Além disso, para o preço esperado Y_e , utilizou-se a média do vetor de preços simulados da cópula com 20.000 observações.

Observe que a taxa pura simulada (para $N = 20.000$) do prêmio de seguro com NC de 60% e 70% correspondem a 4,682 e 7,115, respectivamente. Considerando a taxa comercial para o nível de cobertura de 60% e 70% é 9,365 e 14,230, respectivamente.

Observe que quanto maior o nível de cobertura, maior será a taxa do prêmio de seguro.

Tabela 4.12. Taxa Pura e Comercial para o prêmio do seguro de faturamento bidimensional para os municípios de Toledo e Cascavel utilizando a cópula Frank ($\delta = -0,3358$).

N	NC	60%	65%	70%	75%	80%
454	Taxa Pura	4,499	5,386	6,446	7,653	8,950
	Taxa Comercial	8,999	10,773	12,892	15,306	17,901
20 000	Taxa Pura	4,682	5,842	7,115	8,483	9,966
	Taxa Comercial	9,365	11,684	14,230	16,967	19,932

A Tabela 4.13 apresenta a taxa comercial média do prêmio do seguro de faturamento utilizando a cópula selecionada com $N = 454$ observações e as taxas da seguradora A, para os níveis de cobertura 60 – 69% e 70 – 79%. Observe que a taxa calculada neste trabalho está bem acima daquelas aplicadas pelo mercado segurador. Por exemplo, a taxa da seguradora A é 63,32 % e 40,28% da taxa calculada pela cópula para NC de 60 – 69% e 60 – 69%, respectivamente.

Portanto, há indícios de uma subestimação por parte da seguradora A, o que pode levar a uma perda grande para a seguradora, pois esta pode estar considerando um risco muito menor do que deveria ser levado em consideração.

Tabela 4.13. Taxa Comercial média para o seguro de faturamento usando a cópula e a da seguradora A para Toledo (Cascavel).

Modelo	60 – 69%	70 – 79%
Cópula	9,886	14,099
Seguradora A	6,26	5,68(5,52)

4.6.2 Contrato de Faturamento Tridimensional

Sob o enfoque tridimensional, a cópula selecionada para os dois municípios é a t-Student com parâmetros $(\rho_1, \rho_2, df) = (0,5458; 0,0896; 1,84)$.

A Tabela 4.14 apresenta a taxa pura do prêmio do seguro de faturamento e a taxa comercial com 20% de taxa de carregamento e nível de cobertura 60%, 65%, 70%, 75% e 80% para $N=454$ (preços ajustados) e $N = 20.000$ observações simuladas da cópula selecionada. Observe que para NC = 65% e NC = 80%, a taxa pura do prêmio é, respectivamente, 1,360 e 3,230.

Comparando os resultados das Tabelas 4.12 e 4.14, para todos os níveis de cobertura, a taxa calculada pela cópula bivariada está superestimada quando comparada com a cópula tridimensional.

Além disso, quando se considera o câmbio como variável aleatória, a taxa calculada pelo enfoque tridimensional (NC=80%) é quase três vezes menor do que aquela calculada sob o enfoque bidimensional, no qual não se leva em consideração a variabilidade do dólar.

Portanto, quando se acrescenta a variável dólar na modelagem, há uma diminuição da taxa do prêmio de seguro de faturamento. Essa diminuição pode estar relacionada à incorporação da correlação negativa entre o preço (US\$) e a taxa do dólar no modelo, o que pode estar levando à redução do risco entre essas variáveis. Em outras palavras, a dependência negativa que existe entre os preços e o dólar pode estar reduzindo o risco destas variáveis na modelagem.

A Tabela 4.15 apresenta a taxa comercial para nível de cobertura 60 – 69% e 70 – 79% calculados pela cópula tridimensional e pela seguradora A. Para ambos os intervalos do nível de cobertura, a seguradora A está superestimando a taxa do seguro quando comparada com a metodologia de cópulas.

A superestimação da taxa pode dificultar a massificação do seguro em todo o território brasileiro, além de atrair produtores com maior risco de indenização, e assim, aumentar o problema da seleção adversa.

Tabela 4.14. Taxas pura e comercial (20%) de carregamento para o seguro de faturamento agrícola utilizando a copula t-Student para (Produtividade Toledo, Preço CME, Dólar)

	N	60%	65%	70%	75%	80%
N=454	Taxa Pura	0,955	1,410	1,937	2,589	3,378
	Taxa Comercial	1,911	2,820	3,874	5,178	6,755
N=20000	Taxa Pura	0,947	1,360	1,869	2,486	3,230
	Taxa Comercial	1,894	2,719	3,738	4,973	6,459

Tabela 4.15. Taxas médias comerciais para o seguro de faturamento agrícola utilizando a copula t-Student e a seguradora A.

NC	Cópula	Seguradora A
60 – 69%	2,36	6,26
70 – 79%	4,52	5,68(5,52)

4.7 Conclusão

No mercado segurador brasileiro, existem apenas duas empresas seguradoras que oferecem o seguro de faturamento agrícola. A companhia de seguros A leva em consideração a série de produtividades anuais municipais e o preço físico (indicador CEPEA/ESALQ) na precificação do seu produto. Por outro lado, a seguradora B se diferencia da primeira por utilizar a série de preços futuros negociados na CME, já transformada em reais.

Por isso, neste trabalho, métodos alternativos foram propostos para o cálculo da taxa do prêmio do seguro de faturamento. O método de cópulas foi abordado sob dois enfoques: bidimensional e tridimensional. O enfoque bidimensional leva em consideração o risco da produtividade e do preço em reais. Por outro lado, o enfoque tridimensional leva em consideração as variáveis produtividade, preço em dólares e a taxa do câmbio. Além disso, no contexto de risco agrícola, não foi encontrado na literatura estudos relacionando essas três variáveis.

Neste trabalho utilizou-se as séries de produtividades anuais de soja para os municípios de Toledo e Cascavel, os preços de soja negociados na CBOT/CME com vencimento de março de 2017 e a cotação do dólar neste mesmo período. A produtividade foi simulada utilizando a distribuição OLLN e as séries de preços e dólar foram ajustadas pelo modelos GARCH(1,1) com erros t-Student para modelagem da volatilidade.

Sob o enfoque bidimensional, a estrutura de dependência que melhor se ajustou aos dados foi a cópula Frank com o parâmetro de dependência negativo ($\delta = -0,3358$). Os resultados sugerem que o mercado segurador subestima a taxa do prêmio do seguro quando comparado com a metodologia de cópula. A subestimação da taxa leva a seguradora a correr riscos de perdas, o que pode levar a consideradas perdas financeiras. Além disso, as seguradoras levam em consideração um risco menor do que deveria ser levado em consideração no momento da precificação do produto.

Sob o enfoque tridimensional, a cópula que melhor se ajustou aos dados foi a t-Student com parâmetros $(\rho_1, \rho_2, df) = (0,5458; 0,0896; 1,84)$. A taxa calculada sob este enfoque é menor do que aquela apresentada sob o enfoque bidimensional, pois neste caso, levou-se em consideração a variável aleatória taxa do dólar na modelagem. Com isso, há a incorporação de uma variável exógena ao modelo, mas que tem grande influência no preço da soja no contexto brasileiro. Essa diminuição pode estar relacionada à incorporação da dependência negativa entre o preço (US\$) e a taxa do dólar no modelo, o que pode levar à redução do risco entre essas variáveis.

Além disso, os resultados sugerem que as taxas aplicadas pelas seguradoras podem estar superfaturadas quando comparadas com o modelo de cópula tridimensional. Este superfaturamento pode atrair produtores de alto risco, aumentando o problema da seleção adversa além de dificultar a massificação do

produto no território brasileiro.

Referências

- AKAIKE, H., 1974 A new look at the statistical model identification. *Automatic Control, IEEE Transactions on* **19**: 716–723.
- BARCHART, 2017 Advanced commodities service. Disponível em : <http://acs.barchart.com/> Acessado em: 05/02/2017.
- BOLLERSLEV, T., 1986 Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics* **31**: 307–327.
- BRISOLARA, C. S., 2013 *Proposições para o desenvolvimento do seguro de receita agrícola no Brasil: do modelo teórico ao cálculo das taxas de prêmio*. Ph.D. thesis, Universidade de São Paulo.
- CHEN, X. and Y. FAN, 2006 Estimation of copula-based semiparametric time series models. *Journal of Econometrics* **130**: 307–335.
- CHERUBINI, U., E. LUCIANO, and W. VECCHIATO, 2004 *Copula methods in finance*. John Wiley & Sons.
- DUARTE, G. V., A. BRAGA, D. L. MIQUELLUTI, and V. A. OZAKI, 2017 Modeling of soybean yield using symmetric, asymmetric and bimodal distributions: implications for crop insurance. *Journal of Applied Statistics* pp. 1–18.
- ENGLE, R. F., 1982 Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society* pp. 987–1007.
- FISHER, N. and P. SWITZER, 2001 Graphical assessment of dependence: Is a picture worth 100 tests? *The American Statistician* **55**: 233–239.
- GALLAGHER, P., 1987 Us soybean yields: estimation and forecasting with nonsymmetric disturbances. *American Journal of Agricultural Economics* **69**: 796–803.
- GENEST, C. and A.-C. FAVRE, 2007 Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of hydrologic engineering* **12**: 347–368.
- GENEST, C., K. GHOUDI, and L.-P. RIVEST, 1995 A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* **82**: 543–552.
- GENEST, C. and B. RÉMILLARD, 2004 Test of independence and randomness based on the empirical copula process. *Test* **13**: 335–369.
- GOODWIN, B. K. and O. MAHUL, 2004 Risk modeling concepts relating to the design and rating of agricultural insurance contracts. Disponível em: <http://elibrary.worldbank.org/doi/book/10.1596/1813-9450-3392> Acessado em : 25 junho de 2016.
- IPARDES, 2017 Instituto paranaense de desenvolvimento econômico e social. base de dados do estado (bdweb). Acessado em: 20/dezembro/2017.
- IPEA, 2017 Câmbio - séries históricas- ideadata. Disponível em: <http://www.ipeadata.gov.br/Default.aspx> Acessado em: 05/02/2017.
- JOE, H., 1997 *Multivariate models and multivariate dependence concepts*. CRC Press.

- JOE, H., 2014 *Dependence modeling with copulas*. CRC Press.
- JOE, H. and J. J. XU, 1996 The estimation method of inference functions for margins for multivariate models. Technical report, Technical report.
- KOJADINOVIC, I. and J. YAN, 2011 Tests of serial independence for continuous multivariate time series based on a möbius decomposition of the independence empirical copula process. *Annals of the Institute of Statistical Mathematics* **63**: 347–373.
- MAPA, 2017 Atlas do seguro rural. indicador das taxas. Acessado em 21 de fevereiro de 2017.
- MAPA, 2018 Zoneamento agrícola de risco (zarc). Disponível em: <http://indicadores.agricultura.gov.br/zarc/index.htm> Acessado em 22/01/2018., Disponível em: <http://indicadores.agricultura.gov.br/zarc/index.htm> Acessado em 22/01/2018.
- MIQUELETO, G. J., 2011 *Contribuições para o desenvolvimento do seguro agrícola de renda para o Brasil: evidências teóricas e empíricas..* Ph.D. thesis, Escola Superior de Agricultura "Luiz de Queiroz -Universidade de São Paulo.
- MORETTIN, P. A., 2008 *Econometria financeira: um curso em séries temporais financeiras*. Edgard Blücher.
- NELSON, D. B., 1991 Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society* pp. 347–370.
- SCHWARZ, G. *et al.*, 1978 Estimating the dimension of a model. *The annals of statistics* **6**: 461–464.
- SKLAR, M., 1959 *Fonctions de répartition à n dimensions et leurs marges*. Université Paris 8.
- TEJEDA, H. A., B. K. GOODWIN, *et al.*, 2008 Modeling crop prices through a burr distribution and analysis of correlation between crop prices and yields using a copula method. In *annual meeting of the Agricultural and Applied Economics Association, Orlando, FL, USA*, volume 83, pp. 643–649, Citeseer.
- TURSUNALIEVA, A. and P. SILVAPULLE, 2007 Assessing and modeling the changes in dependence between exchange rates. Disponível em: https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=ESAM07&paper_id=130.

ANEXO

Anexo A -Estimação por Máxima Verossimilhança

Para estimar os parâmetros das distribuições apresentadas na seção 2.5, utilizamos o método de máxima verossimilhança. Este apêndice descreve a inferência estatística usada neste trabalho para os modelos Skew -t Student (ST), sendo esse procedimento o mesmo para as distribuições remanescentes. Considere $y_1, \dots, y_n$, uma amostra aleatória de tamanho n da distribuição ST. Então, o logaritmo da função de verossimilhança para o vetor dos parâmetros $\boldsymbol{\theta} = (\mu, \sigma, \nu, \tau)^T$, é dado por:

$$l(\boldsymbol{\theta}) = \sum_{i=1}^n \log \left\{ \frac{2}{\tau + \frac{1}{\tau}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \sigma(\pi\nu)^{\frac{1}{2}}} \times \left[\left(1 + \frac{z^2}{\nu}\right) \frac{1}{\tau^2} \right]^{-\frac{(\nu+1)}{2}} \right\} \quad (4.17)$$

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

As estimativas de máxima verossimilhança $\hat{\boldsymbol{\theta}}$ do vetor de parâmetros podem ser obtidas ao maximizar a Equação 4.17. Nesta etapa, utilizamos os métodos "Nelder-Mead" e "L-BFGS B", fornecidos na função "Optim" do *software* R, (R CORE TEAM, 2016), no qual os valores iniciais para estimação dos parâmetros podem ser obtidos, primeiro, estimando os parâmetros para o modelo t-Student padrão e com estes valores são estimados os parâmetros do modelo ST.

Além disso, a função "Optim" fornece a opção "gr" na qual o usuário pode fornecer o vetor score, o que torna o algoritmo mais eficiente. Nesse caso, os componentes do vetor score $U(\boldsymbol{\theta})$ são dados por:

$$\begin{aligned} U_{\nu}(\boldsymbol{\theta}) &= \sum_{i=1}^n \left\{ \frac{\psi\left(\frac{\nu+1}{2}\right) \sqrt{\pi\nu}\tau^2 + \Gamma\left(\frac{\nu}{2}\right) [\psi\left(\frac{\nu+1}{2}\right) \pi\nu - \psi\left(\frac{\nu}{2}\right) \pi\nu - \pi]}{2\sqrt{\pi\nu}} \right\} \\ &\quad + \sum_{i=1}^n \left\{ -\frac{1}{2} \ln \left[\frac{\nu\sigma^2 + (y_i - \mu)^2}{\nu\sigma^2\tau^2} \right] + \frac{(\nu+1) + (y_i - \mu)^2}{2\nu[\nu\sigma^2 + (y_i - \mu)^2]} \right\}, \\ U_{\mu}(\boldsymbol{\theta}) &= \sum_{i=1}^n \left\{ \frac{\left(\frac{\nu+1}{2}\right) (-2y_i + 2\mu)}{\nu\sigma^2 + (y_i - \mu)^2} \right\}, \\ U_{\tau}(\boldsymbol{\theta}) &= \sum_{i=1}^n \left\{ \frac{-\tau^2 + \Gamma\left(\frac{\nu}{2}\right) \sqrt{\pi\nu} - (\nu+1) [\tau^2 + \Gamma\left(\frac{\nu}{2}\right) \sqrt{\pi\nu}]}{\tau(\tau^2 + \Gamma\left(\frac{\nu}{2}\right) \sqrt{\pi\nu})} \right\}, \\ U_{\sigma}(\boldsymbol{\theta}) &= \sum_{i=1}^n \left\{ \frac{1}{\sigma} - \frac{(\nu+1)(y_i - \mu)^2}{\sigma[\nu\sigma^2 + (y_i - \mu)^2]} \right\}. \end{aligned}$$

Igualando estas equações a zero e resolvendo-as, ao mesmo tempo, obtemos as estimativas de máxima verossimilhança (EMV) dos parâmetros, usando métodos numéricos.

Sob certas condições de regularidade, o vetor de parâmetro $\boldsymbol{\theta}$, no seu espaço de parâmetros, tem a distribuição assintótica $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$, Normal multivariada $N_4(0, K(\boldsymbol{\theta})^{-1})$, em que $K(\boldsymbol{\theta})$ é a matriz de informação esperada.

A matriz $K(\boldsymbol{\theta})^{-1}$ de covariâncias assintóticas de $\hat{\boldsymbol{\theta}}$ pode ser aproximada pela matriz inversa de informações observada $(4) \times (4)$, $-\ddot{\mathbf{L}}(\boldsymbol{\theta})$. Assim, os EMV podem fornecer regiões de confiança usando a normalidade assintótica.

Então, as inferências assintóticas para o vetor dos parâmetros $\boldsymbol{\theta}$ podem ser realizadas usando a aproximação Normal $N_4(0, -\ddot{\mathbf{L}}(\boldsymbol{\theta})^{-1})$ para $\hat{\boldsymbol{\theta}}$ e erros padrão para os EMV podem ser obtidos a partir da raiz quadrada dos elementos na diagonal principal da matriz de informação observada inversa, e pode-se

formular hipóteses a serem testadas. Os elementos da matriz de informação observada são dados por:

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

para $i = 1, \dots, I$, que são obtidos numericamente. A distribuição Normal assintótica $N_4(0, -\ddot{\mathbf{L}}(\boldsymbol{\theta})^{-1})$ pode ser usada para criar regiões aproximadas de confiança para o vetor de parâmetros $\boldsymbol{\theta}$.

A normalidade assintótica também é útil para testar a qualidade do ajuste de alguns submodelos. As estatísticas da razão de verossimilhança (RV) podem ser usadas para comparar as distribuições ST $(\mu, \sigma, \lambda, \alpha)$ com os submodelos Normal, t-student e Cauchy.

Então, obtemos o valor da estatística w da função log-verossimilhança do modelo restrito sob o modelo irrestrito ainda por testar. Em alguns casos, a hipótese do tipo $H_0 : \boldsymbol{\psi} = \boldsymbol{\psi}_0$ versus $H : \boldsymbol{\psi} \neq \boldsymbol{\psi}_0$ é testada, em que $\boldsymbol{\psi}$ é o vetor formado com alguns componentes de $\boldsymbol{\theta}$ e $\boldsymbol{\psi}_0$ é um vetor específico.

Por exemplo, a hipótese $H_0 : \tau = 1$ versus $H : H_0$ não é verdadeira é equivalente a comparar ST (μ, σ, τ, ν) para a distribuição t-Student e a estatística W é obtida por:

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

em que $\hat{\alpha}$, $\hat{\mu}$ e $\hat{\sigma}$ são os EMV sob H e $\tilde{\mu}$ e $\tilde{\sigma}$ são as estimativas sob H_0 .

APÊNDICE

Apêndice I- Script da distribuição OLLN para o software R.

Este apêndice apresenta o script do programa em R para o cálculo das funções de densidade (pdf_olln) e de distribuição acumulada (cdf_olln) do modelo OLLN.

```
pdf_olln=function(par,x){
  mu = par[1]
  sigma = par[2]
  alpha = par[3]
  z = (x - mu)/sigma
  f= dnorm(z, mean =0, sd = 1)
  F = pnorm(z)
  num= alpha*f*( F*(1-F))^(alpha-1)
  den = (sigma)*(F^(alpha) + (1-F)^(alpha))^2
  num/den
}

cdf_olln=function(par,x){
  mu = par[1]
  sigma = par[2]
  alpha = par[3]
  z = (x - mu)/sigma
  G = pnorm(z)
  (G^(alpha))/((G^(alpha))+(1-G)^(alpha))
}

require(AdequacyModel)
aj=goodness.fit(pdf=pdf_olln,cdf=cdf_olln, starts=c(mean(y),sd(y),1.2),data=y, method="Nelder-
Mead", domain=c(0,Inf))
```