

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO, CONTABILIDADE E ATUÁRIA
DEPARTAMENTO DE CONTABILIDADE E ATUÁRIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CONTROLADORIA E CONTABILIDADE

Rafael Tonet Rensi

De Basileia para o Campo: estimando a estrutura de dependência de *default* em portfólios de crédito rural por meio de Cópulas

São Paulo
2022

Prof. Dr. Carlos Gilberto Carlotti Júnior
Reitor da Universidade de São Paulo

Profa. Dra. Maria Dolores Montoya Diaz
Diretora da Faculdade de Economia, Administração, Contabilidade e Atuária

Prof. Dr. Valmor Slomski
Chefe do Departamento de Contabilidade e Atuária

Prof. Dr. Renê Coppe Pimentel
Coordenador do Programa de Pós-Graduação em Controladoria e Contabilidade

Rafael Tonet Rensi

De Basileia para o Campo: estimando a estrutura de dependência de *default* em portfólios de crédito rural por meio de Cópulas

Dissertação apresentada ao Programa de Pós-Graduação em Controladoria e Contabilidade do Departamento de Contabilidade e Atuária da Faculdade de Economia, Administração, Contabilidade e Atuária da Universidade de São Paulo como requisito parcial para obtenção do título Mestre em Ciências.

Orientador: Prof. Dr. João Vinícius de França Carvalho

Versão Corrigida

(versão original disponível na Biblioteca da Faculdade de Economia, Administração e Contabilidade)

São Paulo

2022

Catálogo na Publicação (CIP)
Ficha Catalográfica com dados inseridos pelo autor

Rensi, Rafael Tonet

De Basileia para o Campo: estimando a estrutura de dependência de *default* em portfólios de crédito rural por meio de Cópulas / Rafael Tonet Rensi – São Paulo, 2022.

141p.

Dissertação (Mestrado) – Universidade de São Paulo, 2022.

Orientador: Prof. Dr. João Vinícius de França Carvalho.

1.Risco de Crédito. 2. Crédito Rural. 3. Probabilidade de Default.
4.Cópulas 5. Estruturas de Dependência de Funções Conjuntas.
Universidade de São Paulo. Faculdade de Economia, Administração,
Contabilidade e Atuária

Rafael Tonet Rensi

De Basileia para o Campo: estimando a estrutura de dependência de *default* em portfólios de crédito rural por meio de Cópulas.

Dissertação apresentada à Faculdade de Economia, Administração, Contabilidade e Atuária da Universidade de São Paulo para obtenção do título de Mestre em Ciências.

Aprovado em: 15 de junho de 2022

Banca Examinadora

Prof^a. Dra. Chang Chiann

Instituição: IME – USP

Julgamento: Aprovado

Prof. Dr. Luiz Paulo Lopes Fávero

Instituição: FEA – USP

Julgamento: Aprovado

Prof. Dr. Wesley Mendes da Silva

Instituição: FGV – SP – Externo

Julgamento: Aprovado

Agradecimentos

Agradeço à minha mãe Helena pelo exemplo de dedicação e busca por conhecimento, pela educação que me forneceu formal e informal e por ter me apresentado o mundo em todo o seu esplendor.

À minha irmã Lili que me inspira a ser uma pessoa melhor e desbravou diversos caminhos, pavimentando-os para facilitar a minha passagem.

Ao meu pai Rensi, por ter me ensinado pelo exemplo a ter empatia.

Aos meus tios Tereza e Júnior, que sempre vibraram com as minhas conquistas e acreditam em todo o meu potencial.

À minha tia Malu, pelos profundos conselhos que me dá.

À Beatriz, pelo companheirismo durante toda esta jornada.

Aos meus amigos e colegas de trabalho, que de alguma forma ao longo desta jornada me estenderam a mão nos momentos em que precisei.

Aos gestores de risco Mário e Gláucio, por acreditarem neste projeto de pesquisa. Sem o apoio e patrocínio de vocês isto não seria possível.

Por fim, ao meu orientador e amigo João Vinícius, pela relação próxima e humana. Obrigado pelos ensinamentos. Sob a sua supervisão aprendi e amadureci muito nestes dois anos, me tornando capaz de abordar temas que antes eram demasiadamente complexos. Ressalto todo o apoio e esforço despendido para a finalização deste trabalho. A sua atuação como docente é um exemplo e inspiração.

*Every step that I have taken in my life
has brought me here, now
Milano, 2019*

De Basileia para o Campo: estimando a estrutura de dependência de *default* em portfólios de crédito rural por meio de Cópulas

O crédito rural é fundamental ao desenvolvimento da agricultura, sendo um instrumento catalisador do objetivo humanitário de erradicação da fome e da miséria. Em países com vocação para o agronegócio, como o Brasil, o financiamento dos produtores rurais é estratégico e tema de políticas públicas. A melhora na qualidade da concessão de crédito rural, e de seu gerenciamento de riscos, proporciona ganhos sociais e econômicos. Apesar da relevância deste tipo de financiamento, há grande escassez na literatura de modelos de risco de crédito específicos ao segmento rural, incluindo a regulação do Acordo de Basileia. Sob a hipótese de que o risco de um portfólio de crédito rural possa apresentar aumento da dependência entre *defaults* individuais em cenários extremos, provocados por desastres naturais, e que esta é uma característica idiossincrática deste tipo de crédito não captada em modelos de risco de crédito usuais, o objetivo desta pesquisa é identificar qual é a melhor forma de se representar a distribuição de probabilidade de *default* de um portfólio de crédito rural, considerando: (i) o pressuposto da normalidade adotado por modelos de risco de crédito tradicionais, como o proposto por Basileia, e; (ii) distribuições de probabilidade assimétricas, havendo interesse no comportamento das distribuições com dependência de cauda. A partir de 74.964.734 observações de operações de crédito de um grande *player* do mercado de crédito rural, foram construídas 60 observações longitudinais de sete linhas de crédito rural. As estruturas de dependência dessas séries foram estimadas por dez diferentes famílias de cópulas. Cópulas são funções estatísticas que generalizam o conceito de dependência, para além das tradicionais medidas-resumo. Para cada par de linha de crédito rural, a melhor representação de sua distribuição de probabilidade de *default* foi verificada pela cópula com menor erro quadrático médio dos quantis em relação à sua respectiva distribuição empírica. Os resultados indicam que, de forma geral, as distribuições geradas por meio de cópulas assimétricas ajustam-se melhor aos dados reais, em comparação à distribuição Normal bivariada, havendo destaque para a cópula de Clayton. Contraintuitivamente, a performance das cópulas foi melhor quando avaliamos a aderência aos dados considerando todo o domínio da distribuição empírica do que quando observamos somente os quantis superiores, região onde as cópulas de valores extremos teriam maior potencial de identificação de dependência. Por fim, embora os experimentos não tenham identificado plena supremacia de uma abordagem (cópulas ou simetria geral ou de Basileia) em detrimento de qualquer outra, restou evidente a fragilidade da premissa de Normalidade da distribuição de probabilidade de *default* conjunta. Assim, os resultados obtidos não apenas contribuem para o argumento da necessidade de refinamento de modelos de risco de crédito voltados ao segmento rural, como principalmente sugerem que as cópulas são relevantes instrumentos para o desenvolvimento destes modelos. Os principais agentes econômicos a se beneficiarem deste refinamento são: (i) *credores rurais*, devido a uma maior eficiência no processo de concessão de crédito; (ii) *seguradoras*, dada a característica idiossincrática do risco de crédito rural de exposição a desastres naturais serem em boa parte absorvidos por essas companhias; (iii) *reguladores* (Banco Central e Susep), pela redução potencial do risco sistêmico oriunda de modelos de risco mais precisos de seus regulados, e; (iv) ao *Setor Público*, pela melhor eficiência no processo de concessão de crédito rural por ele subsidiado, além do auxílio na promoção dos objetivos humanitários relacionados à agricultura.

Risco de Crédito; Crédito Rural; Probabilidade de Default; Cópulas; Estruturas de Dependência de Funções Conjuntas.

From Basel to the Countryside: estimating the default dependence structure in rural credit portfolios through Copulas

Rural credit is fundamental to the development of agriculture, being a catalyst for the humanitarian objective of eradicating hunger and misery. In countries with vocation for agribusiness, such as Brazil, the financing of rural producers is strategic and a subject of public policies. The improvement in the quality of rural credit concession and risk management provides social and economic gains. Although the relevance of this type of credit, there are few previous studies focused on rural credit, including the regulation of the Basel Accord. We assume the hypothesis that a rural credit portfolio risk may increase the dependence between individual *defaults* in extreme scenarios, caused by natural disasters, being an idiosyncratic characteristic of this type of credit, not considered in usual credit risk models. This, the objective of this work is to identify the best way to represent the default probability distribution of a rural credit portfolio, assuming: (i) the assumption of normality adopted by traditional credit risk models, such as the one proposed by Basel, and; (ii) asymmetric probability distributions, with interest in the behavior of tail-dependent distributions. From 74.964.734 observations of credit operations of a major player in the Brazilian rural credit market, we constructed 60 longitudinal observations of seven rural credit lines. The dependence structures of these series were estimated for ten different copulas families. Copulas are functions that generalize the concept of dependence, beyond the traditional summary measures. For each pair of rural credit line, the best representation of its default probability distribution was verified by the copula with the smallest mean squared error of the quantiles in relation to their respective empirical distribution. The results indicate that, in general, the distributions generated through asymmetric copulas fit better to the real data, compared to the bivariate Normal distribution, with the Clayton copula stands out. Counterintuitively, the performance of copulas was better when we evaluated adherence considering the entire domain of the empirical distribution than when we observed only the upper quantiles, a region where copulas with extreme values would have greater potential for identifying dependence. Finally, although the experiments did not identify the full supremacy of one approach (copulas or general symmetry or Basel requirement) to the detriment of any other, the fragility of the Normal assumption of the joint default probability distribution remained evident. Thus, the results obtained not only contribute to the argument of the need to refine credit risk models aimed at the rural segment, but also suggest that copulas are relevant instruments for the development of these models. The main economic agents interested in this refinement are: (i) rural creditors, due to greater efficiency in the credit granting process; (ii) insurance companies, given the idiosyncratic nature of rural credit risk of exposure to natural disasters being largely absorbed by these companies; (iii) regulators (Banco Central do Brasil and Susep), due to the potential reduction of systemic risk arising from more accurate risk models of their regulated entities, and; (iv) the Public Sector, for the better efficiency in the process of granting rural credit subsidized by it, and the effect in promoting humanitarian goals related to agriculture.

Credit risk; Rural credit; Probability of Default; Copulas; Dependence Structures for Joint Distributions.

LISTA DE FIGURAS

Figura 1. Classificação dos modelos de PD aplicados ao crédito rural.....	29
Figura 2. Taxa de default mensal - Aquisição de Ativos Financeiros.....	58
Figura 3. Taxa de default mensal - Linhas de Crédito Rural e SFN	61
Figura 4. Representação conjunta das taxas de default das linhas de crédito rural avaliadas .	64
Figura 5. Teste <i>GoF</i> híbrido $PD(A \times B)$	68
Figura 6. Densidade de probabilidade conjunta das cópulas de maior aderência aos pares de PD	71
Figura 7. Distribuições de frequência e de probabilidade acumulada $PD(A)$, $PD(B)$	73
Figura 8. Função de Distribuição Acumulada – Observações Empíricas e Dados Simulados sob a Distribuição Normal e Cópulas	74
Figura 9. Teste <i>GoF</i> híbrido $PD(A \times C)$	101
Figura 10. Teste <i>GoF</i> híbrido $PD(A \times D)$	104
Figura 11. Teste <i>GoF</i> híbrido $PD(A \times E)$	106
Figura 12. Teste <i>GoF</i> híbrido $PD(A \times F *)$	109
Figura 13. Teste <i>GoF</i> híbrido $PD(A \times G)$	112
Figura 14. Teste <i>GoF</i> híbrido $PD(B \times E)$	115

LISTA DE TABELAS

Tabela 1. Resumo da bibliografia de modelos de PD de crédito rural a nível individual.	29
Tabela 2. Ponderação de Risco dos Ativos – Acordo de Basileia I.....	34
Tabela 3. Fontes de estimação de parâmetros para IRB – Acordo de Basileia II.....	38
Tabela 4. Cópulas Elípticas Bivariadas - Funções de distribuição acumulada.....	45
Tabela 5. Cópulas Elípticas Bivariadas: medidas de dependência de Kendall, Spearman e índices de cauda.....	46
Tabela 6. Cópula de Plackett Bivariada – Função de distribuição acumulada, medidas de dependência e medidas de cauda	46
Tabela 7. Cópulas Arquimedianas Bivariadas: Funções de distribuição acumulada	47
Tabela 8. Cópulas Arquimedianas Bivariadas: Medidas de dependência de Kendall, Spearman e de Cauda	47
Tabela 9. Cópulas de valores extremos - Função geradora e função de distribuição acumulada	48
Tabela 10. Cópulas de valores extremos – índices de cauda.....	49
Tabela 11. Resumo da bibliografia de cópulas aplicadas ao risco de crédito.	54
Tabela 12. Informações disponíveis das carteiras de Crédito Rural – Descrição das variáveis.	55
Tabela 13. Rótulos e quantidade de observações das linhas de crédito rural	56
Tabela 14. Estatísticas descritivas - saldo contábil (em BRL) por linha de crédito	56
Tabela 15. Estatísticas descritivas – Taxa de Default por linhas de crédito rural e SFN	57
Tabela 16. Observações e Estatísticas Descritivas – Aquisição de Ativos Financeiros.....	59
Tabela 17. Matrizes de correlações – Coeficientes de Pearson, Kendall e Spearman.....	60
Tabela 18. Assimetria e Curtose – taxas de default por linha de crédito e SFN	62
Tabela 19. Testes de Normalidade – taxas de <i>default</i> por linha de crédito e SFN	63
Tabela 20. Classes e Famílias de cópulas e suas respectivas estruturas de dependência	66
Tabela 21. Testes Goodness-of-Fit agregados pelo GoF Híbrido para cada classe de cópula.....	67
Tabela 22. PD Conjunta – Cópulas de maior aderência segundo testes <i>GoFhb</i> e MME.....	70
Tabela 23. PD Conjunta – Parâmetros da distribuição Normal Bivariada e das cópulas de maior aderência.....	72
Tabela 24. Erro Quadrático Médio – Simulações de normalidade e via Cópulas	77
Tabela 25. Simulações de maior aderência aos dados observados.	79
Tabela 26. Simulações de normalidade e via Cópulas – Diretrizes de Basileia.....	80
Tabela 27. Simulações de normalidade e via Cópulas – Todas as linhas de crédito.....	81
Tabela 28. p-valor dos testes <i>GoF</i> e estatística dos testes <i>MME - PD(A × B)</i>	100
Tabela 29. p-valor dos testes <i>GoF</i> e estatística dos testes <i>MME - PD(A × C)</i>	103
Tabela 30. p-valor dos testes <i>GoF</i> e estatística dos testes <i>MME - PD(A × D)</i>	105
Tabela 31. p-valor dos testes <i>GoF</i> e estatística dos testes <i>MME - PD(A × E)</i>	108
Tabela 32. p-valor dos testes <i>GoF</i> e estatística dos testes <i>MME - PD(A × F *)</i>	110
Tabela 33. p-valor dos testes <i>GoF</i> e estatística dos testes <i>MME - PD(A × G)</i>	113
Tabela 34. p-valor dos testes <i>GoF</i> e estatística dos testes <i>MME - PD(B × E)</i>	116

LISTA DE ABREVIACOES E ACRONIMOS

ANN	<i>Artificial Neural Networks</i>
BIS	<i>Bank of International Settlements</i>
CAR	<i>Capital Adequacy Ratio</i>
CML	<i>Canonical Maximum Likelihood</i>
dTree	<i>Decision Trees</i>
EAD	Exposure at default
EL	<i>Expected loss</i>
EV	<i>Extreme Value</i> - Copulas de valores extremos
GoF	Teste <i>Goodness-of-Fit</i>
GP	<i>Goal Programming</i>
Ipea	Instituto de Pesquisa Econmica Aplicada
IRB	Internal Ratings-Based Approach
kNN	<i>k-Nearest Neighbors</i>
LGD	<i>Loss given default</i>
M	Maturity
MCDA	<i>Multiple criteria decision analysis</i>
MLE	<i>Maximum Likelihood Estimation</i>
MME	Mtodo dos momentos
MPLE	<i>Maximum Pseudo-Likelihood</i>
MtM	<i>Mark to Market</i>
ONU	Organizao das Naes Unidas
PD	Probabilidade de <i>default</i>
RPA	Recursive Partitioning Algorithm
RWA	<i>Risk-weighted assets</i>
S&P	<i>Standard & Poor's</i>
SA	<i>Standardized Approach</i>
SCSRF	<i>Syncretic Cost-Sensitive Random Forest</i>
SFN	Sistema Financeiro Nacional
SGS	Sistema Gerenciador de Sries Temporais
SVM	<i>Support Vector Machines</i>
TL	Total loss
UL	Unexpected loss

SUMÁRIO

1. Introdução.....	16
2. Risco de produção agrícola e desastres naturais.....	19
2.1 Introdução	19
2.2. Risco de produção e desastres naturais.....	19
3. Modelos de PD de crédito rural a nível individual.....	22
3.1. Modelos <i>Credit Scoring</i> aplicados ao crédito rural	23
3.1.1. Modelos paramétricos de <i>Credit Scoring</i>	23
3.1.2. Modelos não paramétricos de <i>Credit Scoring</i>	25
3.2. Modelos de forma reduzida aplicados ao crédito rural.....	27
3.3. Considerações finais sobre os modelos de PD de crédito rural	29
4. Modelo de PD de crédito rural em nível de portfólio – uma perspectiva regulatória	32
4.1. Introdução	32
4.2. Bank of International Settlements e o Capital Regulatório	32
4.3. Acordo de Basileia I	33
4.3. Acordo de Basileia II	36
4.3.1. Abordagem Padronizada (SA)	37
4.3.2 Abordagem por modelos internos (IRB)	37
4.3.3 Função de ponderação de risco da abordagem IRB.....	38
5. Modelos de Cópulas: aspectos teóricos e aplicações.....	41
5.1. Definições e Propriedades Básicas das Cópulas.....	41
5.2. Medidas de Dependência.....	42
5.2.1. Coeficiente ρ de Pearson	43
5.2.2. Medidas de Concordância.....	43
5.2.2.1. Coeficiente τ de Kendall	43
5.2.2.2. Coeficiente ρ de Spearman.....	44
5.2.3. Medidas de Dependência de Cauda ($\lambda u, \lambda l$).....	44
5.3. Classes e Famílias de Cópulas	45
5.3.1. Cópulas Elípticas	45
5.3.2. Cópula de Plackett	46
5.3.3. Cópulas Arquimedianas	46
5.3.4. Cópulas de Valores Extremos.....	48
5.4. Ajuste das cópulas aos dados.....	49
5.4.1. Métodos de estimação de parâmetros	49
5.4.2. Testes <i>Goodness-of-Fit</i> para escolha de cópulas	50
5.6. Aplicação de cópulas ao risco de crédito.....	51
6. Os dados e análise descritiva	55

7. Modelagem e Análises de Resultados	66
7.1. PD de Portfólio - Estimação dos parâmetros e escolha da cópula.....	66
7.2. PD de Portfólio – Simulações de Normalidade e de Cópuas	72
7.3. PD de Portfólio - Diretrizes do Acordo de Basileia.....	79
8. Considerações Finais	83
Referências	86
Anexo A – Estimação do ajuste das cópulas aos dados	99
Anexo B – Códigos do Software R	118

1. Introdução

A gênese da agricultura como forma de produção de alimentos é um dos eventos mais importantes da história da civilização. Registros arqueológicos indicam que a sedentarização de populações viabilizada pela agricultura ocorreu no período entre 10.000 a.C a 5.000 a.C. (Weisdorf, 2005). A transição (parcial) da caça e da coleta para a agricultura, denominada como Revolução Neolítica, proporcionou a produção de excedentes de alimentos, o crescimento demográfico e a maior especialização do trabalho, implicações basilares ao desenvolvimento social e econômico. Tauger (2011) considera que, dentre outras atividades humanas, a agricultura é aquela que tornou a civilização possível.

Sob uma perspectiva contemporânea, o papel da agricultura é de extrema importância devido a sua atuação no combate à desigualdade social (Cervantes-Godoy & Dewbre, 2010). A segurança alimentar e o combate à fome são preocupações de países e entidades internacionais. A Agenda 2030 para o Desenvolvimento Sustentável, promulgada pela ONU em 2015, tem como um de seus pilares a erradicação da fome (Campbell, 2017; Jong & Vijge, 2021).

O plano de ação de erradicação da fome proposto pela Agenda 2030 é subdividido em 8 metas que além de tratarem do objetivo primário (findar com a fome e a má nutrição), abordam três temas auxiliares: (i) aumento da produtividade e renda dos pequenos produtores rurais; (ii) garantia de sistemas de produção sustentáveis, e; (iii) investimento em infraestrutura. O Brasil, por meio do Instituto de Pesquisa Econômica Aplicada (Ipea), adaptou as metas da Agenda 2030 à realidade doméstica (Ipea, 2018), explicitando o acesso a linhas de crédito como um dos meios de se atingir as metas de aumento da produtividade e da renda dos pequenos produtores. Esta importância dada ao crédito é tema recorrente na literatura (Bennouna & Tkiouat, 2019; Condori-Alejo et al., 2021; Kumar et al., 2013; Linh et al., 2019; Twumasi et al., 2020, 2021), a qual atribui o acesso ao crédito como instrumento de ganho econômico-social de pequenos produtores e famílias rurais.

Além do escopo dos pequenos produtores, a oferta de crédito é identificada como um fator responsável pelo aumento da produtividade rural em diversos países (Acclassato Houensou et al., 2021; Chandio et al., 2019; Guirkingner & Boucher, 2008; Hu et al., 2021; Jimi et al., 2019; Kassouri & Kacou, 2022; Li et al., 2020; Manoharan & Varkey, 2021), incluindo o Brasil (de Castro & Teixeira, 2012; S. R. F. Figueira & Rolim, 2020). Isto pode ser explicado, em parte, pela alta dependência deste setor ao financiamento, por conta de suas especificidades. A agricultura é caracterizada pelos ciclos produtivos de safras ou de amadurecimento de rebanho (Baghdasaryan et al., 2021; Katchova & Barry, 2005). Estes ciclos produtivos implicam na recorrência de investimentos de início de produção (e.g., preparo do solo e compra de sementes) e em possíveis descasamentos de fluxo de caixa (tempo entre as despesas de produção e comercialização até o recebimento da venda dos produtos). Assim, a perenidade da produção agrícola depende da capacidade dos produtores de autofinanciamento ou da disponibilidade de crédito. Além disso, com exceção da produção familiar, a agricultura é tipicamente uma indústria de capital intensivo (Barry & Robison, 2001; Weed & Hardy, 1980).

Quer seja pelo ganho econômico-social, quer seja pelo incremento na produtividade, o crédito destinado à agricultura, conhecido como crédito rural, possui papel relevante para o atingimento dos objetivos relacionados à erradicação da fome da Agenda 2030. Portanto, há um interesse social na ampliação da oferta de crédito rural (podendo inclusive justificar subsídios públicos para estas operações). Uma vez que a ampliação da oferta de crédito rural é relevante, todo o processo de disponibilização deste produto financeiro também o é, incluindo o cálculo de risco destas operações.

De forma ampla, o risco de crédito pode ser definido como o risco de um devedor não honrar com as obrigações contratuais pactuadas com o credor (BCBS, 2000), quantificado como a perda financeira em caso de *default* (não cumprimento das obrigações do devedor). Sob o aspecto regulatório, esta perda financeira pode ser mensurada como uma combinação entre a função de probabilidade de *default* (PD), da perda dado o default (LGD) e da exposição temporal ao default (EAD) (Hünseler, 2013).

O foco desta pesquisa é a mensuração da estrutura de dependência da probabilidade de *default* de diferentes carteiras de crédito rural. Apesar da constatada relevância do crédito rural para o desenvolvimento econômico e social, há escassez na literatura de modelos de PD que levem em conta as especificidades deste segmento. Em geral, a PD de uma operação de crédito rural é estimada tal como uma operação de varejo ou de dívida corporativa, sendo tal prática efetuada inclusive por órgãos reguladores (Barry, 2001; Kim, 2005).

Uma característica relevante da produção rural e, por consequência, do risco de crédito dos produtores, é a exposição a fenômenos naturais catastróficos (e.g., seca extrema, chuvas excessivas, furacões) que podem prejudicar um grupo de produtores da região afetada pelo evento. Pesquisas em *environmental economics* relacionadas a fenômenos naturais catastróficos têm identificado que o formato da distribuição de probabilidade de frequência (Malamud et al., 2005; Malamud & Turcotte, 2006) e da distribuição de probabilidade de severidade (Conte & Kelly, 2018, 2021; Kousky & Cooke, 2009; Weitzman, 2009) apresentam caudas pesadas. Distribuições de probabilidade com elevada curtose são um desafio à gestão de risco pois: (i) dados históricos são menos confiáveis para a predição futura do evento, e; (ii) distribuições de probabilidade conjuntas destes eventos também têm cauda pesada (Kousky & Cooke, 2009).

Se as distribuições de probabilidade de frequência e severidade dos fenômenos naturais catastróficos possuem caudas pesadas e estes fenômenos, por sua vez, afetam o retorno financeiro dos produtores rurais, então há a possibilidade de que as distribuições de probabilidade de *default* de operações de crédito rural também apresentem caudas pesadas. Além disso, há a possibilidade de existência de uma forte estrutura de dependência positiva entre as operações, especialmente se (i) um mesmo indivíduo afetado pelos desastres naturais estiver exposto a múltiplas linhas de crédito, ou; (ii) se múltiplos agentes estiverem concentrados em uma mesma região sujeita aos fenômenos.

Por esta razão, é importante que a dependência seja corretamente estimada e considerada em diversas instâncias da gestão de risco. Afinal, como preconiza a Teoria Moderna de Finanças, desconsiderar a estrutura de dependência entre operações pode conduzir à subestimação do risco efetivamente incorrido por agentes econômicos (Fenech et al., 2015; Moreira, 2010; Pertaia et al., 2021; Schäfer & Koivusalo, 2013; Schönbucher & Schubert, 2001).

O objetivo deste trabalho é identificar qual é a melhor forma de se representar a distribuição de probabilidade de *default* de um portfólio de crédito rural. Como justificativas, pode-se elencar diversos fatores como: (i) a importância social do crédito rural; (ii) a escassez de modelos de risco de crédito elaborados especificamente para este segmento, e; (iii) a exposição compartilhada de produtores rurais (tomadores de crédito) a fatores de riscos comuns, como os desastres naturais. A estimação do risco de um portfólio de crédito reverbera em aspectos prudenciais de regulação das instituições financeiras, cuja face mais relevante é o capital regulatório. Por conta disso, será dado enfoque na abordagem do Acordo de Basileia ao risco de crédito rural.

A análise da distribuição de probabilidade de *default* será feita com base em microdados de um portfólio de crédito do principal banco de financiamento rural do Brasil, pelo critério de *market share*. O portfólio avaliado contém sete linhas de crédito (Pronaf, Pronamp, Custeio, Giro, Investimento, Aquisição Ativos Financeiros e; Garantia de Preços e Estocagem). A linha de crédito Pronaf será enfatizada – por ser a mais relevante em número de operações –, e avaliada sua interação com as outras linhas. O Pronaf é um programa de financiamento subsidiado pelo setor público com o objetivo de fortalecimento da agricultura familiar (Brasil, 2021). Estimar a forma funcional mais adequada para a distribuição conjunta da probabilidade de *default* da linha de crédito Pronaf com outras linhas de crédito rural é relevante por dois motivos: (i) o Brasil é um dos principais fornecedores de alimento do mundo, com tendência a aumentar a sua participação até 2030 (OECD & FAO, 2021), e; (ii) a linha de crédito Pronaf auxilia no atingimento da meta da Agenda 2030 de aumento da produtividade e da renda de pequenos produtores.

As distribuições de probabilidade de *default* conjunta serão estimadas por meio de cópulas. Cópulas são funções estatísticas que generalizam o conceito de estruturas de dependência (Nelsen, 2006), devido à flexibilidade de se ajustarem a quaisquer conjuntos de distribuições: (as)simétricas, caudas pesadas, discretas/contínuas.

Para a análise do risco do portfólio, serão estimadas as distribuições de probabilidade conjunta bivariadas, da PD observada nas operações de crédito Pronaf com as PD observadas nas outras seis linhas de crédito rural. Após realizar inferências acerca da cópula que melhor representa a distribuição conjunta empírica, será verificada a sua aderência aos dados observados, comparando-a com a aderência da distribuição Normal bivariada, simétrica, tradicional na literatura e usual nos modelos de risco de crédito (Accornero et al., 2018; Babbal, 1989; Bégin et al., 2019; Grass, 2012; Jorion, 2007; Laurent, 2008; Rosen & Saunders, 2009).

Os resultados obtidos podem contribuir na elaboração de modelos probabilidade de *default* de crédito rural ajustados às especificidades deste setor, auxiliando credores a dimensionar mais fidedignamente o risco de suas operações, resultando em um processo de precificação mais eficiente (a taxas mais próximas da atuarialmente justa) e, em um melhor ambiente prudencial (devido à possibilidade de uma estimacão mais precisa do risco agregado do portfólio e composição de capital econômico). Além dos benefícios supramencionados, direcionados aos credores e reguladores, há a contribuição subjacente da maior precisão dos modelos de PD de crédito rural à ampliação da oferta deste tipo de financiamento, contribuindo ao desenvolvimento da agricultura e à meta humanitária de erradicação da fome.

O trabalho está dividido em 8 capítulos além desta introdução. O próximo capítulo abordará como os riscos de desastres naturais vêm sendo estimados pela literatura e como eles se relacionam com a produção rural. O terceiro capítulo trará uma compilação de modelos de risco de crédito rural, verificando o tratamento dado às especificidades do segmento e à potencial cauda pesada da PD. Já o Capítulo 4 trata de como o Acordo de Basileia determina que seja estimado o risco de um portfólio de crédito rural. O capítulo subsequente traz como a teoria de cópulas vem sendo aplicada na modelagem de risco de crédito. O sexto capítulo apresenta a estrutura de dados utilizada nos modelos empíricos que serão desenvolvidos no Capítulo 7. Por fim, o último capítulo é reservado às conclusões.

2. Risco de produção agrícola e desastres naturais

2.1 Introdução

A produtividade agrícola é crucial para o objetivo de redução da pobreza (Oyinbo et al., 2021). Federico (2005) a explica como uma função de características da terra, da força de trabalho e de inovações técnicas. O aumento da produtividade agrícola também é endereçado por estudos econômicos que podem auxiliar na tomada de decisão dos agentes relacionados à produção rural. A *agricultural economics* é a área da ciência social aplicada que avalia como produtores e consumidores utilizam recursos escassos na produção, processamento, *marketing* e consumo de produtos agrícolas (Penson et al., 2017). A evolução tecnológica e da ciência de dados vêm endereçando os estudos recentes.

Antle (2019) aborda como a ciência computacional e os avanços em estruturas de dados podem acelerar a transformação nos sistemas agrícolas. Novas tecnologias como veículos aéreos não tripulados, robôs e sensores são capazes de coletar um grande volume de dados, reduzindo os custos e tempo dispendidos neste processo. Além disso, há um aumento da interdisciplinaridade do conhecimento, incorporando avanços de áreas como microbiologia do solo e genética. Um exemplo de aplicação são os estudos baseados em genes das plantas, que, com o advento da tecnologia computacional, podem representar complexas interações entre genótipo, ambiente e manejo (Cooper et al., 2016; Technow et al., 2015).

Zilberman (2019) traça um quadro sobre a evolução da pesquisa agrícola, abordando quais são principais linhas de pesquisa da *agricultural economics*, sendo os estudos ambientais uma vertente. Estes, abordam como a utilização dos recursos naturais e as mudanças climáticas afetam a produção rural, por meio da interdisciplinaridade com outras áreas do conhecimento como a biologia, física e geologia.

2.2. Risco de produção e desastres naturais

Além do risco tradicional de mercado de *commodities*, referente às flutuações de preços, de oferta e de demanda dos insumos e produtos agrícolas, os produtores rurais estão expostos ao risco de produção, sendo este uma função de suas decisões de produção, sujeitas às condições climáticas e naturais (Just et al., 2010). As intempéries da natureza afetam a agricultura e toda a rede de agentes econômicos a ela relacionados, conectando-se ao risco de crédito.

Hohl (2019) aborda os riscos de produção agrícola sob a perspectiva de uma seguradora, considerando os riscos catastróficos. O autor classifica este tipo de risco em duas categorias: (i) riscos naturais (seca, enchentes, ciclones, neve e geada, incêndios, terremotos e tsunamis, erupções vulcânicas e temperaturas extremas), e; (ii) biológicos (pestes e doenças). A severidade das perdas ocasionada pela materialização destes riscos é apresentada como uma função das seguintes variáveis: (i) intensidade e duração do fator de risco; (ii) momento do impacto durante a safra, fase de crescimento ou ciclo da pecuária; (iii) eficiência dos agentes responsáveis pelo plano de contingência do risco (e.g., contenção de um incêndio), e; (iv) proficiência dos sistemas de alerta de risco.

Dada a natureza dos riscos relacionados à produção agrícola apresentados por Hohl (2019), o seu potencial de impacto tende a ser regional ou sistêmico. É pouco provável que a materialidade de algum risco afete apenas um único produtor, sem produzir efeitos de qualquer escala em outros agentes relacionados por produto ou região geográfica. A correlação espacial entre os produtores rurais ocorre por compartilharem um padrão climático (Goodwin, 2001). O risco de produção compartilhado entre os produtores implica em uma maior dependência dos riscos de crédito individuais destes produtores. Esta dependência intrínseca entre risco de produção e risco de crédito, torna (ou, ao menos, deveria tornar) objeto de interesse dos

credores, as características dos fenômenos naturais ou biológicos que interferem na produção, de modo a incorporar o comportamento destes fenômenos em seus modelos de risco de crédito.

Deste modo, é relevante para o mercado de crédito rural a incorporação de conhecimentos dos estudos climáticos focados em avaliar a frequência, severidade e estrutura de dependência de fenômenos naturais extremos. Há uma forte (e recente) linha de pesquisa em *environmental economics* sendo desenvolvida a respeito deste tema (Bazrafshan et al., 2020; Cheng et al., 2019; Conte & Kelly, 2018, 2021; Erhardt et al., 2015; Kousky & Cooke, 2009; Malamud et al., 2005; Malamud & Turcotte, 2006; Mesbahzadeh et al., 2020; Weitzman, 2009; Wong & Collins, 2020).

Após ampla revisão de literatura, Conte & Kelly (2021) constatam que o comportamento de variáveis que estimam a amplitude térmica e perdas por desastres naturais apresentam distribuições de probabilidade com maior curtose. Com foco na distribuição de perdas ocasionada por desastres naturais, os autores exemplificam como a dependência de cauda resulta em um risco potencial agregado (e.g., ciclones tropicais que podem desencadear ventos fortes, inundações e tempestades, ampliando os danos potenciais). Ignorar esta estrutura de dependência subestima a potencial severidade dos eventos catastróficos.

Kousky & Cooke (2009) apresentam três aspectos da distribuição de perdas de desastres naturais: cauda pesada, dependência de cauda e microcorrelações. Distribuições de probabilidade com caudas pesadas têm elevados desvios de ocorrência de um evento em relação à sua frequência média. A dependência de cauda ocorre quando duas ou mais variáveis aleatórias têm a sua dependência concentrada nos valores extremos do domínio, de forma que perdas severas sejam mais prováveis de ocorrerem de forma conjunta em comparação com a ocorrência conjunta de perdas com menor severidade. Por fim, as microcorrelações são pequenas correlações de fatores de riscos isolados. Entretanto, quando há alguma agregação das exposições, o risco conjunto é significativamente incrementado (e.g., a correlação do risco de incêndio entre duas casas é menor do que a correlação do risco de incêndio entre um conjunto de vinte casas).

Diversos pesquisadores identificaram distribuições de probabilidade com caudas pesadas em variáveis relacionadas aos riscos naturais e biológicos. De forma não exaustiva, há exemplos de riscos estimados em incêndios florestais (Malamud et al., 2005), enchentes (Malamud & Turcotte, 2006), mudanças climáticas (Weitzman, 2009), ciclones tropicais (Conte & Kelly, 2018) e contágio do vírus COVID-19 (Wong & Collins, 2020).

Sob o escopo da gestão dos riscos naturais e suas aplicações, a atribuição teórica da forma funcional da distribuição de probabilidade das variáveis relacionadas é apenas uma das etapas. Também se faz necessário identificar se há ou não uma estrutura de dependência entre as variáveis de interesse, sobretudo pelo já mencionado potencial sistêmico e de contágio oriundo da materialização dos riscos naturais.

Um dos possíveis métodos de estimação de estruturas de dependência é aplicação de cópulas estatísticas. Devido à sua capacidade de identificar dependência de cauda, os modelos de cópulas vêm apresentando boa aderência aos dados climáticos. Erhardt et al. (2015) abordaram o efeito das características espaciais (distância e altitude) de estações meteorológicas da Alemanha nas temperaturas médias registradas por essas estações. Os autores estimaram dois modelos: um assumindo a distribuição conjunta das variáveis como uma Gaussiana multivariada (normal multivariada), e outro com a aplicação da *R-vine* copula, considerando uma forte dependência de cauda entre as variáveis, obtendo melhor resultado no modelo de *R-vine* cópula.

Modelos baseados em cópulas estatísticas também são utilizados na estimação de secas meteorológicas (baixa precipitação) e agrícolas (baixa umidade do solo). Preocupados com o aumento da seca na província de Yazd (Irã), Mesbahzadeh et al. (2020) elaboraram um modelo de cópulas para inferir sobre a probabilidade de ocorrência de secas futuras, auxiliando a tomada de decisão administrativa para mitigar danos. Os resultados indicam que a província de Yazd deverá passar por maiores secas no futuro em relação ao seu histórico, em termos de duração e severidade. Também observando o Irã, Bazrafshan et al. (2020) construíram um índice preditivo de secas baseado em cópulas, apresentando melhor resultado em períodos extremos em relação a um modelo tradicional sem dependência de cauda.

Cheng et al. (2019) utilizaram um modelo de C-vine copulas para estimar a dependência entre secas e ondas de calor nos Estados Unidos. Wu et al. (2021) também construíram um modelo de C-vine copulas para estudar o fenômeno da seca, estimando a estrutura de dependência entre temperatura, secas agrícolas e meteorológicas na China. Os resultados indicaram que o modelo desenvolvido apresenta boa aderência à predição da seca na maior parte das regiões analisadas.

Este capítulo tratou das características dos riscos naturais relacionados à produção agrícola. Há consenso na literatura que estes riscos possuam distribuições de probabilidade subjacentes de caudas pesadas, e que a estrutura de dependência entre eles não siga uma distribuição Normal simétrica, havendo repercussão na dependência nos valores extremos. Estudos recentes que aplicaram a teoria de cópulas estatísticas para estimar estruturas de dependência entre variáveis ambientais relacionadas aos riscos naturais obtiveram resultados satisfatórios (Bazrafshan et al., 2020; Cheng et al., 2019; Erhardt et al., 2015; Mesbahzadeh et al., 2020; H. Wu et al., 2021). Espera-se que modelos de PD de crédito rural que incorporem a característica de distribuição de probabilidade de cauda pesada dos riscos naturais relacionados à produção obtenham ganhos de precisão na estimação do risco de crédito.

3. Modelos de PD de crédito rural a nível individual

Os modelos de risco de crédito visam estimar a probabilidade de melhora ou piora na qualidade de crédito, sendo o *default* (inadimplência) o evento de grande interesse nas análises, mas não o único. Além da inadimplência, a migração entre diferentes faixas de risco, a duração do atraso nos pagamentos e a perda condicional ao *default* também podem ser os objetos de interesse. Há grande variedade de modelos de risco de crédito que se diferenciam por suas premissas. Entretanto, o objetivo comum destes modelos é prever a função de distribuição de perdas das carteiras de crédito (Hayden & Porath, 2006; Lopez & Saidenberg, 2000; Ryser & Denzler, 2009). Os modelos mais usuais derivam de duas abordagens: (i) estrutural, e; (ii) de forma reduzida (Altman & Saunders, 1998; Bessis, 2015; Jeanblanc & Rutkowski, 1999).

Modelos estruturais auxiliam a compreender quais são as causas do *default*. Tais modelos consideram que o *default* ocorre quando o devedor não cumpre com suas obrigações por falta de capacidade (ou vontade), sob influência de sua estrutura de capital. Por basearem-se em informações contábeis de ativos e passivos dos devedores, estes modelos são considerados endógenos aos agentes. Ademais, esta família de modelos é usualmente aplicada ao risco de crédito de empresas de capital aberto, uma vez que as oscilações de valor de seu patrimônio – marcada a mercado (MtM) – são *proxies* da capacidade de pagamento de seus passivos. Exemplos de modelos estruturais são propostos por Black & Scholes (1973), Geske (1977), Merton (1973) e Vasicek (1984).

Já os modelos de forma reduzida têm por objetivo estimar quando o *default* irá ocorrer. Neste caso, o evento *default* deriva de fatores exógenos ao devedor, regidos por um processo estocástico (Processo de Poisson), cujo parâmetro é a intensidade da frequência de *default* (Arora et al., 2005; Darrell & Schaefer, 2003). Assim, não é necessário ter acesso à estrutura de capital dos devedores para estimar a probabilidade de *default* (Klieštk & Cúg, 2015). Exemplos de modelos de forma reduzida são os modelos desenvolvidos por Darrell & Schaefer (2003) e Jarrow & Turnbull (1995).

Os modelos de *credit scoring* podem ser considerados como uma terceira via, pois apesar de se assemelharem aos modelos estruturais, são passíveis de incorporar outros fatores endógenos diferentes da estrutura de capital dos devedores, que influenciam no *default*. Além disso, podem ser elaborados com auxílio de modelos de forma reduzida. Em síntese, estes modelos são estimados com base em representações multivariadas, utilizando as características observáveis dos devedores, ou potenciais tomadores de recursos, como variáveis explicativas em funções estatísticas que atribuem uma nota (*score*). O *score* é usado para a concessão dos empréstimos, indicando a capacidade individual em honrar com a obrigação assumida e dando início ao ciclo do produto (Bliss, 2002; Mester, 1997). Em sua forma mais simples, o *score* é binário, segregando os devedores em bons ou maus pagadores, e em análises mais completas, há a estratificação dos devedores, indicando graus de capacidade de pagamento. A atribuição de classificações de risco padronizada para faixas de *score* viabiliza a análise de risco de crédito de maneira massificada, sendo tal metodologia mais adequada às exposições de varejo. Além disso, o *score* auxilia no gerenciamento do portfólio de crédito e na precificação de operações (Featherstone et al., 2006).

O foco desta pesquisa (i.e., a verificação da melhor forma de se representar a distribuição de PD de um portfólio de crédito rural), visa auxiliar o desenvolvimento de modelos de PD de crédito rural mais ajustados aos fatores de risco destas operações. Os modelos usuais carecem de metodologias próprias, sendo comumente elaborados com base em modelos de risco de crédito corporativo ou de varejo (Kim, 2005).

Entretanto, os empréstimos rurais possuem características específicas que podem afetar a probabilidade de *default* de uma carteira de forma diferente do esperado em outros tipos de

financiamento. O fluxo de receita dos produtores rurais e, por consequência, a sua capacidade de pagamento, é condicionado às safras (Barry, 2001; Katchova & Barry, 2005). Além disso, é de se esperar elevada correlação entre os resultados de produtores rurais, não apenas por proximidade geográfica, como também por sazonalidade das safras e similaridades de fatores de produção e exposição a riscos naturais e biológicos.

Os modelos tradicionais de risco de crédito não incorporam estas especificidades do crédito rural. Pesquisadores que objetivaram estimar o risco de crédito rural ou ignoraram este problema ou propuseram alguma forma de ajuste para melhorar as suas estimativas. As próximas seções deste capítulo trarão uma revisão de literatura detalhada sobre como o risco de crédito rural é abordado pela literatura.

3.1. Modelos *Credit Scoring* aplicados ao crédito rural

Chhikara (1989) fez uma revisão sobre os modelos de *credit scoring* utilizados na análise de crédito rural à época. A primeira divisão dos modelos é em relação à utilização de parâmetros na estimação: modelos paramétricos e não paramétricos. São modelos paramétricos os modelos: (i) de probabilidade linear; (ii) *probit*; (iii) *logit*, e; (iv) de análise discriminante. Os modelos de probabilidade linear são pouco utilizados devido aos problemas de heterocedasticidade. Os modelos *probit* e *logit* restringem a forma funcional da variável explicada ao intervalo (0,1), seguindo as distribuições de probabilidade normal (*probit*) e logística (*logit*), havendo aplicação ao crédito rural constatada pelo autor (Lufburrow et al., 1984; Miller & LaDue, 1988). Os trabalhos empíricos de modelos de *credit scoring* aplicados ao crédito rural serão apresentados, nas seções a seguir, sob a dicotomia de modelos paramétricos e não paramétricos.

3.1.1. Modelos paramétricos de *Credit Scoring*

Análise discriminante

A análise discriminante consiste em aplicar uma função de classificação dos devedores com base em uma amostra e parâmetros pré-definidos, assumindo o pressuposto de que os parâmetros aplicados à função são normalmente distribuídos. Johnson & Hagan (1973) e Weed & Hardy (1980) utilizaram este método para classificar os proponentes de crédito rural como aceitáveis ou não. Quando utilizada de forma isolada, a análise discriminante é pouco robusta, devido à necessidade de atendimento do pressuposto de normalidade e de linearidade (na maioria dos modelos). Além disso, não é possível verificar o poder discriminante de cada variável no modelo isoladamente (Dillon, 1979; Eisenbeis, 1978). Entretanto, quando combinada com outras técnicas, a análise discriminante pode produzir resultados com maior confiabilidade.

O trabalho de Bennouna & Tkouat (2019) é um exemplo de aplicação conjunta de análise discriminante com modelos de regressão *Logit*. Os autores combinaram as técnicas em um único modelo de *credit scoring* de microcrédito rural. O microcrédito é destinado às pessoas de baixa renda com pouco ou nenhum acesso ao sistema financeiro, sendo uma forma de inclusão econômica (Bauer et al., 2012). Devido ao perfil dos usuários, as operações de microcrédito dificilmente têm alguma garantia vinculada, sendo fundamental que os credores elaborem uma análise de risco assertiva. O modelo proposto foi elaborado com dados de 1500 operações individuais de instituições de microcrédito no Marrocos. Para classificar os devedores como bons ou maus, os autores aplicaram dois modelos *logit*. O primeiro para avaliar a influência das variáveis demográficas e econômicas dos devedores, como sexo, estado civil e renda, e o segundo para avaliar as características do histórico de empréstimos, como número de empréstimos contratados e prazo da dívida. O *score* obtido em cada um dos modelos foi

combinado, via análise discriminante. A área sob a curva ROC do modelo combinado foi superior à área dos modelos *logit* individuais, indicando maior eficiência global.

Logit

A utilização de modelos *logit* na construção de *credit scoring* é a mais recorrente na bibliografia. Khoju & Barry (1993) aplicaram o modelo em um estudo baseado em dados financeiros de fazendas de grãos do estado de Illinois, EUA. Turvey & Weersink (1997) buscaram maior compreensão sobre os fatores que influenciam a oferta e demanda de crédito rural na *Farm Credit Corporation* (empresa de fomento ao crédito rural de propriedade da coroa canadense). Os autores consideram que a curva de oferta de crédito é oriunda da distribuição de perdas, equivalente à variável dependente dos modelos de *credit scoring*. Assim, estimam um modelo *logit* para a probabilidade de *default*, com a preocupação de diferenciar os tipos de produtores de acordo com o que produzem (grãos, oleaginosas, leite, pecuária e horticultura), devido às diferenças de exposição ao risco de cada atividade.

Entender quais são os fatores que influenciam a inadimplência é tarefa de maior interesse dos credores, e cuja consequência é viabilizar uma melhor e maior oferta de crédito, auxiliando no desenvolvimento rural. É com este racional que Chaudhary & Ishfaq (2003) propuseram um modelo para mensurar a probabilidade de *default* de famílias rurais no Paquistão, tendo como variáveis explicativas características: (i) contratuais do empréstimo (valor; prazo; taxa de juros; garantias vinculadas; tipo de credor, e; motivo do crédito); (ii) financeiras do proponente (histórico de dívidas e patrimônio) e; (iii) demográficas (educação, ocupação e sexo). A variável ocupação é importante por informar se o proponente do crédito é assalariado ou trabalhador rural, capturando sua capacidade de pagamento. Os resultados indicam que a probabilidade de *default* aumenta quando o devedor possui a ocupação de produtor rural, *caeteris paribus*. Além disso, o trabalho contribui no entendimento sobre a vinculação de garantias às operações. Segundo os autores, no Paquistão, as operações de crédito são tradicionalmente condicionadas a garantias excessivas e que muitas vezes impossibilitam a contratação da operação. Entretanto, os resultados indicam que elementos qualitativos como educação e ocupação podem ser mais significantes para a probabilidade de *default* do que a vinculação de garantias.

A utilização de modelos *logit* para a elaboração de *credit scoring* também foi aplicada sob a ótica de capital regulatório. Featherstone et al. (2006) tratam do incentivo que o Acordo de Basileia II deu às instituições financeiras para utilizarem modelos internos de classificação de risco (IRB), devido à possibilidade de redução de requerimento de capital. Preocupados com a falta de comparabilidade entre os modelos dos credores rurais e por consequência, do risco entre as instituições, os autores propõem um modelo de classificação de risco seguindo os critérios de *rating* praticados pela *Standard & Poor's (S&P)*, passível de ser replicado por todas as instituições. Elaborado com dados de operações de crédito rural de uma instituição financeira norte-americana, o *default* foi definido como a falta de pagamento superior a 90 dias da data da obrigação, estimado por meio de um modelo *logit* e então classificado seguindo os critérios da *S&P*.

Durguner et al. (2006) comentam sobre a adequação do *Farm Credit System* ao Acordo de Basileia II. O *Farm Credit System* é um sistema de cooperativas de crédito rural e o maior credor do segmento nos Estados Unidos. Os autores consideram que o sistema de crédito é consistente com o Acordo por já praticar uma estratificação na mensuração de risco, ajustando os modelos de *credit scoring* aos tipos de operações (*commercial farm loan, farm real estate loan, agribusiness loan, rural housing loan, e small loans*). Entretanto, os autores questionam se as diferenças entre as categorias de fazendas não deveriam ser consideradas na mensuração de risco de crédito, uma vez que a atividade praticada (e.g., agricultura ou pecuária) afeta de

forma significativa a taxa de retorno dos ativos, liquidez e índices de alavancagem. A verificação empírica é feita com dados de crédito rural do estado de Illinois, EUA, estimando dois modelos *logit*: um para as fazendas de agricultura e outro para as fazendas de pecuária. Os resultados apresentaram coeficientes e níveis de significância diferentes entre as variáveis explicativas dos modelos, indicando que distinguir os devedores de acordo com o tipo de atividade da fazenda reduz os erros de classificação e torna o processo de avaliação de crédito mais eficiente.

Tal como Bennouna & Tkiouat (2019), outros pesquisadores têm dedicado os seus estudos para estimar o risco de operações de microcrédito rural. O campo de microfinanças mostrou-se um forte instrumento de redução de desigualdade social (Baghdasaryan et al., 2021; Bennouna & Tkiouat, 2019; Medina-Olivares et al., 2021). O microcrédito é um instrumento relevante ao desenvolvimento de comunidades rurais (Condori-Alejo et al., 2021), e entender os fatores que influenciam a inadimplência nesta linha de crédito é fundamental para a perenidade da sua oferta.

Hering & Musshoff (2017) dissertam sobre o acesso ao microcrédito rural no Azerbaijão. Instituições que operam com microcrédito sofrem com elevado grau de assimetria informacional, devido à elevada imprecisão sobre as características de renda e de produção dos devedores. Como medida de melhora na qualidade de risco, os credores operam com o que é conhecido como *empréstimo progressivo*, o qual racionam a disponibilidade de crédito aos novos clientes, que, conforme vão honrando seus compromissos, têm o seu limite de crédito ampliado.

Os autores avaliam se a progressão no limite de crédito ocorre da mesma maneira entre devedores rurais e não rurais. Os resultados indicam que os proponentes ao crédito rural enfrentam maior restrição ao crédito do que os outros proponentes, sob o mesmo nível de garantia vinculado à operação. Para verificar se a diferença na oferta de crédito dá-se de fato pelo risco, os autores estimaram um modelo *logit* para explicar o atraso ou *default* da operação por: (i) características do empréstimo e do mutuário; (ii) controles de sazonalidade, e; (iii) a *dummy Farmer* (igual a 1 quando o proponente é do grupo rural e 0 caso contrário). A *dummy Farmer* foi significativa ao nível de confiança de 99%, com coeficiente negativo, indicando que, *caeteris paribus*, os proponentes rurais apresentam menor probabilidade de atraso ou *default* da operação contratada quando comparados aos proponentes não rurais. Tal resultado torna a assimetria ao acesso ao microcrédito rural mais exacerbada.

Considerando o risco sistêmico do microcrédito rural, devido à exposição compartilhada aos riscos naturais, Medina-Olivares et al. (2021) elaboram um modelo *logit* de *credit scoring* adicionando um componente de localização espacial dos devedores. O modelo foi testado empiricamente com os dados de uma das maiores instituições de microcrédito chinesa. Os resultados indicaram aumento no poder explicativo dos modelos com a inclusão da componente espacial e uma influência positiva no parâmetro de concentração espacial com a probabilidade de *default*.

3.1.2. Modelos não paramétricos de *Credit Scoring*

Goal Programming (GP)

Os modelos não paramétricos apresentados por Chhikara (1989) são o *Goal Programming* e o *Recursive Partitioning Algorithm* (RPA). O *Goal Programming* (GP) resolve um problema de otimização linear, com uma função objetivo submetida a um conjunto restrições. Hardy & Adrian (1985) aplicaram o GP na elaboração de um modelo de *credit scoring* de operações rurais de forma a maximizar o desvio do *credit score* atribuído a cada operação do valor crítico do *cutoff* que divide as operações como aceitáveis ou não.

Recursive Partitioning Algorithm (RPA)

O *Recursive Partitioning Algorithm* (RPA) é uma ferramenta de classificação de variáveis que efetua uma série de divisões binárias no conjunto de dados sob algum critério e repete este processo até que nenhuma subdivisão seja possível. Novak & LaDue (1999) avaliam o RPA como uma alternativa ao modelo *logit* na classificação de devedores rurais. A comparação dos modelos é feita com base na proporção dos erros de classificação do tipo 1 (emprestar para um mau pagador) e tipo 2 (não emprestar para um bom pagador), dando preferência ao erro do tipo 2, pois os autores assumem que o prejuízo esperado deste erro seja menor do que o prejuízo esperado do erro tipo 1. Em todas as simulações, o modelo RPA apresentou menor proporção dos erros de classificação. Entretanto, parte dos resultados têm menor poder explicativo quando comparados ao modelo logístico, sendo inconclusiva a decisão de qual é a melhor metodologia, ao menos para este estudo.

Mandala et al. (2012) aplicam um modelo RPA para identificar quais são os fatores que deveriam ser considerados por um banco rural em sua análise de crédito. Por meio de um estudo de caso de um banco de Bali, os autores encontraram uma taxa de 11,99% de *non performing loans*, muito acima do limite de 5% permitido pelo órgão regulador. O algoritmo utilizado foi o C5.0, devido à sua capacidade de gerar múltiplos galhos em cada ramo, superando a limitação de ramificação binária do algoritmo CART. O modelo decisório estimado incluiu as seguintes variáveis: (i) existência de operações anteriores; (ii) garantias vinculadas à operação; (iii) poupança dos proponentes e; (iv) submissão de toda a documentação requerida. Ao aplicar o algoritmo C5.0, foi gerada uma árvore de decisão que, além da análise das garantias, leva em conta o período do empréstimo e o tamanho da empresa (para os casos de empréstimos corporativos), tendo como resultado uma taxa de *non performing loans* de 3%.

Na China, os pequenos produtores rurais enfrentam dificuldade de acesso ao crédito em bancos devido à sua produção de pequena escala; a sazonalidade de seu fluxo de caixa; a incapacidade de atendimento dos requisitos de garantia, e; ao baixo histórico financeiro (Rao et al., 2020). Uma alternativa ao acesso ao crédito são as plataformas online de empréstimo P2P, nas quais os produtores podem obter empréstimos diretamente com os investidores. Rao et al. (2020) utilizaram os dados da *Pterosaur Loan*, plataforma de crédito P2P, para estimar o risco de crédito de seus usuários com base em um modelo *Syncretic Cost-Sensitive Random Forest* (SCSRF). O SCSRF é um modelo RPA sensível aos custos de erros de classificação, considerando que há diferença na severidade entre os erros do tipo 1 e do tipo 2, aplicando cortes nos dados sob o critério do menor custo de classificação.

Multiple criteria decision analysis (MCDA)

A tomada de decisão efetuada sob *multiple criteria decision analysis* (MCDA) considera um grupo de critérios para avaliar as alternativas, levando em conta os prós e contras que cada critério tem sobre estas alternativas. Uma das premissas do MCDA é a possibilidade de existência de estruturas de dominância entre os critérios (J. Figueira et al., 2005). Chavira et al. (2017) aplicaram o algoritmo ELECTRE-III em um problema de MCDA, para determinar quais variáveis são relevantes na análise de crédito rural de uma instituição financeira não bancária no México. A opção pela MCDA ocorre pelo entendimento de que os modelos *logit* e de análise discriminante necessitam de um volume muito grande de dados para se ter uma indicação razoável do perfil das operações em *default*. Entretanto, este tipo de dado é escasso. Além disso, o MCDA permite a recalibragem, sendo passível de ajustes nos parâmetros pelo analista de crédito.

Entre as vantagens do algoritmo ELECTRE-III está a possibilidade de inclusão de variáveis quantitativas e qualitativas, ambas necessárias na análise de crédito. Ademais, o ELECTRE-III é um modelo não compensatório, i.e., o resultado indesejado em algum dos critérios não pode ser compensando pelo resultado positivo em outro critério (e.g., se uma

operação entra na zona de rejeição pelo critério *tipo de plantação*, este resultado negativo não pode ser compensando por uma boa performance no critério *ativos do proponente*).

Fuzzy Set

Bai et al. (2019) aplicam a lógica *fuzzy set* para verificar a relação entre características de fazendas, cenário competitivo e risco de crédito rural. Os dados utilizados pertencem a um banco estatal Chinês, contendo 2044 produtores rurais com ampla diversificação geográfica. A cada produtor são associadas 43 variáveis condicionais. Os autores aplicaram o *Fuzzy C-Means Algorithm* (FCM) para gerar *clusters*. A escolha do FCM para realizar o agrupamento se deu pelo fato de que o algoritmo transforma os dados contínuos em discretos, sendo essa uma condição necessária para os próximos passos da análise. Cada *cluster* representa um nível de gradação de cada variável. Então, cada produtor foi associado a cada grupo com um peso que varia de 0 a 1. Quanto mais elevado o peso, mais forte é a associação do produtor ao *cluster*.

Por fim, o produtor é vinculado somente ao *cluster* em que ele apresenta maior peso. As regras criadas pelo modelo acertaram 93,7% das classificações na pior faixa de risco. Os autores consideram que a vantagem deste método é a facilidade de aplicação (uma vez que as instituições financeiras não precisam ter nenhuma informação adicional para além das que já possui em sua base de dados) e ser escalável, pois as regras podem ser facilmente aplicadas a um grande volume de dados. Além disso, o modelo fornece ganhos computacionais, uma vez que elimina as variáveis não significantes (no caso, das 43 variáveis, o modelo final utilizou 12), sem precisar efetuar uma série de iterações, como nos modelos paramétricos.

Artificial Neural Networks (ANN)

Condori-Alejo et al. (2021) compararam a eficiência de modelos tradicionais e de *Machine Learning* para classificação de operações de microcrédito rural no Peru. Os dados pertencem a uma instituição localizada na região de Puno, contendo 17.454 observações de 15.015 clientes. A performance dos modelos foi comparada pela área abaixo da curva ROC. Os modelos estimados e as suas respectivas áreas foram: (i) *Artificial Neural Networks* (ANN), 93,72%; (ii) *Decision Trees* (dTree), 88,80%; (iii) *logit*, 86,07% (iv) *Support Vector Machines* (SVM), 84,44%; (v) *k-Nearest Neighbors* (kNN), 65,98%.

O poder preditivo de técnicas de *machine learning* também foi avaliado por Baghdasaryan et al. (2021). Os autores comparam os modelos econométricos *probit* e *logit* a um modelo ANN. Os dados utilizados pertencem a uma instituição financeira armênia, contendo informações de 9.196 empréstimos rurais entre os anos de 2012 e 2017. Foi observado melhor desempenho no modelo de redes neurais (área abaixo da curva ROC). Entretanto, os resultados do modelo ANN pouco explicam sobre os dados, sugerindo que os modelos ANN e econométricos podem ser complementares à análise de crédito. A ANN pode ser utilizada em decisões automatizadas de concessão de empréstimo enquanto os modelos econométricos podem servir às análises sobre o comportamento e perfil dos devedores.

3.2. Modelos de forma reduzida aplicados ao crédito rural

Mortality Function

Sob uma perspectiva atuarial, Barry et al. (2000) desenvolvem um modelo de forma reduzida para estimar o de risco de crédito de hipotecas rurais objetos de contratos de seguro. Os seguros de hipoteca protegem os credores caso os devedores não honrem com seus pagamentos, transferindo o ônus da inadimplência à seguradora. A probabilidade de *default* foi estimada via função de mortalidade, a qual considera que o evento de *default* (morte) ocorre quando o valor de uma dívida é inferior à capacidade de pagamento do mutuário. Já os eventos de cauda foram estimados de acordo com a quantidade de *defaults* observados em relação aos

defaults esperados ao final da observação, sendo a diferença distribuída no tempo futuro do empréstimo proporcional à densidade da distribuição neste tempo.

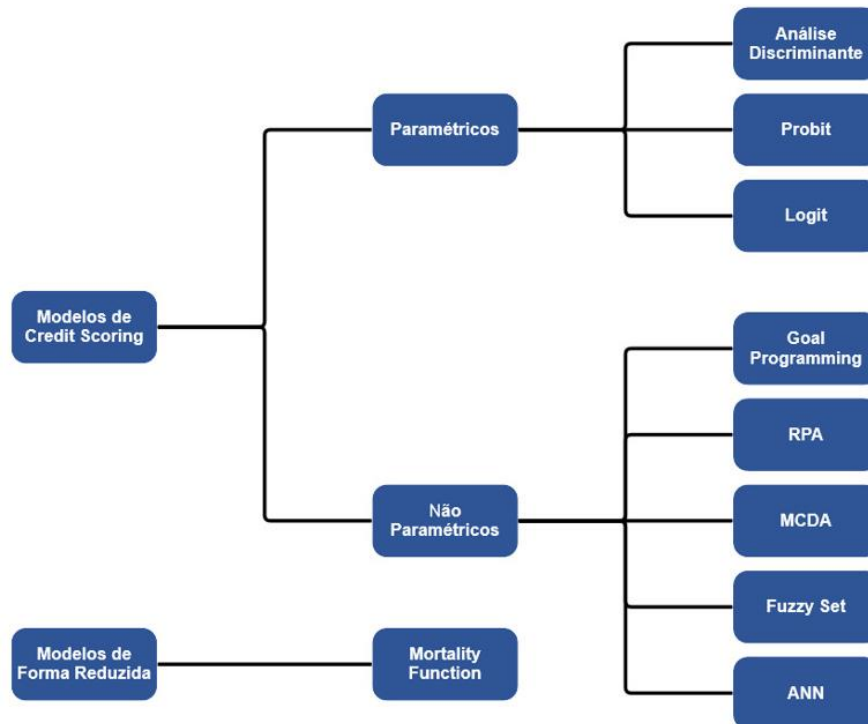
No esforço de adequar as diretrizes do Acordo de Basileia II aos credores rurais, Zech & Pederson (2004) buscam a melhor forma de adaptação dos modelos utilizados pelo mercado. A análise de crédito rural dificilmente é passível de ser efetuada via modelos de dívida corporativa devido à ausência de dados necessários para estes modelos, como uma fonte pública do histórico financeiro dos devedores e classificações de risco externas (fornecidas pelas agências de *rating*). Para propor um modelo adaptado ao crédito rural, os autores apresentam quatro modelos amplamente utilizados por credores: (i) *KMV Portfolio Manager*, criado pela Moody's KVM Corporation; (ii) *CreditMetrics*, criado pelo RiskMetrics Group vinculado ao J.P Morgan; *CreditPortfolioView*, criado pela *McKinsey & Company*, e; (iv) *CreditRisk+*, criado pela *Credit Suisse Financial Products*.

Após avaliar as premissas de cada modelo em termos de como o *default* é estimado, os autores consideram o *CreditRisk+* (de abordagem atuarial) mais adequado à disponibilidade de dados de operações rurais. Neste modelo de forma reduzida, o *default* é estimado via função de mortalidade, construída com as informações históricas de *default*. Para testar a adequação do modelo proposto, os autores utilizaram dados da cooperativa de crédito rural *AgStar Financial Services*, atuante nos estados de Minnesota e Wisconsin nos Estados Unidos, que à época detinha aproximadamente US\$ 2,3 bilhões em ativos.

Assim como Zech & Pederson (2004), Katchova & Barry (2005) adaptaram modelos tradicionais de risco de crédito às operações rurais. Entretanto, os autores diferem do trabalho predecessor, ao adaptar os modelos *CreditMetrics* e *KMV Portfolio Manager*. Estes modelos derivam do modelo de opções de Merton (1973), o qual assume que o *default* de uma empresa ocorre quando o valor presente de seus ativos se torna inferior ao valor de seus passivos. O *CreditMetrics* classifica os devedores de acordo com o seu risco e estima a probabilidade de migração dos devedores para classificações melhores ou piores em um horizonte de tempo. As classificações de risco e a matriz de transição são obtidos externamente, por meio das agências de *rating*. Como não há *rating* externo de operações rurais, os autores elaboraram a classificação com base na relação de ativos e passivos dos proponentes e a probabilidade de *default* foi atribuída como a proporção de operações em *default* em cada classe, em relação ao total de operações.

O *KMV Portfolio Manager* assume que o *default* decorre de um processo endógeno. Katchova & Barry (2005) agruparam os devedores de acordo com número de desvios padrões entre a diferença média dos ativos e passivos. Quanto mais desvios padrões houver, maior será a distância até o *default* e menor o risco. Embora ambos os modelos estejam adequados às diretrizes do Acordo de Basileia II, não houve convergência entre os requisitos de capital estimados. Tal resultado indica que a adoção discricionária de modelos de risco de crédito, resulta em diferentes requisitos de capital para o mesmo nível de risco.

A Figura 1 apresenta a classificação dos modelos de probabilidade de *default* aplicados ao crédito rural supramencionados.

Figura 1. Classificação dos modelos de PD aplicados ao crédito rural

A figura apresenta a classificação dos oito modelos aplicados ao crédito rural identificados na bibliografia. Os modelos são seccionados em modelos de *Credit Scoring* (Paramétricos e Não Paramétricos) e de forma reduzida.

Fonte: elaborado pelo autor

3.3. Considerações finais sobre os modelos de PD de crédito rural

Os modelos de risco de crédito rural evoluíram com o avanço das técnicas de gestão e mensuração de risco e de técnicas de *machine learning*. Ao longo do presente capítulo foram apresentadas aplicações práticas dos modelos de *Credit Scoring* (paramétricos e não paramétricos) e dos modelos de forma reduzida. Alguns destes modelos preocuparam-se de algum modo, com as especificidades do crédito rural. Entretanto, em toda a bibliografia observada sobre o tema, até o melhor de nosso conhecimento, somente um trabalho estimou a probabilidade de *default* a nível individual considerando a possibilidade de a distribuição de probabilidade de *default* apresentar cauda pesada (Barry et al., 2000). A Tabela 1 apresenta o resumo da bibliografia de modelos de risco de crédito rural.

Tabela 1. Resumo da bibliografia de modelos de PD de crédito rural a nível individual.

Ordem	Autores	Período dos dados	Região dos dados	Modelo de Risco	Método	Considera características rurais?	Considera cauda pesada?
1	Johnson & Hagan (1973)	1969–1970	EUA	<i>Credit Scoring</i>	An. Discr.	N	N
2	Weed & Hardy (1980)	1974–1978	EUA	<i>Credit Scoring</i>	An. Discr.	N	N
3	Lufburrow et al. (1984)	1982	EUA	<i>Credit Scoring</i>	<i>Probit</i>	N	N
4	Hardy & Adrian (1985)	1980	EUA	<i>Credit Scoring</i>	GP	N	N

5	Miller & LaDue (1988)	1984	EUA	<i>Credit Scoring</i>	<i>Logit</i>	N	N
6	Khoju & Barry (1993)	1985–1990	EUA	<i>Credit Scoring</i>	<i>Logit</i>	N	N
7	Turvey & Weersink (1997)	1981–1988	Canadá	<i>Credit Scoring</i>	<i>Logit</i>	S	N
8	Novak & LaDue (1999)	1985 - 1993	EUA	<i>Credit Scoring</i>	<i>Logit; RPA</i>	N	N
9	Barry et al. (2000)	1979–1992	EUA	Forma Reduzida	<i>Mortality Function</i>	N	S
10	Chaudhary & Ishfaq (2003)	1996–1997	Paquistão	<i>Credit Scoring</i>	<i>Logit</i>	S	N
11	Zech & Pederson (2004)	1997–2002	EUA	Forma Reduzida	<i>Mortality Function</i>	S	N
12	Katchova & Barry (2005)	1995–2002	EUA	Forma Reduzida	<i>Mortality Function</i>	S	N
13	(Durguner et al. (2006)	1995–2004	EUA	<i>Credit Scoring</i>	<i>Logit</i>	S	N
14	Featherstone et al. (2006)	1995–2002	EUA	<i>Credit Scoring</i>	<i>Logit</i>	N	N
15	Mandala et al. (2012)	2002–2012	Indonésia	<i>Credit Scoring</i>	RPA	N	N
16	Chavira et al. (2017)	Não informado	México	<i>Credit Scoring</i>	MCDA	S	N
17	Hering & Musshoff (2017)	2007–2012	Azerbaijão	<i>Credit Scoring</i>	<i>Logit</i>	N	N
18	Bai et al. (2019)	2009	China	<i>Credit Scoring</i>	<i>Fuzzy Set</i>	S	N
19	Bennouna & Tkiouat (2019)	Não informado	Marrocos	<i>Credit Scoring</i>	An. Discr; <i>Logit</i>	N	N
20	Rao et al. (2020)	2018	China	<i>Credit Scoring</i>	RPA	S	N
21	Baghdasaryan et al. (2021)	2012–2017	Armenia	<i>Credit Scoring</i>	ANN; <i>Logit</i>	N	N
22	Condori-Alejo et al. (2021)	2017–2018	Peru	<i>Credit Scoring</i>	ANN; <i>Logit; RPA</i>	S	N
23	Medina-Olivares et al. (2021)	2017–2018	China	<i>Credit Scoring</i>	<i>Logit</i>	S	N

Esta tabela apresenta o resumo da bibliografia de modelos de probabilidade de *default* de crédito rural a nível individual. 23 trabalhos são referenciados indicando o período dos dados utilizados; região dos dados; modelo de risco aplicado; método; consideração de características específicas ao segmento rural e; consideração de distribuição de probabilidade de *default* com cauda pesada.

Fonte: elaborado pelo autor

A estrutura teórica dos modelos de probabilidade de *default* individual apresentada neste capítulo é relevante ao objetivo desta pesquisa (i.e. identificar qual é a melhor forma de se representar a distribuição de probabilidade de *default* conjunta de um portfólio de crédito rural), pois a estimação da probabilidade de *default* de um portfólio depende da estimação das probabilidades de *default* individuais. Conforme evidenciado, além do baixo enfoque dado ao crédito rural pela literatura, a maior parte dos trabalhos destinados a este segmento não

consideram suas características idiossincráticas, tornando ainda mais relevante que ao menos no nível de risco de portfólio, isto seja abordado e desenvolvido.

Conforme apresentado no Capítulo 2, a performance dos produtores rurais é altamente sujeita ao risco de desastres naturais, sendo estes abordados em estudos de *environmental economics*, os quais identificaram que as distribuições probabilísticas de frequência e de severidade apresentam cauda pesada. Assim, devido à forma da distribuição com que um dos principais fatores de risco dos produtores rurais é estimado, é razoável esperar que as características da distribuição de probabilidade de desastres naturais também se manifestem na probabilidade de *default* de operações de crédito rural.

4. Modelo de PD de crédito rural em nível de portfólio – uma perspectiva regulatória

4.1. Introdução

A estrutura teórica dos modelos de estimação da PD de um portfólio de operações de crédito considera, além das características idiossincráticas das operações, a estrutura de dependência entre elas. Jacobson & Roszbach (2003) afirmam que a estimação da PD individual é uma análise limitada, pois a relação do risco individual das operações com a perda agregada do portfólio não é tão clara.

Se por um lado a estimação da PD a nível individual auxilia as instituições financeiras na decisão de se conceder ou não um crédito (Xia et al., 2021; Zhou et al., 2021) e de se precificar estas operações (Brezigar-Masten et al., 2021; Dinh & Kleimeier, 2007; Featherstone et al., 2006; Luo et al., 2016), por outro a estimação da PD a nível de portfólio contribui para decisões gerenciais de alocação de capital (Bohn & Stein, 2009; Bülbül et al., 2019; Fraise & Laporte, 2022) e é determinante para o cálculo da reserva de capital regulatório (Moreira, 2010; Thomas, 2009).

Os modelos de PD em nível de portfólio podem assumir que o formato da distribuição de probabilidade de *default* conjunta tenha forma simétrica (Accornero et al., 2018; Jacobson & Roszbach, 2003) ou de cauda (Crook & Moreira, 2011; Moreira, 2015; Tente et al., 2019). Neste capítulo, será apresentado como o Comitê de Basileia tem abordado a estimação da PD de portfólio no cálculo do capital regulatório, dando ênfase ao tratamento dado ao crédito rural.

Enfatizar a abordagem da PD no nível de portfólio do Comitê é relevante devido à sua atuação na composição do capital regulatório, além de ampla adoção por órgãos reguladores de 28 países incluindo o Brasil (BCBS, 2021b). Além disso, a estimação da PD em nível de portfólio é utilizada nos modelos gerenciais de alocação de capital, que devem respeitar a restrição de valores mínimos determinados pelo regulador.

4.2. Bank of International Settlements e o Capital Regulatório

A solidez e a resiliência do sistema bancário são fundamentais para o crescimento econômico sustentável, devido à atuação dos bancos como intermediários financeiros prestadores de serviços essenciais aos consumidores, empresas e governos (BCBS, 2011). A composição de uma reserva de capital de risco é o principal instrumento para a promoção destes objetivos. Maiores níveis de capital estão associados à maior estabilidade devido ao efeito *ex-post*, agindo como reserva em momentos adversos, garantido que os bancos cumpram com as suas obrigações, e; *ex-ante*, por disciplinar os acionistas e administradores no compartilhamento de perdas (Mariathasan & Merrouche, 2014; Perotti et al., 2011).

O *Bank of International Settlements* (BIS) é uma organização internacional que auxilia os bancos centrais dos países associados na busca pela estabilidade monetária e financeira, atuando como um fórum de discussão de práticas regulatórias e promovendo a cooperação entre os órgãos reguladores (BCBS, 2019). Os conhecidos tratados internacionais *Acordos de Basileia* são a síntese deste processo. Uma premissa subjacente à elaboração destes tratados é a percepção da necessidade de um acordo de fortalecimento do sistema bancário internacional, de modo que nenhuma instituição financeira poderia escapar de supervisão, devendo tal supervisão ser adequada e consistente entre todas as jurisdições (BCBS, 2015).

Visando à promoção da estabilidade monetária e financeira, os *Acordos de Basileia* exigem que as instituições financeiras mantenham ativos para fazer face às suas perdas para além do esperado (Jorion, 2007; Moreira, 2010; Yanaka, 2009). Estes recursos patrimoniais, que devem ter lastro em ativos garantidores, compõem o que é denominado como capital regulatório, e deve ser igual ou superior ao determinado pelos órgãos reguladores. Este valor

estabelecido como limite inferior é conhecido como capital mínimo requerido. O capital regulatório requisitado pelos órgãos reguladores, diferencia-se do capital econômico, estimado por modelos gerenciais e vinculado à estratégia de gestão de capital e de risco das instituições financeiras.

Chorafas (2004) define o capital regulatório como o valor mínimo necessário para uma instituição financeira ter a sua licença de funcionamento e o capital econômico como a quantidade média de capital necessário para se estar no negócio. Tiesset & Troussard (2005) explicam as diferenças dos tipos de capital devido à natureza de sua utilização: enquanto o capital regulatório tem como função garantir a solvência das instituições, o capital econômico serve a um processo gerencial que influencia as decisões de investimento com o objetivo de maximização de lucro dos acionistas. Por derivar de uma estratégia de negócio, o processo de composição do capital econômico pode ser conflitante com o objetivo de estabilidade financeira do capital regulatório, o que ajuda a explicar a diferença de métricas para dimensioná-los.

Elizalde & Repullo (2007) associam o capital econômico à maximização de uma função de utilidade. Os autores partem da definição de capital econômico como uma quantidade escolhida de maneira discricionária pelos acionistas caso não houvesse a imposição do capital regulatório. Em seguida, constroem um modelo decisório de alocação de capital econômico que considera variáveis de risco (*probability of default*; *loss given default*; e a correlação do ativo com o portfólio) e variáveis da margem de intermediação (custo de capital próprio e de terceiros, e; a taxa de juros dos empréstimos). A conclusão é que se não houvesse a restrição do capital regulatório, o capital econômico aproximar-se-ia do mínimo regulatório somente em um cenário de normalidade, devido à diferente sensibilidade dos tipos de capital em relação às variáveis de risco e de margem de intermediação.

O capital regulatório dos *Acordos de Basileia* evoluiu em relação à tipologia dos riscos e à sua forma de mensuração. Em sua primeira versão, de 1988, o Acordo considerava somente o risco de crédito. Entre os acordos de Basileia I e II houve a incorporação do risco de mercado; em Basileia II, o risco operacional, e; em Basileia III, o risco de liquidez (Gonzales, 2012). Este arcabouço regulatório institui que o capital mínimo requerido de cada banco deve ser uma função de seus riscos idiossincráticos, sendo o risco de crédito o mais significativo nesta composição (BCBS, 2016; Berg & Koziol, 2017), responsável, em média, por mais de 64% da composição da reserva de capital de grandes bancos internacionais¹ (BCBS, 2021a).

4.3. Acordo de Basileia I

Em junho de 1974, o banco *Bankhaus Herstatt* foi liquidado pelo órgão regulador alemão após sofrer elevadas perdas cambiais que o tornaram insolvente. Em outubro do mesmo ano, e também relacionado a perdas cambiais, o banco norte-americano *Franklin National Bank* tornou-se insolvente, sendo a maior falência de uma instituição financeira na história dos Estados Unidos à época (McNamara et al., 2015).

Criado em 1974, o Comitê de Regulação Bancária e Práticas de Supervisão de Basileia, posteriormente renomeado para Comitê de Supervisão Bancária de Basileia, foi uma reação dos países do G-10 à instabilidade presenciada nos mercados financeiros globais (BCBS, 2015), tendo como objetivo estabelecer diretrizes para a cooperação entre autoridades monetárias e promover a melhoria nas práticas de regulação (BCBS, 1975). Na década de 1980, as mudanças estruturais nos mercados financeiros globais, com a internacionalização dos bancos e o crescimento dos mercados de derivativos, elevaram a preocupação dos órgãos reguladores do sistema financeiro com a solidez de seus regulados (Crouhy et al., 2014).

¹ Baseado em uma amostra de 105 bancos com saldo de patrimônio líquido e lucros retidos superior à 3 bilhões de euros.

Em 1988, o Comitê de Basileia elabora o Acordo de Basileia I. Nele, pactuou-se a mensuração de um capital mínimo, necessário para as instituições financeiras protegerem-se da exposição ao risco de crédito de seus ativos. A convergência regulatória tinha como finalidades: (i) reforçar a resiliência e estabilidade do sistema bancário internacional, e; (ii) reduzir a vantagem comparativa das instituições financeiras proveniente da assimetria regulatória entre as diversas autoridades monetárias (BCBS, 1988). Para isto, o Acordo definiu três fatores: (i) composição do capital regulatório; (ii) sistema de ponderação de risco, e; (iii) taxa mínima do capital requerido.

O capital regulatório foi dividido em dois níveis, de acordo com a fonte de recursos que o compõem: *Tier 1* (ou *Core Capital*) e o *Tier 2* (ou *Supplementary Capital*). O *Tier 1* é composto pelo capital próprio e reservas incluídas no balanço (*disclosed reserves*), com as seguintes deduções: (i) *Goodwill* (ativo intangível representado pelo ágio por expectativa de rentabilidade futura em combinações de negócios), e; (ii) investimentos em empresas não consolidadas no balanço. O *Goodwill* é deduzido por ser considerado um elemento de capital de avaliação subjetiva e de baixo valor em caso de liquidação da companhia. Já a dedução das empresas não consolidadas no balanço evita que mais de uma instituição financeira considere tais ativos em sua composição de capital (Balthazar, 2006). O *Tier 2* inclui ativos com menor previsibilidade de realização, permitindo em sua composição as reservas não incluídas no balanço (*undisclosed reserves*); reservas de reavaliação de ativos; provisões; instrumentos de capital híbridos e dívidas subordinadas. O Acordo determina que o capital regulatório deveria ser composto de pelo menos 50% de capital do *Tier 1*.

O segundo fator refere-se à forma de mensuração dos riscos conforme a contraparte e taxonomia contábil dos ativos, dividindo-os em cinco faixas que vão de 0 a 100% apresentadas na Tabela 2.

Tabela 2. Ponderação de Risco dos Ativos – Acordo de Basileia I

Fator de ponderação	Ativo
	(a) Caixa
0%	(b) Ativos em Governos e Bancos Centrais denominados e financiados em moeda nacional* (c) Ativos em Governos e Bancos Centrais de países membros da OECD*
	(a) Ativos em bancos de países membros da OECD e bancos multilaterais de desenvolvimento*
20%	(b) Ativos em bancos de países não pertencentes à OECD com <i>maturity</i> inferior a um ano (c) Empréstimos com prazo inferior a um ano garantidos por bancos de países não pertencentes à OECD (d) Ativos de entidades não domésticas do setor público de países membros da OECD* excluindo o governo central (e) Itens de caixa em processo de coleta
50%	(a) Empréstimos totalmente garantidos por hipotecas residenciais
100%	(a) Ativos do setor privado (b) Ativo imobilizado (c) Ativos Imobiliários (d) Todos os outros ativos (e.g. ativos de crédito rural)

Esta tabela apresenta o fator de ponderação de risco conforme o tipo de ativo sob o escopo do Acordo de Basileia I.

* Ou por estes garantido.

Fonte: elaborada pelo autor a partir de BCBS (1988).

Assim, os ativos provenientes de operações de crédito rural eram enquadrados na categoria “todos os outros ativos” e ponderados por 100% do seu valor. O sistema de ponderação de risco proposto também considera as exposições fora do balanço (*off-balance*

sheet exposures). Para Balthazar (2006), esta é uma inovação do modelo em relação ao que era até então praticado. Durante a década de 1980, houve um rápido aumento deste tipo de exposição (Hull, 1989) por produzir retornos com baixo ou nenhum comprometimento de capital (Lewis, 1988).

O último item abordado pelo Acordo é a taxa mínima do capital requerido, também conhecida como Índice de Basileia ou *Capital Adequacy Ratio* (CAR). Ela é o quociente da divisão do requisito de capital regulatório pelos ativos ponderados pelo risco (*risk-weighted assets* – RWA), sendo determinado que o seu valor deve ser maior ou igual a 8%, respeitando a restrição de que ao menos metade do capital regulatório deve ser composto pelo *Tier 1 Capital*, conforme a Equação 1.

$$CAR = \frac{\text{Capital Regulatório}}{\text{RWA}}, \quad (1)$$

de forma que $CAR \geq 8\%$ e capital regulatório $\geq 50\%$ do *Tier 1*. Assim, caso o *Capital Adequacy Ratio* (CAR) seja fixado, um incremento nos ativos ponderados pelo risco (RWA) implica em um aumento de igual proporção no requisito de capital regulatório. A implementação dessas três abordagens (composição do capital regulatório; sistema de ponderação de risco, e; taxa mínima do capital requerido) formam a estrutura de capital regulatório de Basileia I.

Exemplificando a aplicação do modelo, suponha que uma instituição possua três ativos: saldo em caixa; operações de crédito imobiliário com garantia hipotecária e; operações de crédito rural, com os respectivos valores (x, y, z) , o requisito de capital regulatório, com base nos ativos ponderados pelo risco e um CAR de 8% é de:

$$\text{Capital Regulatório} = [(x \times 0\%) + (y \times 50\%) + (z \times 100\%)] \times 8\% \quad (2)$$

O Comitê de Basileia tinha ciência sobre certas limitações do Acordo. Ao tratar somente do risco de crédito, o Comitê apontava que outros tipos de risco deveriam ser considerados pelos reguladores locais (BCBS, 1988). À época, outras limitações não foram identificadas no tratado. Atualmente, devido ao processo dinâmico de inovações dos mercados financeiros, dispondo de técnicas avançadas de modelagem, além da experiência histórica da implementação das mecânicas de cálculo, é de razoável consenso na literatura que as limitações metodológicas tornaram-se latentes com o passar do tempo (Balthazar, 2006).

Contudo, apenas quatro anos após a publicação do Acordo, Iken (1992) já defendia o cálculo do capital regulatório pela ponderação de risco dos ativos, mas criticava o modelo proposto por Basileia I. O autor comparou o custo de oportunidade do provisionamento de capital a um prêmio de seguro contra perdas financeiras pago pelos bancos. Ocorrendo sinistro, o capital regulatório seria consumido e, caso não fosse suficiente, o excedente de perdas seria absorvido pelo governo. Assim, o requisito de capital reduz o custo social de falências bancárias, sendo a princípio interessante ao governo a sua elevação. Entretanto, requisitos de capitais elevados geram consequências negativas à viabilidade da intermediação financeira e à economia real, sendo necessário uma medida de risco eficiente que gere solidez às instituições sem onerar o seu custo de capital (Borgo, 2020).

Segundo Iken (1992), o problema da ponderação dos ativos pelo risco proposta pelo Acordo de Basileia I é que ela homogeneiza o requisito de capital dos ativos, pelo tipo de entidade emissora. Como exemplo, dois títulos emitidos por bancos comerciais com riscos diferentes poderão ter o mesmo efeito no capital requerido. Como o prêmio a ser pago era o mesmo, criar-se-ia um incentivo à aquisição de ativos com maiores níveis de risco e retorno, opondo-se à busca por maior solidez dos sistemas financeiros. Por fim, outra falha reside em considerar o risco individual dos ativos, ignorando o potencial efeito da diversificação de portfólio na redução do risco da carteira.

Crouhy et al. (2014) elencam outras deficiências do primeiro Acordo. Os autores consideram que a ponderação de risco proposta foge da realidade por não considerar o risco específico de cada contraparte. Isto cria distorções não apenas dentro do mesmo grupo (governos, bancos e corporações), pois, independente do perfil individual, terão o mesmo risco, como também entre os grupos. A razão é que a escala de ponderação considera que ativos emitidos por governos possuem menor ponderação de risco do que ativos emitidos por bancos e estes por sua vez possuem menor ponderação de risco do que as corporações. Entretanto, é esperado que o risco de um grande banco ou corporação seja menor do que o de um título soberano de algum país periférico.

Outra crítica ao primeiro Acordo é a abordagem superficial do efeito da *maturity* dos ativos em relação ao seu risco: a *maturity* influencia apenas duas classes de ativos, dividindo-os em operações com menos de um ano e com mais de um ano. Deste modo, o peso de risco para uma operação com *maturity* de dois anos é o mesmo do que de operações de mais de vinte anos, por exemplo. Logo, não é razoável a concepção de que um ativo de longo prazo tenha a mesma previsibilidade em sua realização do que um ativo de curto prazo com mais de um ano. Por fim, o modelo de risco proposto pelo Acordo não incentivou adoção de técnicas de mitigação de risco.

Em 1996, foi publicado o documento *Amendment to the Capital Accord to incorporate market risks* (BCBS, 1996), introduzindo a necessidade de carregamento de capital para o risco de mercado nas seguintes exposições: taxas de juros; ações; câmbio, e; *commodities*. A emenda inovou ao permitir que os bancos optassem entre duas abordagens para mensurar o seu risco de mercado: a abordagem padronizada (*standardized*) ou a abordagem por modelos internos (*internal models*). Na abordagem padronizada, o requerimento de capital é determinado de acordo com a categoria do ativo, em um processo semelhante ao já praticado para o risco de crédito. Entretanto, há maior atenção ao impacto da *maturity* no risco das operações.

A segunda opção de abordagem, por modelos internos, permite que os bancos utilizem modelos próprios mediante aprovação do regulador que estabelece a necessidade de cumprimento de requisitos, premissas e parâmetros. Nestes modelos, é possível considerar os efeitos da diversificação de ativos no portfólio, o que incentiva práticas de gerenciamento de risco (Balthazar, 2006; BCBS, 1996; Crouhy et al., 2014; McNamara et al., 2015).

Um ano antes da emenda ao acordo de capital, o Comitê de Basileia publicou um artigo (BCBS, 1995) abordando o desenvolvimento de uma abordagem de ponderação de risco dos ativos por meio modelos internos. Tal documento foi uma resposta à indústria bancária sobre uma proposta de implementação de um requisito de capital padronizado para o risco de mercado. Segundo a publicação, diversos bancos atestaram que seus modelos internos possuíam maior precisão na mensuração dos riscos, assim como que a proposta de uma abordagem padronizada não proveria incentivos às práticas de gerenciamento de risco. Adicionalmente, não levava em conta a diversificação dos ativos nem era compatível com os sistemas internos de mensuração de risco dos bancos (BCBS, 1995). A evolução da regulação prudencial de Basileia teve como um de seus pilares o aprimoramento das abordagens por modelos internos.

4.3. Acordo de Basileia II

A atualização do Acordo de Basileia ocorreu em 2004, com a publicação do *Revised Capital Framework*, fruto de um processo consultivo de seis anos com os países membros do Comitê e seus supervisores locais. O novo acordo objetivou tornar o requisito de capital mais sensível aos riscos, devido às críticas sobre o requisito de capital do primeiro acordo (Bliss, 2002). Os avanços de Basileia II estão em aumentar a precisão do agrupamento de devedores em classes de risco (Katchova & Barry, 2005). Para reforçar a promoção da estabilidade financeira, a nova estrutura proposta, conhecida como *Acordo de Basileia II*, adicionou ao

capital mínimo requerido de Basileia I orientações para a prática da supervisão bancária e de disciplina de mercado. O requisito de capital, as práticas de supervisão e a disciplina de mercado formam os três pilares de Basileia II.

As inovações no capital mínimo requerido ocorreram em duas frentes: (i) o refinamento das metodologias de ponderação de risco, e; (ii) a inclusão de capital requerido para o risco operacional, juntando-se com os requisitos já existentes de risco de crédito e mercado. Em Basileia I, o capital regulatório para fazer face ao risco de crédito era obtido por uma única abordagem estabelecida pelo Comitê. O Acordo de Basileia II possibilita que os bancos escolham entre duas abordagens para a ponderação deste risco: a padronizada (*Standardized Approach*, SA) e a por modelos internos (*Internal Rating-Based*, IRB), trazendo maior convergência entre o capital regulatório e o capital econômico (Tieset & Troussard, 2005).

4.3.1. Abordagem Padronizada (SA)

A abordagem padronizada se aproxima da abordagem de Basileia I, mas inova ao considerar que a ponderação de risco dos ativos é uma função não só das características da contraparte (se são classificados como governos, bancos, empresas ou pessoas físicas), como também do risco atribuído à operação por uma agência de *rating* externa (Balthazar, 2006). A inclusão do *rating* na ponderação faz com que a abordagem seja mais sensível ao risco. Para exemplificar, uma dívida corporativa em Basileia I seria ponderada a 100% do valor de sua exposição, independente do risco da contraparte. Na abordagem padronizada de Basileia II, a ponderação de risco de uma dívida corporativa varia entre 20%, para dívidas com classificações superiores à “AA-”², e 150%, para classificações inferiores à “BB-”. A maior estratificação desta abordagem trouxe regras específicas para as exposições soberanas, bancárias, corporativas e de varejo, sem dar um direcionamento para exposições em ativos de crédito rural. Embora tenha apresentado uma evolução em relação ao Acordo de Basileia I, a abordagem padronizada continuou a onerar o requisito de capital (E. Altman, Saunders, & Bharath, 2002).

4.3.2 Abordagem por modelos internos (IRB)

A principal motivação por trás da implementação do IRB advém da hipótese da informação interna, pela qual entende-se que os bancos possuem melhores condições para mensurar os próprios riscos em comparação aos agentes externos (Beltratti & Paladino, 2016). Entretanto, há na literatura estudos sobre os efeitos adversos desta abordagem, relacionados à uma maior opacidade e heterogeneidade no cálculo do risco das instituições financeiras (Avramova & le Leslé, 2012; Beltratti & Paladino, 2016; EBA, 2013; Haldane, 2011; Mariathasan & Merrouche, 2014).

Mais complexa do que a abordagem padronizada, a abordagem por modelos internos (IRB) mensura o requisito de capital por meio de uma função de ponderação de risco, a qual tem ao menos um de seus parâmetros calculado pelos bancos. A aplicação dos parâmetros de risco à função de ponderação é subordinada à classificação dos ativos em uma das cinco classes determinadas: corporativo; soberano; bancário; varejo, e; ações.

É interessante notar que o Acordo não deu nenhum tratamento específico aos empréstimos rurais, havendo a possibilidade de classificá-los como dívida corporativa (subcategoria *commodities*) ou operação de varejo (Bandyopadhyay, 2007; Kim, 2005). Segundo Kim (2005), os empréstimos rurais são, em sua maioria, destinados à indivíduos ou pequenas empresas e as operações individuais são passíveis de serem agregadas em grandes

² O Acordo foi redigido com base nas classes de *rating* da agência de classificação de risco *Standard & Poor's*. Notas de outras agências de classificação de risco são aceitas, por meio de critérios de equivalência.

grupos (*pools*). Essas características são mais adequadas à classe de varejo do que à classe corporativa.

A função de ponderação de risco possui os seguintes parâmetros: *probability of default* (PD); *loss given default* (LGD); *exposure at default* (EAD); *maturity* (M); e coeficiente de correlação linear (ρ). A Tabela 3 apresenta as possibilidades de estimação dos parâmetros, de acordo com a classe dos ativos.

Tabela 3. Fontes de estimação de parâmetros para IRB – Acordo de Basileia II

Exposição	IRB <i>Foundation</i>		IRB <i>Advanced</i>	
	Modelos Internos	Regulador	Modelos Internos	Regulador
Corporativa				
Soberana	PD	LGD, EAD, M	PD, LGD, EAD, M	-
Bancária				
Varejo		PD, LGD EAD, M		
Ações	PD/LGD <i>Approach</i> ou <i>Market-Based Approach</i>			

Esta tabela apresenta quais parâmetros podem ser estimados pelas instituições financeiras e quais parâmetros são determinados pelo regulador para as abordagens por modelos internos IRB *Foundation* e IRB *Advanced* para cada tipo de exposição (corporativa, soberana, bancária, varejo e ações). Nota: o coeficiente de correlação linear (ρ) sempre é determinado pelo regulador.

Fonte: adaptado de Balthazar (2006).

As exposições corporativas, soberanas e bancárias possuem duas opções de IRB: a *foundation* e a *advanced*. Na IRB *foundation*, os bancos estimam a PD por modelos internos e os outros parâmetros de risco são determinados pelo órgão regulador local. Na IRB *advanced* todos os parâmetros de risco são estimados por modelos internos. As exposições em varejo devem ter todos os seus parâmetros de risco calculados pela instituição. Por fim, as exposições em ações possuem abordagens específicas: a *PD/LGD Approach*, que se assemelha à IRB *foundation*, e; a *Market-Based Approach*, que é calculada pela instituição, considerando a taxa livre de risco. Com exceção das exposições em ações calculadas pela *Market-Based Approach*, todas as outras exposições possuem a mesma função de ponderação de risco, divergindo somente na magnitude da correlação linear entre os ativos (ρ), sempre determinado pelo regulador.

4.3.3 Função de ponderação de risco da abordagem IRB

A função de ponderação de risco da abordagem IRB (RWA_{IRB}) é uma adaptação do modelo de único fator do risco de *default* de um portfólio de Vasicek (Vasicek, 1984, 1987), que, por sua vez, baseia-se na estrutura teórica de Merton (Merton, 1973). Os modelos fatoriais estimam uma relação linear entre uma variável explicada e uma ou mais variáveis explicativas (fatores). Aplicado à Finanças, estes modelos associam retornos de um conjunto de ativos aos fatores comuns que eles possuem, estimando a volatilidade do portfólio e a correlação entre os *defaults* (Bessis, 2015; Bluhm et al., 2010).

Conforme apresentado, o capital regulatório tem como função proteger as instituições financeiras das perdas para além do esperado (Yanaka, 2009). Uma perda é considerada além da esperada quando o seu valor observado é maior do que a expectativa monetária formada *ex-ante* para a perda futura, sob algum nível de confiança e conjunto informacional disponível (Jorion, 2007; Moreira, 2010). Sob esta dicotomia, a perda total (TL) efetiva de um portfólio de crédito pode ser entendida como o resultado da soma da perda esperada (EL) com a perda para além do esperado (UL):

$$TL = EL + UL \quad (3)$$

Na RWA_{IRB} , partindo do modelo de Vasicek (Vasicek, 1984, 1987), a perda total de um conjunto de operações é estimada com base na PD, no LGD e no coeficiente de correlação linear, assumindo que a distribuição de probabilidade da PD seja Normal, sob um intervalo de confiança de 99,9%:

$$TL = \Phi \left(\frac{\Phi^{-1}(PD) + \sqrt{\rho} \times \Phi^{-1}(0,999)}{\sqrt{1-\rho}} \right) \times LGD, \quad (4)$$

em que Φ é a função acumulada da distribuição Normal, Φ^{-1} é a função inversa da função acumulada da distribuição Normal e ρ o coeficiente de correlação linear. Já a perda esperada é o produto da PD com a LGD:

$$EL = PD \times LGD, \quad (5)$$

substituindo as Equações (4) e (5) na Equação (3):

$$UL = \left(LGD \times \Phi \left(\frac{\Phi^{-1}(PD) + \sqrt{\rho} \times \Phi^{-1}(0,999)}{\sqrt{1-\rho}} \right) \right) - (LGD \times PD), \quad (6)$$

por fim, a RWA_{IRB} ajusta a perda para além do esperado (UL) à *maturity* (M) do portfólio, resultando na equação final proposta pelo Acordo de Basileia:

$$RWA_{IRB} = \left[\left(LGD \times \Phi \left(\frac{\Phi^{-1}(PD) + \sqrt{\rho} \times \Phi^{-1}(0,999)}{\sqrt{1-\rho}} \right) \right) - (LGD \times PD) \right] \times M, \quad (7)$$

chegando ao valor do capital regulatório (mínimo requerido):

$$Capital\ Regulat\acute{o}rio = RWA_{IRB} \times EAD \times CAR \quad (8)$$

Assim, o modelo de PD no nível de portfólio do Acordo de Basileia diferencia o tratamento dado aos tipos de operação de crédito (corporativo; soberano; bancário, e; varejo) por meio do coeficiente de correlação linear ρ . Como não há um tratamento específico para o crédito rural, a estimação da PD de um portfólio desta linha de crédito deve ser elaborada com base nos coeficientes ρ das exposições corporativas e de varejo, definidos pelo Acordo como:

$$\rho_{Corporativa} = 0,12 \times \left(\frac{1 - e^{(-50 \times PD)}}{1 - e^{(-50)}} \right) + 0,24 \times \left(1 - \frac{1 - e^{(-50 \times PD)}}{1 - e^{(-50)}} \right) \quad (9)$$

$$\rho_{Varejo} = 0,15 \quad (10)$$

Moreira (2010) indica que este modelo pode não ser adequado. A razão é que se não for válida a premissa de Normalidade para distribuição de probabilidade da PD, há a possibilidade de desbalanceamento no requisito de capital: excessivo em cenários de expansão econômica e insuficiente em cenários de estresse. Outro potencial fator de erro na estimação do capital regulatório é a estrutura de dependência entre as operações, engessada pelas Equações (9) e (10) para os casos de dívidas corporativas e de varejo. Mesmo que a estrutura de dependência seja linear, o coeficiente de correlação pode divergir do valor previamente estabelecido pela norma. Ademais, caso a distribuição da PD conjunta possua assimetrias ou curtoses elevadas, a estimação da estrutura de dependência entre as operações via coeficientes de correlação linear é inadequada.

Sob a ótica crédito rural, tais fatores são ainda mais sensíveis. Apesar das suas especificidades, tal modalidade de crédito não recebeu nenhum tratamento. Se os coeficientes ρ de dívida corporativa e de varejo não forem aderentes, sua utilização para o crédito rural pode não refletir adequadamente a realidade. Outro agravante advém da possibilidade da distribuição

de PD conjunta do crédito rural apresentar caudas pesadas, influenciada pelo risco de desastres naturais a que os produtores estão expostos.

A análise comparativa do poder preditivo de modelos de PD conjunta de crédito rural, assumindo distribuições de probabilidade simétricas e de caudas pesadas, proposta por este trabalho, ajudará na avaliação da adequação do requerimento de capital regulatório para o crédito rural. Caso os resultados obtidos indiquem a melhor adequação de modelos de probabilidade conjunta que apresentem caudas pesadas, restará evidente a necessidade de o modelo regulatório de Basileia (i) considerar estruturas de dependência não lineares e distribuições de probabilidade de *default* de cauda, e; (ii) dar um tratamento específico para o crédito rural.

5. Modelos de Cópulas: aspectos teóricos e aplicações

5.1. Definições e Propriedades Básicas das Cópulas

As cópulas são funções estatísticas que generalizam o conceito de medidas de dependência por meio da associação de distribuições de probabilidade conjuntas às suas respectivas distribuições de probabilidade marginais (Nelsen, 2006). Estas características fazem as cópulas serem de interesse dos estatísticos e pesquisadores de áreas aplicadas por duas principais razões: são uma forma de se estudar medidas de dependência livres de escala e são um ponto de partida para a construção de distribuições bivariadas (Fisher, 2006). A teoria de cópulas é remetida aos trabalhos basilares de Fréchet (1951) e Sklar (1959). Também contribuíram para o seu desenvolvimento os trabalhos de Hoeffding (1940, 1941, 1942), pela abordagem de propriedades básicas de distribuições de probabilidade conjuntas. Devido ao fato de estas publicações terem ocorrido durante a Segunda Guerra Mundial, em periódicos alemães pouco conhecidos pela comunidade estatística, elas foram menos conhecidas (Fisher, 2006). Atualmente, estes trabalhos estão acessíveis na compilação e tradução para o inglês da obra de Hoeffding feita por Fisher & Sen (1994).

As definições, propriedades básicas e resultados gerais da teoria de Cópulas apresentadas nesta seção seguem de perto Nelsen (2006), no qual todas as demonstrações dos teoremas podem ser encontradas. Eventualmente, as definições de Nelsen (2006) serão complementadas por definições de outros autores oportunamente citados.

Sejam $F(x, y)$ a função de distribuição conjunta entre duas variáveis aleatórias X e Y dada por $F(x, y) = \mathbb{P}[X \leq x, Y \leq y]$, $F_X(x) = \mathbb{P}[X \leq x]$ e $F_Y(y) = \mathbb{P}[Y \leq y]$, as distribuições marginais de X e Y . Denotando $u = F_X(x)$ e $v = F_Y(y)$, uma cópula é uma função $C: [0,1]^2 \rightarrow [0,1]$ tal que

$$F(x, y) = C(F_X(x), F_Y(y)) = C(u, v) \quad (11)$$

Assim, a função cópula associa a distribuição conjunta às distribuições marginais unidimensionais uniformes, para qualquer forma funcional das distribuições $F_X(x)$ e $F_Y(y)$, devido à Transformada Integral de Probabilidade. Baseando-se em Anjos et al. (2004), dado que X é uma variável aleatória com distribuição $F_X(x) = \mathbb{P}[X \leq x]$, se $F_X(x)$ for contínua, então para $u \in (0,1)$ existe um valor mínimo único $x(u)$, tal que $F_X(x(u)) = u$. Assim,

$$x(u) = F_X^{-1}(u) = \inf \{x: F_X \geq u\}, \quad (12)$$

sendo $F_X^{-1}(u)$ a função de distribuição inversa. Assim, $F_X(x) \leq u$ se e somente se $x \leq F_X^{-1}(u)$, condição atingida se $F_X(x)$ for contínua e não-decrescente.

Dado que $F_X(x)$ é contínua e não-decrescente, $F_X^{-1}(u)$ também será, com $u \in (0,1)$. Então,

$$\mathbb{P}[F_X(x) \leq u] = \mathbb{P}[X \leq F_X^{-1}(u)] = F_X(F_X^{-1}(u)) = u \quad (13)$$

Deste modo, $F_X(x)$ tem distribuição uniforme em $(0,1)$, i.e., $F_X(x) \sim U(0,1)$. A transformação $U = F_X(x)$ é conhecida como *Transformada Integral de Probabilidade*. As cópulas transformam a distribuição conjunta das variáveis aleatórias em uma distribuição uniforme, i.e., $C: [0,1]^n \rightarrow [0,1]$, com $n \geq 2$. A intenção é criar um sistema de transformação invariante em escala, de forma que qualquer variável aleatória tenha a sua distribuição marginal representada em uma distribuição uniforme entre 0 e 1 (Fisher, 2006). Uma condição necessária para que a cópula produza essa transformação é a de que a função de distribuição de probabilidade conjunta também atenda às condições de continuidade e não decrescimento da

transformada integral de probabilidade (Hoeffding, 1940). Para maiores detalhes sobre essas propriedades ver Nelsen (2006).

Entre os anos de 1959 e 1974, a maior parte das pesquisas envolvendo cópulas ocorreu no desenvolvimento da teoria do espaço métrico probabilístico, sobretudo no estudo de operações binárias no espaço de distribuições de probabilidade. Então, foi descoberto que cópulas bidimensionais poderiam ser usadas para estimar medidas de dependência não paramétricas entre pares de variáveis aleatórias (Schweizer, 1991). O ponto de partida da aplicação da teoria de copulas às finanças é a interpretação probabilística entre a relação das funções cópulas e as funções de distribuição de probabilidade. O Teorema de Sklar é um dos pilares desta interpretação. O Teorema mostra que, não só as cópulas são funções de distribuições conjuntas, como também o inverso: as funções de distribuição conjunta podem ser reescritas como cópulas (Cherubini et al., 2004).

Teorema (Sklar, (1959)). Seja H uma função de distribuição conjunta com as distribuições marginais F e G . Então, existe uma cópula C para qualquer x, y em \bar{R} ,

$$H(x, y) = C(F(x), G(y)) \quad (14)$$

Se F e G são contínuas, então a cópula C é única, caso contrário, C é determinado no conjunto imagem $RanF \times RanG$. Reciprocamente, se C é uma cópula e F e G são funções de distribuição de probabilidade, então a função $H(x, y)$ definida por (15) é uma função de distribuição de probabilidade conjunta de F e G .

Demonstração. Ver Nelsen (2006).

A parte recíproca do teorema de Sklar é importante para a modelagem de distribuições multivariadas, pois implica que podemos acoplar distribuições univariadas, de quaisquer tipos, com qualquer cópula, e teremos definida uma distribuição multivariada válida (Anjos et al., 2004).

5.2. Medidas de Dependência

A seção 5.1 apresentou a estrutura teórica das funções de cópulas, evidenciando como estas funções estimam a distribuição de probabilidade conjunta de variáveis aleatórias a partir de distribuições marginais quaisquer. A dependência entre as variáveis aleatórias é totalmente definida pela função de distribuição conjunta dessas variáveis, no entanto, a distribuição conjunta não informa especificamente como é esta dependência nem a mensura (Anjos et al., 2004). Ademais, para ajustar uma função de cópulas a um conjunto de dados, a estimação da estrutura de dependência entre as variáveis se faz necessária (Carvalho, 2011; Carvalho & Chiann, 2013). As medidas de dependência são formas de se estimar esta associação.

No campo das finanças modernas, as medidas de dependência são um tema recorrente e passíveis de incompreensão. Parte desta incompreensão advém do uso literal da palavra correlação para se referir a toda relação de dependência, enquanto para estatísticos, a correlação é apenas uma medida de dependência estocástica particular, entre outras (Embrechts et al., 2002). A dependência entre variáveis é multifacetada, passível de ser observada sob diversas óticas como a correlação linear, medidas de concordância e dependência de cauda.

Ter ciência plena dos conceitos, especificidades e significados das medidas de dependência é fundamental para a escolha de qual medida utilizar na construção de cópulas, pois cada medida é apropriada para mensurar um aspecto da dependência e que, caso seja aplicada para mensurar um aspecto diverso, pode levar a conclusões e interpretações errôneas (Anjos et al., 2004). Esta seção apresentará as características gerais das medidas de dependência: coeficiente de Pearson; medidas de concordância, e; dependência de cauda.

5.2.1. Coeficiente ρ de Pearson

O coeficiente de Pearson, também conhecido como coeficiente de correlação linear é uma medida paramétrica de dependência linear entre as variáveis, sendo invariante sob transformações lineares crescentes. É uma medida de dependência natural nas distribuições Normais multivariadas. Ele é definido como:

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}, \quad (15)$$

em que a $Cov(X, Y)$ a covariância entre X e Y , $Var(X)$ a variância de X e $Var(Y)$ a variância de Y . Uma das limitações do coeficiente de correlação linear é que ele não é invariante sob transformações não lineares estritamente crescentes ($T: \mathbb{R} \rightarrow \mathbb{R}$), ou seja:

$$\rho(T(X), T(Y)) \neq \rho(X, Y) \quad (16)$$

Outra limitação do coeficiente de Pearson ocorre com a análise do coeficiente $\rho(X, Y) = 0$. A independência entre duas variáveis aleatórias (X, Y) implica que elas não possuem correlação, logo, $\rho(X, Y) = 0$. Entretanto, $\rho(X, Y) = 0$ não necessariamente implica em independência. Um contraexemplo clássico no qual a correlação desaparece mesmo havendo estrita dependência entre as variáveis aleatórias é: dado $X \sim N(0, 1)$ e $Y = X^2$, então $\rho(X, Y) = 0$ (Embrechts et al., 2002).

5.2.2. Medidas de Concordância

As medidas de concordância, também conhecidas como correlações de classificação (ou postos), são alternativas que visam tratar as limitações do coeficiente de correlação linear. Variáveis aleatórias são consideradas concordantes quando a posição de suas observações concordam em termos de escala, ou seja, maiores valores de uma distribuição são observados em conjunto com maiores valores da outra distribuição, ocorrendo de forma análoga com observações de valores baixos e intermediários. De maneira mais formal, considere (x_i, y_i) e (x_j, y_j) duas observações do vetor (X, Y) de variáveis aleatórias contínuas. Dizemos que (x_i, y_i) e (x_j, y_j) são:

1. *concordantes* se $(x_i < x_j)$ e $(y_i < y_j)$ ou se $(x_i > x_j)$ e $(y_i > y_j)$,
2. *discordantes* se $(x_i < x_j)$ e $(y_i > y_j)$ ou se $(x_i > x_j)$ e $(y_i < y_j)$.

Tal relação também pode ser apresentada na forma de produtos:

$$(x_i - x_j)(y_i - y_j) > 0, \text{ para concordância}, \quad (17)$$

$$(x_i - x_j)(y_i - y_j) < 0, \text{ para discordância}. \quad (18)$$

Devido à invariância das cópulas às transformações monotônicas (Nelsen, 2006), a inferência sobre medidas de dependência baseadas medidas de concordância permite a comparação da dependência entre variáveis aleatórias com diferentes distribuições marginais (Genest et al., 2011). As principais medidas de concordância comumente relacionadas às cópulas – coeficiente τ de Kendall e coeficiente ρ de Spearman – serão detalhadas a seguir.

5.2.2.1. Coeficiente τ de Kendall

O coeficiente τ de Kendall é definido como a diferença entre as probabilidades de concordância e discordância. Em sua versão amostral, o coeficiente é calculado da seguinte maneira: seja $\{(x_1 - y_1), (x_2 - y_2) \dots (x_n - y_n)\}$ uma amostra aleatória de n observações do vetor (X, Y) de variáveis aleatórias contínuas, havendo $\binom{n}{2}$ pares distintos (x_i, y_i) e (x_j, y_j) , que podem ser concordantes ou discordantes. Seja c o número de pares concordantes e d o número de pares discordantes, então, o coeficiente τ de Kendall representado por τ , é:

$$\tau = \frac{c-d}{c+d} = \frac{c-d}{\binom{n}{2}} \quad (19)$$

Para a versão populacional, o coeficiente τ de Kendall é obtido por meio do cálculo da diferença da probabilidade das equações (20) e (21). Para isto, considerando que (X_1, Y_1) e (X_2, Y_2) são vetores independentes e identicamente distribuídos, a diferença entre as probabilidades de concordância e discordância destes vetores é:

$$\tau_{X,Y} = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0] \quad (20)$$

Dado que as probabilidades de pares concordantes e discordantes são complementares, a Equação (23) pode ser reescrita como:

$$\tau = \tau_{X,Y} = -1 + 2P[(X_1 - X_2)(Y_1 - Y_2) > 0] \quad (21)$$

Se (X, Y) forem variáveis aleatórias contínuas, podemos apresentar o coeficiente τ de Kendall como uma função da cópula C que associa as distribuições de probabilidade marginais de X e Y :

$$\tau = \tau_{X,Y} = -1 + 4 \cdot \int_{[0,1]^2} C(u, v) dc(u, v), \quad (22)$$

sendo u e v as Transformadas Integrais de Probabilidade $u = F(x)$ e $v = G(y)$.

5.2.2.2. Coeficiente ρ de Spearman

Assim como o coeficiente τ de Kendall, o coeficiente ρ de Spearman é uma medida de concordância que mensura a correlação de classificação entre duas variáveis, ou seja, a associação das variáveis em função não apenas do seu valor, mas também da sua posição em sua distribuição de probabilidade.

Conforme apresentado, o coeficiente τ de Kendall pode ser definido como a diferença da probabilidade de concordância e discordância entre os vetores (X_1, Y_1) e (X_2, Y_2) . Por sua vez, o coeficiente ρ de Spearman mensura a diferença de probabilidade de ocorrência de concordância e discordância entre os vetores (X_1, Y_1) e (X_2, Y_3) , um par de vetores com as mesmas margens. Entretanto somente (X_1, Y_1) possui a distribuição de probabilidade conjunta H , enquanto (X_2, Y_3) é formado por variáveis independentes:

$$\rho_{X,Y} = 3(P[(X_1 - X_2)(Y_1 - Y_3) > 0] - P[(X_1 - X_2)(Y_1 - Y_3) < 0]) \quad (23)$$

Quando (X, Y) forem variáveis aleatórias contínuas, podemos apresentar o coeficiente ρ de Spearman como uma função da cópula C que associa as distribuições de probabilidade marginais de X e Y :

$$\rho_{X,Y} = 12 \int_{[0,1]^2} uv dC(u, v) - 3, \quad (24.1)$$

$$\rho_{X,Y} = 12 \int_{[0,1]^2} C(u, v) dudv - 3, \quad (24.2)$$

sendo u e v as Transformadas Integrais de Probabilidade $u = F(x)$ e $v = G(y)$.

5.2.3. Medidas de Dependência de Cauda (λ_u, λ_l)

Assim como as medidas de concordância, as medidas de dependência de cauda mensuram como um conjunto formado pelo par de variáveis aleatórias evolui em quantis elevados. Contudo, neste caso, o objetivo é mensurar a dependência das distribuições em situações extremas (McNeil et al., 2005), identificadas nos quadrantes superior direito e inferior esquerdo, do quadrado unitário da distribuição conjunta (Nelsen, 2006).

Os coeficientes de dependência de cauda são obtidos ao avaliarmos a probabilidade condicional de as distribuições marginais excederem um limiar de quantis extremos. Dado um par de variáveis aleatórias (X, Y) , com distribuições de probabilidade $F(X)$ e $G(Y)$, o coeficiente de dependência na cauda superior λ_u é o limite da probabilidade condicional de que Y seja maior do que o percentil limítrofe de G , dado que X é maior do que o percentil limítrofe de F quando (X, Y) se aproximam de 1 pela esquerda:

$$\lambda_u = \lim_{t \rightarrow 1^-} P[Y > G^{-1}(t) | X > F^{-1}(t)]. \quad (25)$$

Se $\lambda_u \in (0, 1]$, (X, Y) possuem dependência na cauda superior. Caso contrário, quando $\lambda_u = 0$, (X, Y) são assintoticamente independentes na cauda superior. De forma análoga, o coeficiente de dependência na cauda inferior λ_l é o limite da probabilidade condicional de que Y seja menor ou igual ao percentil limítrofe de G , dado que X é maior ou igual ao percentil limítrofe de F quando (X, Y) se aproximam de 0 pela direita:

$$\lambda_l = \lim_{t \rightarrow 0^+} P[Y \leq G^{-1}(t) | X \leq F^{-1}(t)] \quad (26)$$

Em ambos os casos, a dependência da probabilidade condicional pode ser trocada, ou seja, os coeficientes podem ser obtidos pela probabilidade condicional de X ser maior, menor ou igual ao percentil limítrofe de F , dado que Y é maior, menor ou igual ao percentil limítrofe de G . Se (X, Y) forem distribuições contínuas, então é possível escrevermos λ_u e λ_l em termos de cópulas:

$$\lambda_u = 2 - \lim_{t \rightarrow 1^-} \frac{1 - C(t, t)}{1 - t} \quad (27.1)$$

$$\lambda_l = \lim_{t \rightarrow 0^+} \frac{C(t, t)}{t} \quad (27.2)$$

5.3. Classes e Famílias de Cópulas

As definições e propriedades das cópulas apresentadas em possibilitam a construção de diversas cópulas de acordo com a forma funcional da *função de ligação*. Nesta seção, veremos as principais características das famílias de cópulas mais comumente utilizadas na literatura: Elípticas, Arquimediana e de Valores Extremos.

5.3.1. Cópulas Elípticas

As cópulas elípticas derivam das distribuições de probabilidade homônimas, caracterizadas pela simetria no formato da distribuição. As definições e propriedades das distribuições elípticas são detalhadas em Fang et al. (1990) e Gupta et al. (2013). As principais famílias de cópulas elípticas são a Gaussiana e t -Student, geradas a partir das distribuições que as nomeiam. A Tabela 4 apresenta as funções de distribuição acumulada (*cdf*) dessas cópulas para o caso bivariado (X, Y) .

Tabela 4. Cópulas Elípticas Bivariadas - Funções de distribuição acumulada

Cópula	Função de distribuição acumulada
Normal (Gaussiana)	$C_{\theta}^{Ga}(u, v) = \Phi_{\theta}^2(\Phi^{-1}(u), \Phi^{-1}(v))$
t -Student	$C_{n, \theta}^t(u, v) = t_{n, \theta}^2(t_n^{-1}(u), t_n^{-1}(v))$

Fonte: elaborado pelo autor.

Os termos da Tabela 4 são assim representados: $u = F_X(x)$ e $v = F_Y(y)$, Φ^2 a distribuição normal padronizada, Φ^{-1} a inversa da distribuição normal padronizada, θ a medida de dependência, t^2 a distribuição t -Student padronizada, t_n^{-1} a inversa da distribuição t -Student padronizada e n o número de graus de liberdade da distribuição e da inversa da distribuição t -

Student. As cópulas Gaussianas não possuem dependência de cauda superior e inferior, exceto quando as distribuições marginais apresentem correlação perfeita mensurada pelo coeficiente de Pearson. Assim:

$$\lambda_u = \lambda_l = \begin{cases} 0 & \text{se } \rho < 1 \\ 1 & \text{se } \rho = 1 \end{cases} \quad (28)$$

A Tabela 5 apresenta a relação entre coeficientes de Kendall e Spearman e os parâmetros das cópulas elípticas bivariadas, assim como as suas respectivas e as medidas de dependência de cauda.

Tabela 5. Cópulas Elípticas Bivariadas: medidas de dependência de Kendall, Spearman e índices de cauda

Cópula	τ	ρ
Gaussiana	$\tau(\theta) = \frac{2}{\pi} \cdot \arcsen(\theta)$	$\rho(\theta) = \frac{6}{\pi} \cdot \arcsen\left(\frac{\theta}{2}\right)$
<i>t</i> -Student	$\tau(\theta) = \frac{2}{\pi} \cdot \arcsen(\theta)$	$\rho(\theta) = \frac{6}{\pi} \cdot \arcsen\left(\frac{\theta}{2}\right)$
Cópula	λ_l	λ_s
Gaussiana	0	0
<i>t</i> -Student	$2t_n \left(-\sqrt{\frac{(n+1)(1-\rho)}{1+\rho}} \right)$	$2t_n \left(-\sqrt{\frac{(n+1)(1-\rho)}{1+\rho}} \right)$

Fonte: elaborado pelo autor.

5.3.2. Cópula de Plackett

A cópula de Plackett é o exemplo mais conhecido de cópulas construídas pelo método algébrico, o qual utiliza a razão de chances (*odds ratio*) ao invés do coeficiente de correlação, como medida de associação da cópula com as distribuições marginais (Wu, 2015). Esta classe de cópula apresenta distribuição simétrica sem dependência de cauda. A Tabela 6 apresenta os seguintes componentes da cópula de Plackett: a função de distribuição acumulada e as medidas de dependência de Kendall, Spearman e de cauda.

Tabela 6. Cópula de Plackett Bivariada – Função de distribuição acumulada, medidas de dependência e medidas de cauda

Componentes da cópula de Plackett	Expressão
Função de Distribuição Acumulada	$C_{\theta}^{Pl}(u, v) = \frac{n^{-1}}{2} \cdot \left(1 + n \cdot (u + v) - \sqrt{1 + n \cdot (u + v)^2 - 4\theta n u v} \right)$
τ	*
ρ	$\frac{\theta + 1}{\theta - 1} - \frac{2\theta \log(\theta)}{(\theta - 1)^2}$
λ_l	0
λ_s	0

*Não há expressão para o Kendall Tau desta família

Fonte: elaborado pelo autor.

Os termos da Tabela 6 são assim representados: $u = F_X(x)$ e $v = F_Y(y)$, θ a medida de dependência e $n = \theta - 1$.

5.3.3. Cópulas Arquimedianas

As cópulas Arquimedianas para o caso bivariado (X, Y) podem ser definidas como:

$$C_{\theta}^A(u, v) = \phi^{-1}(\phi(u) + \phi(x)), \quad (29)$$

em que θ é o parâmetro da cópula e ϕ a função geradora da respectiva família de cópula, que deve satisfazer as seguintes condições: $\phi(1) = 0$ e $\phi(t)$ é decrescente e convexa para todo $t \in (0,1)$. Esta classe de cópula apresenta as seguintes propriedades (Anjos et al., 2004):

1. $C_\theta^A(u, v)$ é simétrica (permutável), ou seja, $C_\theta^A(u, v) = C_\theta^A(v, u)$;
2. $C_\theta^A(u, v)$ é associativa, i.e., $C(C_\theta^A(u, v), w) = C(u, C_\theta^A(v, w))$, para todo u, v, w em $(0,1)$.
3. Para cada função geradora ϕ de uma família de cópula, existe alguma constante $k > 0$, em que $k\phi$ é mantida como função geradora.

As características das cópulas arquimedianas possibilitam que elas sejam construídas como medidas representativas de distribuições de probabilidade condicional conjunta (Cossette et al., 2018), com ampla aplicação nos campos de finanças e ciências atuariais (Araujo & Carvalho, 2021; Cherubini et al., 2004). A representação da função distribuição acumulada para o caso bivariado (X, Y) de quatro famílias arquimedianas é apresentada na Tabela 7.

Tabela 7. Cópulas Arquimedianas Bivariadas: Funções de distribuição acumulada

Cópula	Função de distribuição acumulada
Frank	$C_\theta(u, v) = -\frac{1}{\theta} \cdot \ln \left(1 + \frac{(e^{-u\theta} - 1)(e^{-v\theta} - 1)}{e^{-\theta} - 1} \right)$
Clayton	$C_\theta(u, v) = \max \left([u^{-\theta} + v^{-\theta} - 1]^{\frac{1}{\theta}}, 0 \right)$
Gumbel	$C_\theta(u, v) = \exp \left\{ -\left[(-\ln(u))^\theta + (-\ln(v))^\theta \right]^{\frac{1}{\theta}} \right\}$
Joe	$C_\theta(u, v) = 1 - \left[(1-u)^\theta + (1-v)^\theta - (1-u)^\theta \cdot (1-v)^\theta \right]^{\frac{1}{\theta}}$

Fonte: elaborado pelo autor.

A Tabela 8 apresenta a relação entre coeficientes de Kendall e Spearman e os parâmetros das cópulas arquimedianas e suas respectivas e as medidas de dependência de cauda.

Tabela 8. Cópulas Arquimedianas Bivariadas: Medidas de dependência de Kendall, Spearman e de Cauda

Cópula	τ	ρ
Frank	$\tau(\theta) = 1 - \frac{4}{\theta} + \frac{4}{\theta^2} \int_0^\theta \frac{t}{e^t - 1} dt$	$\rho(\theta) = 1 - \frac{12}{\theta^3} \int_0^\theta \frac{\theta t - 2t^2}{e^t - 1} dt$
Clayton	$\tau(\theta) = \frac{\theta}{\theta + 2}$	*
Gumbel	$\tau(\theta) = \frac{1}{1 - \theta}$	*
Joe	$\tau(\theta) = 1 + \frac{4}{\theta} \int_0^1 \frac{\ln(1-t^\theta) \cdot (1-t^\theta)}{t^{\theta-1}} dt$	*
Cópula	λ_I	λ_S
Frank	0	0
Clayton	0	$2 - \frac{1}{2\theta}$
Gumbel	0	$2 - \frac{1}{2\theta}$
Joe	0	$2 - \frac{1}{2\theta}$

Nota: * embora a relação do coeficiente com o parâmetro da cópula não possa ser apresentada analiticamente, ele pode ser estimado por métodos paramétricos e não paramétricos. Fonte: elaborado pelo autor.

5.3.4. Cópulas de Valores Extremos

Em áreas como finanças, seguros e *environmental economics*, eventos extremos podem causar severos impactos e precisam ser modelados cuidadosamente (Mendes-Da-Silva et al., 2021; Oliveira & Carvalho, 2020). As cópulas de valores extremos, fundamentadas na teoria dos valores extremos, são uma alternativa para modelar a estrutura de dependência destes eventos (Gudendorf & Segers, 2010).

Conforme apresentado por (Araujo & Carvalho, 2021), uma cópula C_* pertence à família das cópulas para valores extremos se, e somente se, existe uma cópula C tal que

$$C_*(u, v) = \lim_{n \rightarrow \infty} C^n \left(u^{\frac{1}{n}}, v^{\frac{1}{n}} \right). \quad (30)$$

Se uma cópula C pertence à família de valores extremos, então C pode ser escrita como

$$C(u, v) = \exp \left\{ \ln(u \cdot v) \cdot A \left(\frac{\ln(v)}{\ln(u \cdot v)} \right) \right\}, \quad (31)$$

para uma escolha apropriada da função $A(\cdot)$, chamada de *função geradora* da cópula para eventos extremos C . A Tabela 9 apresenta a *função geradora* $A(t)$ e a função de distribuição acumulada das cópulas de valores extremos Galambos, Husler-Reiss e t -EV.

Tabela 9. Cópulas de valores extremos - Função geradora e função de distribuição acumulada

Cópula	Função A(t)
Galambos	$1 - (t^{-\theta} + (1-t)^{-\theta})^{-\frac{1}{\theta}}$
Husler-Reiss	$t \cdot \Phi \left(\theta^{-1} + \frac{\theta}{2} \cdot \ln \left(\frac{t}{1-t} \right) \right) + (1-t) \cdot \Phi \left(\theta^{-1} - \frac{\theta}{2} \cdot \ln \left(\frac{t}{1-t} \right) \right)$
t -EV	$t_{n+1}^2 \left\{ \frac{\left[\left(\frac{t}{(1-t)} \right)^{\frac{1}{n}} - \rho \right]}{\sqrt{\frac{(1-\rho^2)}{(n+1)}}} \right\} + (1-t) \cdot t_{n+1}^2 \left\{ \frac{\left[\left(\frac{(1-t)}{t} \right)^{\frac{1}{n}} - \rho \right]}{\sqrt{\frac{(1-\rho^2)}{(n+1)}}} \right\}$
Cópula	Função de distribuição acumulada
Galambos	$C_\theta(u, v) = u \cdot v \cdot \exp \left\{ \left[(-\log(u))^{-\theta} + (-\log(v))^{-\theta} \right]^{-\frac{1}{\theta}} \right\}$
Husler-Reiss	$C_{HR,\theta}(u, v) = \exp \left\{ -\ln(u) \cdot \Phi \left(\theta^{-1} + \frac{\theta}{2} \cdot \ln \left(\frac{\ln(u)}{\ln(v)} \right) \right) + \ln(v) \cdot \Phi \left(\theta^{-1} + \frac{\theta}{2} \cdot \ln \left(\frac{\ln(v)}{\ln(u)} \right) \right) \right\}$
t -EV	$C_\theta(u, v) = \exp \left\{ t_{n+1}^2 \left[-\frac{\rho}{\theta} + \frac{1}{\theta} \left(\frac{\log(u)}{\log(v)} \right)^{\frac{1}{n}} \right] \log(u) + t_{n+1}^2 \left[-\frac{\rho}{\theta} + \frac{1}{\theta} \left(\frac{\log(u)}{\log(v)} \right)^{\frac{1}{n}} \right] \log(v) \right\}$

Para t -EV, t^2 é distribuição t -Student padronizada, n é o número de graus de liberdade e ρ o coeficiente de correlação.

Fonte: elaborado pelo autor.

Os coeficientes de Kendall e Spearman podem ser obtidos com base nas *funções geradoras* $A(t)$ aplicadas às seguintes equações:

$$\tau = \int_0^1 \frac{t \cdot (1-t)}{A(t)} dA'(t) \quad e \quad \rho = -3 + 12 \cdot \int_0^1 \frac{1}{(1+A(t))^2} dt \quad (32)$$

Por fim, a Tabela 10 apresenta as medidas de dependência de cauda das cópulas de valores extremos.

Tabela 10. Cópulas de valores extremos – índices de cauda

Cópula	λ_l	λ_s
Galambos	$\begin{cases} 0 \text{ se } A(1/2) > 1/2 \\ 1 \text{ se } A(1/2) = 1/2 \end{cases}$	$2[1 - A(1/2)]$
Husler-Reiss	$\begin{cases} 0 \text{ se } A(1/2) > 1/2 \\ 1 \text{ se } A(1/2) = 1/2 \end{cases}$	$2[1 - A(1/2)]$
t -EV	$2 - \lambda_s$	$2t_{n+1}^2 \left(\sqrt{\frac{(n+1)(1-\rho)}{(1+\rho)}} \right)$

Fonte: elaborado pelo autor.

5.4. Ajuste das cópulas aos dados

Conforme argumentado, as cópulas têm por objetivo precípuo a capacidade de associar distribuições de probabilidade marginais a uma distribuição de probabilidade conjunta que terá a sua forma influenciada pela sua função geradora e seu parâmetro. Para a aplicação de uma função de cópula a um conjunto de dados empíricos, é necessário estimar o parâmetro que dará formato à função de distribuição de probabilidade conjunta. Deste modo, é possível extrair uma medida desta associação das distribuições marginais para cada família de cópula.

A escolha de qual família de cópula melhor se ajusta a um conjunto de dados reais pode ser feita por meio de testes que verificam qual família e seu respectivo parâmetro estimado são mais aderentes à distribuição dos dados. Esta seção tem por objetivo apresentar os principais métodos de estimação de parâmetro e testes de aderência e qualidade de ajuste.

5.4.1. Métodos de estimação de parâmetros

Método dos Momentos (MME)

Como há uma relação entre os parâmetros das famílias de cópulas supramencionadas e suas respectivas medidas de concordância (τ de Kendall e ρ de Spearman), é possível estimar o parâmetro destas cópulas partindo destas medidas. Assim, assumindo uma cópula bivariada de parâmetro único dada por $H(x, y) = C_\theta(F(x), G(y))$, nós temos a relação do entre coeficiente τ e a distribuição de probabilidade acumulada conjunta: $\tau(X, Y) = f(\theta)$.

Para ajustar a cópula ao parâmetro estimado, o *Método dos Momentos* (MME) calcula o valor amostral de do coeficiente τ , (r^τ) e então resolve a equação $r^\tau = f(\hat{\theta})$, assumindo que $\hat{\theta}$ esteja contido no intervalo válido do parâmetro da cópula. Como exemplo, conforme a Tabela 8, para a cópula de Clayton, $\tau(\hat{\theta}) = \frac{\theta}{\theta+2}$. Para maiores detalhes, ver McNeil et al. (2015).

Método da Máxima Verossimilhança (MLE)

O Método da Máxima Verossimilhança (*Maximum Likelihood Estimation* – MLE) é outra maneira de estimar o parâmetro verdadeiro. Seguindo Weiß (2011), e sem perda de generalidade usando o caso bivariado como exemplo, o parâmetro $\hat{\theta}^{MLE}$ é obtido por meio da maximização da função de verossimilhança de um conjunto de dados:

$$L_C(\theta) = \sum_{i=1}^n \log c(U_{i,1}, U_{i,2} | \theta), \quad (33)$$

em que c é a densidade da cópula cujas entradas são $U_i \equiv (U_{i,1}, U_{i,2})^t \in [0,1]^2$, com i a t -ésima observação da amostra. A função densidade da cópula é obtida por meio de:

$$c(u_1, u_2 | \theta) = \frac{\partial^2 c(u_1, u_2 | \theta)}{\partial u_1 \partial u_2}, \quad u_1, u_2 \in [0; 1] \quad (34)$$

Por fim, o estimador de Máxima Verossimilhança (MLE) é:

$$\hat{\theta}^{MLE}(U) \equiv \arg \max L_C(\theta). \quad (35)$$

O *Canonical Maximum Likelihood* (CML) e o *Maximum Pseudo-Likelihood* (MPLE) são formas alternativas de se obter o MLE. O CML possui dois estágios: no primeiro, as distribuições marginais (no presente estudo, as taxas de *default*) são convertidas em variáveis aleatórias uniformes, dispensando a estimação dos parâmetros das distribuições marginais. Então, no segundo estágio, o parâmetro da cópula é estimado por uma função de máxima verossimilhança que inclui as margens uniformes e o parâmetro da cópula. Já no método MPLE, os parâmetros das cópulas são estimados por uma função que maximiza conjunta e simultaneamente a verossimilhança das variáveis uniformes e dos parâmetros das cópulas (Crook & Moreira, 2011).

Em estudos empíricos, o método de estimação dos parâmetros de cópulas é uma das escolhas metodológicas que o pesquisador precisa fazer. Essa escolha discricionária pode afetar substancialmente os resultados obtidos. Porém, estudos comparativos dos métodos de estimação auxiliam esta decisão. Durrleman et al. (2000) estimaram a distribuição conjunta do retorno das *commodities* alumínio e cobre negociados na *London Metal Exchange*, observando uma melhor precisão do parâmetro obtido pela CML em comparação aos parâmetros obtidos por MME e MPLE. Frey & McNeil (2003) conduziram um experimento com dados de risco de crédito da *S&P*. Ao comparar os parâmetros estimados por MME e MLE, os autores verificaram a melhor performance do método MLE, cada vez maior à medida que a qualidade do crédito é elevada. Cherubini et al. (2004), Durrleman et al. (2000) e Kole et al. (2007) também consideram que o método MLE apresenta maior precisão, apesar de ponderarem que este método possui um elevado custo computacional. Todos consideram o MME uma forma mais rápida e viável para se estimar os parâmetros de cópulas de um grande volume de dados.

5.4.2. Testes *Goodness-of-Fit* para escolha de cópulas

Os testes *Goodness-of-Fit* (*GoF*) aplicados às cópulas avaliam a aderência da cópula testadas a um conjunto de dados. Esses testes são realizados sob a seguinte hipótese nula:

$$H_0: C \in C_\theta, \quad (36)$$

em que C representa a distribuição de probabilidade acumulada conjunta dos dados e C_θ a família de cópula estimada sob o parâmetro θ , que, por sua vez, a depender do teste (*GoF*), pode ser estimado via MME ou MLE. Assim, dado um conjunto de cópulas que se deseja verificar a melhor aderência à distribuição empírica, elege-se aquela que apresentar maior *p-valor* no teste, indicando a não rejeição da hipótese nula de aderência à cópula testada. Embora a escolha da melhor cópula pela não rejeição da hipótese nula possa levar a um erro de interpretação, uma vez que a não rejeição da hipótese não implica na sua aceitação, este método de seleção é amplamente utilizado (Okhrin et al., 2021).

Existe uma ampla variedade de testes *GoF*, classificados por Okhrin et al. (2021) em sete grupos: (i) *Empirical Copula Process*; (ii) *Kendall's process*; (iii) Transformação de *Rosenblatt*; (iv) Transformação de Cópulas Arquimedianas; (v) Estimativa de densidade de Kernel; (vi) Teste White, e; (vii) Teste de Validação Cruzada. Os trabalhos de Breymann et al., (2003), Genest et al. (2006), (2009), C. Hering & Hofert (2015) Huang & Prokhorov (2014), Scaillet (2007) e Zhang et al. (2016) abordam um ou mais testes *GoF* pertencentes aos sete grupos.

Os testes pertencentes a cada grupo possuem vantagens e desvantagens em termos de poder explicativo, podendo culminar em resultados opostos para o mesmo conjunto de dados e cópula estimada C_θ (Okhrin et al., 2021). Face a esta característica, Zhang et al. (2016) propuseram o teste *Goodness-of-Fit* híbrido (GoF_{hb}), visando consolidar em uma única medida o poder explicativo de um conjunto de os testes *GoF*. O teste GoF_{hb} é estruturado da seguinte

forma: seja q o número de testes GoF que irão compor o teste GoF_{hb} , com os respectivos p -valores p^1, \dots, p^q , o p -valor combinado do teste é

$$p^{hb} = \min\{q \times \min(p^1, \dots, p^q), 1\} \quad (37)$$

5.6. Aplicação de cópulas ao risco de crédito

Os dados financeiros de ativos de crédito geralmente apresentam não normalidade em relação à distribuição de seus retornos, em particular, pela sua característica de caudas pesadas (Clemente & Romano, 2004). Devido à capacidade das cópulas em modelar estruturas de dependência assimétricas, a sua aplicação a modelos de risco de crédito vem sendo tratada em trabalhos teóricos e empíricos, usando tanto dados simulados como reais.

Dentre os trabalhos teóricos, Schönbucher & Schubert (2001) propuseram um modelo de intensidade de *default* que leva em conta o risco de títulos de dívidas corporativas de forma individual e conjunta, sendo o segundo estimado por cópulas. Seguindo a mesma linha, Frey & Mcneil (2003) analisam os modelos de risco de crédito de portfólio utilizados pela indústria financeira, dando ênfase à abordagem da estrutura de dependência entre os *defaults*. Os modelos tradicionais são baseados em distribuições normais multivariadas, e os autores então mostram que é possível mensurar a estrutura de dependência por meio de cópulas.

Por sua vez, Glasserman et al. (2007) evidenciaram como a cópula Normal estima eventos de cauda em modelos de risco de crédito de portfólio em cenários de baixo e elevado índice de *default*. Os autores demonstram, por meio de teoremas, que uma das propriedades da cópula Normal é não apresentar dependência de caudas extremas, implicando que as variáveis aleatórias agregadas pela cópula Normal são independentes nos extremos.

Trabalhos que aplicam cópulas a dados de risco de crédito são mais recorrentes. Partindo do entendimento que os bancos possuem informações não públicas sobre devedores corporativos, Hamilton et al. (2001) propuseram um modelo de *credit rating* que integra a análise de risco interna das instituições financeiras com as classificações de risco externas (agências de *rating*). Com essas informações foi possível estimar o risco individual de um título de crédito e o risco de um portfólio, neste caso com auxílio de funções cópulas. Os autores utilizaram a família de cópula Ali-Mikhail-Haq (da classe de cópulas Arquimedianas) para ajustar a distribuição estimada aos dados de dívida corporativa da *Moody's Corporate Bond Default database*.

Os trabalhos de Clemente & Romano (2004), Crook & Moreira (2011), Fenech et al. (2015) e Koziol et al. (2015) efetuaram comparação da aderência de estimativas de *default* entre modelos que assumem a normalidade da distribuição conjunta dos dados e modelos baseados em cópulas. Clemente & Romano (2004) avaliaram a sensibilidade de modelos de mensuração de risco de crédito e de otimização de portfólio à adoção de diferentes famílias de cópulas.

As famílias de cópulas testadas foram a Gaussiana (Normal), t -Student, *grouped t* e Clayton. A construção das cópulas família *grouped t* se assemelha à da t -Student, tendo como diferença principal a segregação da variável aleatória em grupos com níveis de dependência de cauda próprios, ao invés de um único parâmetro global proposto pela t -Student (Daul et al., 2003). Em relação aos modelos de mensuração de risco de crédito, foi possível observar que as cópulas com dependência de cauda geraram medidas de risco de cauda (i.e., *maximum loss*; CreditVaR, e; *expected shortfall*) mais elevadas no 99º percentil de suas distribuições em comparação com a cópula Gaussiana.

Entretanto, ao verificar o efeito das famílias de cópulas na otimização do portfólio de crédito, sob o escopo da teoria moderna do portfólio, observou-se que a fronteira de alocação eficiente na relação risco e retorno foi muito semelhante para as cópulas Gaussiana, t -Student

e *grouped t*. Apenas houve divergência quando da estimação via cópula de Clayton (a qual apresentou uma fronteira eficiente com maior coeficiente de inclinação). Tal resultado reforça o entendimento de que, apesar do auxílio à melhor estimação de modelos de risco de crédito que as cópulas podem proporcionar, a escolha de qual cópula utilizar deve ser bem fundamentada. Afinal, o principal risco envolvido é de provocar maiores distorções em relação à realidade do que os modelos tradicionalmente utilizados.

Crook & Moreira (2011) estimaram a ocorrência conjunta de *default* extremos em um portfólio de cartão de crédito pertencente a um banco do Reino Unido. A estrutura de dependência da PD das operações foi estimada de acordo com o *rating* atribuído pelo banco. Partindo das cinco classificações de *rating* (de A a E), dez distribuições conjuntas bivariadas foram estimadas $[A \times B, \dots, D \times E]$ sob a premissa da normalidade e via diferentes famílias de cópulas. No modelo de cópulas, foram avaliadas as aderências de dez famílias para cada par de *rating* de operações foi testada, por meio de três testes *Goodness-of-Fit* (*Empirical Copula Process*; *Kendall's process* e Transformação de *Rosenblatt*). A precisão dos modelos de cópulas e da distribuição Normal foi verificada usando proximidade da PD acumulada estimada da PD acumulada observada nos percentis extremos. Os resultados apresentaram maior precisão da estimativa por cópulas em 63,33% dos cenários.

Fenech et al. (2015) avaliaram a distribuição conjunta do risco de crédito de quatro linhas de varejo (aquisição de auto; 1ª hipoteca; 2ª hipoteca, e; cartão de crédito), no período anterior e posterior à crise do *subprime*, tendo como fonte de dados o índice *S&P Experian Consumer Credit Default*. Os autores construíram distribuições bivariadas a partir dos seis subconjuntos formados pelos pares das quatro linhas de crédito. Testes de aderência de cópulas aos dados reais foram efetuados para as cópulas de Gumbel, Clayton e Frank. Na maior parte dos testes, os resultados apontaram para a melhor performance da cópula de Clayton.

Na contramão dos estudos de riscos de crédito conjuntos, Koziol et al. (2015) verificaram boa aderência da distribuição Normal às distribuições empíricas, de modo que este tipo de abordagem não deve ser descartado a priori em detrimento de uma modelagem via cópulas. Os autores utilizaram dados de 17 bancos alemães para aplicar um modelo multifatorial de risco de crédito baseado no *CreditMetrics* mensurando a estrutura de dependência entre as operações via cópulas elípticas e arquimedianas (Normal, *t*-Student e Clayton). A cópula Normal apresentou melhor aderência em dois cenários: (i) quando o *cut-off* de valores extremos é baixo, causado por uma alta dependência entre os valores inferiores da distribuição, e; (ii) quando as distribuições marginais de risco de crédito são altamente correlacionadas.

Sob a perspectiva do capital regulatório, diversos trabalhos verificaram ganhos de precisão na estimação do capital requerido sob uma abordagem de cópulas (Andrievskaya & Penikas, 2012; Clemente, 2020; Embrechts et al., 2014; Lu et al., 2017; Moreira, 2010, 2015; Penikas, 2020; Tente et al., 2019). Moreira (2010) demonstra que o modelo proposto por Basileia II, ao assumir que a variável latente do risco de *default* é normalmente distribuída, é incapaz de identificar dependência de cauda e, por consequência, a ocorrência de eventos extremos conjuntos. Para tratar deste *gap*, o autor propôs a utilização de cópulas na estimação da perda inesperada. Embora o modelo proposto possibilite a aplicação de qualquer família de cópula, o autor optou pela utilização da cópula de Clayton devido à sua capacidade de representar relações de dependência na cauda inferior da distribuição conjunta, uma vez que a sua variável de interesse era a probabilidade de sobrevivência (não *default*) das operações.

O estudo é feito por meio de simulações do requisito de capital de um portfólio de crédito sob duas abordagens: (i) o modelo de Basileia, e; (ii) o modelo alternativo via cópula. Ao todo, 660 cenários foram elaborados a partir de variações na probabilidade de *default*, no formato das distribuições marginais de probabilidade de *default* e nos parâmetros da cópula.

Para comparar a performance entre os modelos, o autor calculou a distância entre a perda efetiva (que também fora simulada) e o requisito de capital de cada abordagem, sem distinguir se o erro da estimativa foi conservador (requisito de capital maior do que a perda efetiva) ou com risco de insolvência (requisito de capital menor do que a perda efetiva). Os resultados apresentaram, em média, melhor performance da abordagem por cópulas para a estimação do requisito de capital.

Moreira (2015) abordou novamente o modelo de capital regulatório de Basileia, estimando as perdas de operações de crédito pela abordagem tradicional e por cópulas, mas desta vez baseado em dados agregados de bancos norte-americanos. A falta de granularidade dos dados impossibilitou o cálculo das medidas de dispersão entre as operações da linha de crédito. Este problema foi contornado pela utilização de coeficientes de correlação estipulados pelo Comitê de Basileia. Três linhas de crédito foram testadas sob a abordagem tradicional e via cópula *t*-Student e Clayton. As linhas de cartão de crédito e corporativa foram melhor representadas pelas cópulas, e a linha de crédito hipotecário foi melhor representada pelo modelo tradicional de Basileia. Embora a ausência da estimação dos coeficientes de correlação pudesse ter sido prejudicial à abordagem de cópulas, os resultados são relevantes. Afinal, há evidências de que, mesmo que a abordagem por cópulas tenha um bom poder explicativo sob uma perspectiva de cálculo de capital regulatório, não há supremacia absoluta desta abordagem em relação ao modelo tradicional de Basileia.

Andrievskaya & Penikas (2012) investigaram se há uma característica pró-cíclica no modelo de requisito de capital implementado pelo Acordo de Basileia II, de modo que o índice de capital das instituições financeiras seja elevado em momentos de expansão econômica (devido à uma subestimação do risco total). Além disso, quando há a entrada em um período de recessão, a percepção de risco é reajustada, havendo uma redução na adequação de capital. Porém, tal redução, gera um efeito endógeno na percepção de risco, atuando de forma cíclica na sua elevação. Para essa finalidade, os autores utilizaram dados agregados de risco de crédito, de mercado e operacional dos bancos russos, e ajustaram cópulas tridimensionais das classes elípticas e arquimedianas. Os autores não identificaram uma característica pró-cíclica no modelo proposto por Basileia II.

Devido à atuação das funções de cópulas na construção distribuições de probabilidade multivariadas, sua aplicação ao risco de crédito ocorre sobretudo em uma perspectiva de portfólio. Todavia, as suas funcionalidades também podem ser aplicadas, e com resultados positivos, em modelos de risco de crédito a nível individual. Calabrese et al. (2019) apresentam um modelo de risco de crédito individual que estima a probabilidade de *default* em plataformas de empréstimo *peer-to-peer* condicionado à existência ou não de *default* registrado em *credit bureaus*, estimando a dependência entre as probabilidades de *default* via cópulas.

Penikas (2020) propõe um modelo de probabilidade de *default* individual de análise discriminante baseado em cópulas (*Copula Discriminant Analysis – CODA*), o qual consiste nas seguintes etapas: (i) estimar a distribuição marginal de todas as variáveis explicativas do modelo e suas respectivas funções de distribuição acumulada; (ii) utilizar as funções de distribuição acumulada marginais para estimar o parâmetro das famílias de cópulas a serem testadas (Clayton, Gumbel, Frank e Joe); (iii) escolha da cópula com maior verossimilhança; (iv) separação da amostra entre treino e teste; (v) cálculo da probabilidade de *default* e de não *default* na amostra de treino, e; (vi) cada observação é classificada como *default* ou não *default* baseada no valor da função de máximo verossimilhança.

A Tabela 11 apresenta um resumo da bibliografia de aplicação de cópulas ao risco de crédito levantada por esta pesquisa.

Tabela 11. Resumo da bibliografia de cópulas aplicadas ao risco de crédito.

Ordem	Autores	Risco de Crédito	Dados	Cópulas
1	Hamilton et al. (2001)	Corporativo	Empírico	AMH
2	Schönbucher & Schubert (2001)	Corporativo		N, Cl, Gb
3	Frey & Mcneil (2003)	Corporativo		Cl
4	Clemente & Romano (2004)	Corporativo	Sintético	N, t, G-t, Cl
5	Glasserman et al. (2007)			N
6	Moreira (2010)	Varejo	Sintético	Cl
7	Crook & Moreira (2011)			N, t, Pk, Cl, Gb, Joe, F,
		Varejo	Empírico	FGM, Ga, HR
8	Andrievskaya & Penikas (2012)	Portfólio Bancário	Empírico	N, Cl, Gb, F
9	Fenech et al. (2015)	Varejo	Empírico	N, Cl, Gb, F
10	Berg & Koziol (2017)	Portfólio Bancário	Empírico	N, t, Cl
11	Moreira (2015)	Varejo e Corporativa	Empírico	t, Cl
12	Calabrese et al. (2019)	Varejo	Empírico	N, Gb
13	Penikas (2020)	-	Teórico	Cl, Gb, Joe, F

Esta tabela apresenta o resumo da bibliografia apurada de cópulas aplicadas a modelos de risco de crédito. Os treze trabalhos são apresentados em função do tipo de exposição ao risco de crédito (corporativo, varejo, portfólio bancário ou nula), o tipo de dado (empírico, sintético ou trabalho teórico) e as famílias de cópulas utilizadas (N: Normal; t: t-Student; G-t: Grouped-t; Pk: Placket; Cl: Clayton; Gb: Gumbel; Joe: Joe; F: Frank; FGM: Farlie–Gumbel–Morgenstern; AMH: Ali-Mikhail-Haq; Ga: Galambos; HR: Hüsler–Reiss).

Fonte: elaborado pelo autor.

6. Os dados e análise descritiva

A análise empírica das probabilidades de *default* de operações de crédito rural estimadas por modelos tradicionais (assumindo a normalidade da distribuição) e por modelos baseados em cópulas foi elaborada utilizando dados de operações de crédito do principal banco de financiamento rural do Brasil, pelo critério de *market share*. A autorização de uso e a disponibilização destes dados para esta pesquisa foi condicionada a um processo de avaliação da instituição financeira que considerou: (i) a viabilidade do projeto de pesquisa; (ii) o interesse da instituição em relação aos resultados obtidos; (iii) o interesse da sociedade em relação aos resultados obtidos, e; (iv) a proteção de informações sensíveis e sigilosas. Uma vez aprovada a utilização pelas áreas internas da instituição, a disponibilização dos dados ocorreu seguindo as diretrizes da Lei Geral de Proteção de Dados (LGPD) sancionada em 2018, sob a égide da jurisdição federal brasileira. As variáveis qualitativas de identificação dos devedores e das operações foram transformadas por uma chave única sigilosa. Assim, o anonimato das operações é preservado sem inviabilizar a identificação de uma mesma operação ao longo do tempo e de operações diferentes para o mesmo devedor.

Os dados correspondem ao acompanhamento mensal de uma amostra de empréstimos rurais, no período entre julho de 2014 a setembro de 2019³. Os critérios de extração da informação original foram definidos internamente pelo banco. As observações mensais não são balanceadas, uma vez que há a saída de operações em caso de vencimento ou contabilizadas como perda (a quantidade de dias em atraso para contabilização como perda varia de acordo com o tipo da operação) e a entrada de operações, por conta de novas contratações. Ao todo, a amostra possui 74.964.734, contendo as variáveis descritas na Tabela 12.

Tabela 12. Informações disponíveis das carteiras de Crédito Rural – Descrição das variáveis.

Variável	Descrição
(1) Data Ob.	Data da observação
(2) Id Devedor	Identificação do devedor
(3) Id. Operação	Identificação da operação
(4) Id. Linha	Identificação da linha de crédito
(5) Data Cont.	Data da contratação da operação
(6) Nat. Jr.	Natureza jurídica do devedor (pessoa física ou pessoa jurídica)
(7) Dias de Atraso	Nº de dias que a operação está em atraso

Esta tabela apresenta as variáveis das 74.964.734 observações individuais de operação de crédito rural disponibilizadas pela instituição financeira, referente ao período entre julho de 2014 a setembro de 2019. A identificação destas variáveis tem como propósito subsidiar uma análise preliminar da carteira de crédito e auxiliar na elaboração de *clusters* que serão utilizados a abordagem por cópulas, não tendo a função de variáveis aplicadas a algum modelo.

Fonte: elaborado pelo autor.

A amostra contém sete linhas de crédito: Pronaf; Pronamp; Custeio; Giro; Investimento; Aquisição Ativos Financeiros, e; Garantia de Preços e Estocagem. A linha de crédito Pronaf destina-se a estimular a geração de renda e melhorar o uso da mão de obra familiar, por meio do financiamento de atividades e serviços rurais agropecuários e não agropecuários desenvolvidos em estabelecimento rural. O Pronamp é destinado Financiamento para custeio e investimentos dos médios produtores rurais em atividades agropecuárias. A linha de Custeio atende ao financiamento de despesas normais da atividade agrícola ou pecuária, mediante orçamento, plano ou projeto, por sua vez, a linha de Giro é um crédito rotativo de capital de

³A primeira observação mensal é de julho de 2014, mas nela, assim como nos meses subsequentes, há o acompanhamento de operações contratadas em períodos anteriores. A operação mais antiga e ainda ativa até a data da observação foi contratada em 06/08/2008.

giro destinado ao produtor rural para aplicação na atividade agropecuária. A linha de Investimento é destinada à aquisição de bens e serviços voltados à produção agrícola, como tratores ou contratação de consultoria. A linha de Aquisição de Ativos Financeiros destinada à antecipação de recursos ao produtor via cédula do produtor rural – CPR. Por fim, a linha de Garantia de Preços e Estocagem financia as despesas com estocagem para além da que se era esperada pelo produtor, devido à problemas na comercialização de seu insumo.

O número de observações conforme a linha de crédito é apresentada na Tabela 13. Para simplificar a identificação destas linhas no decorrer da análise, foram atribuídos rótulos de “A” a “G”.

Tabela 13. Rótulos e quantidade de observações das linhas de crédito rural

Rótulo	Linha de Crédito	Número de Observações
A	Pronaf	55.745.916
B	Pronamp	9.868.078
C	Custeio	6.097.575
D	Giro	403.501
E	Investimento	2.194.904
F	Aquisição Ativos Financeiros	344.501
G	Garantia de Preços e Estocagem	310.259

Fonte: elaborado pelo autor.

As estatísticas descritivas do saldo contábil das operações apresentadas na Tabela 14, servem como *proxies* da dimensão dos valores emprestados em cada linha de crédito.

Tabela 14. Estatísticas descritivas - saldo contábil (em BRL) por linha de crédito

Linha de Crédito	Média	Mediana	Valor Máximo	Desvio Padrão
A	33.949	21.601	34.980.050	76.525
B	115.262	84.891	42.450.638	125.760
C	328.016	133.616	412.201.842	1.797.500
D	99.329	52.435	100.69.567	201.396
E	390.361	150.899	249.147.638	1.649.853
F	439.434	200.319	899.353.873	4.486.938
G	1.344.089	275.091	2.435.857.740	25.765.452

Fonte: elaborado pelo autor.

Conforme esperado, a linha de crédito Pronaf é a que apresenta menor valor de saldo contábil médio e mediano, por ser destinada a pequenos produtores, o que implica em menores financiamentos. O valor máximo observado nesta linha é um *outlier* (456 desvios padrões acima da média), justificado pela existência da subcategoria “Pronaf Agroindústria Investimento”, linha de crédito inserida no programa Pronaf, para financiamento de investimentos, inclusive em infraestrutura. Entre o público elegível, encontram-se as cooperativas constituídas por agricultores familiares, que podem financiar até R\$ 35 milhões, respeitado o limite individual de R\$ 45 mil por cooperado.

A variável *dias de atraso* foi utilizada para classificar as operações como em andamento ou em *default*. O *default* foi definido como o atraso no pagamento superior a 90 dias. A taxa de *default* do mês t da linha de crédito k , foi calculada pela seguinte equação:

$$\text{Taxa de default}_{k,t} = \frac{n.\text{default}_{k,t}}{n_{k,(t-3)}}, \quad (38)$$

sendo $n.\text{default}_{k,t}$, o total de *defaults* na linha de crédito k no mês t e $n_{k,(t-3)}$ o número de operações em andamento da linha de crédito k no mês $t - 3$. Para impedir que as operações sejam consideradas como em *default* por mais de um mês, superestimando a taxa de *default*, as operações com mais de 120 de atraso foram excluídas da amostra. Assim, as operações

classificadas como em *default* em t , são aquelas que atingiram a marca de 90 dias de atraso no mês em questão, originadas das operações em andamento em $t - 3$, e totalizam 256.380 observações.

Como as observações mensais iniciam em julho de 2014, a taxa de *default* começa a ser calculada em outubro de 2014, mês em que há o primeiro histórico de operação em andamento em $t - 3$. Deste modo, temos uma amostra de 60 taxas *defaults* mensais das sete linhas de crédito rural, referentes ao período entre out/2014 a set/2019. A Tabela 15 traz as médias, variâncias e coeficientes de variação da taxa de *default* das linhas de crédito rural avaliadas. Para auxiliar no entendimento das características idiossincráticas destas linhas, também são apresentadas as respectivas informações de todas as operações de crédito registradas no Sistema Financeiro Nacional no mesmo período, com base nos dados disponíveis no Sistema Gerenciador de Séries Temporais -SGS⁴, de curadoria do Banco Central do Brasil.

Tabela 15. Estatísticas descritivas – Taxa de *Default* por linhas de crédito rural e SFN

Linha de Crédito	Média	Desvio Padrão	Coefficiente de Variação
A	0,37%	0,17%	0,4744
B	0,36%	0,15%	0,4144
C	0,22%	0,12%	0,5552
D	0,83%	0,61%	0,7369
E	0,22%	0,12%	0,5732
F	2,12%	5,12%	2,4171
G	0,25%	0,20%	0,7726
SFN	3,30%	0,35%	0,1057

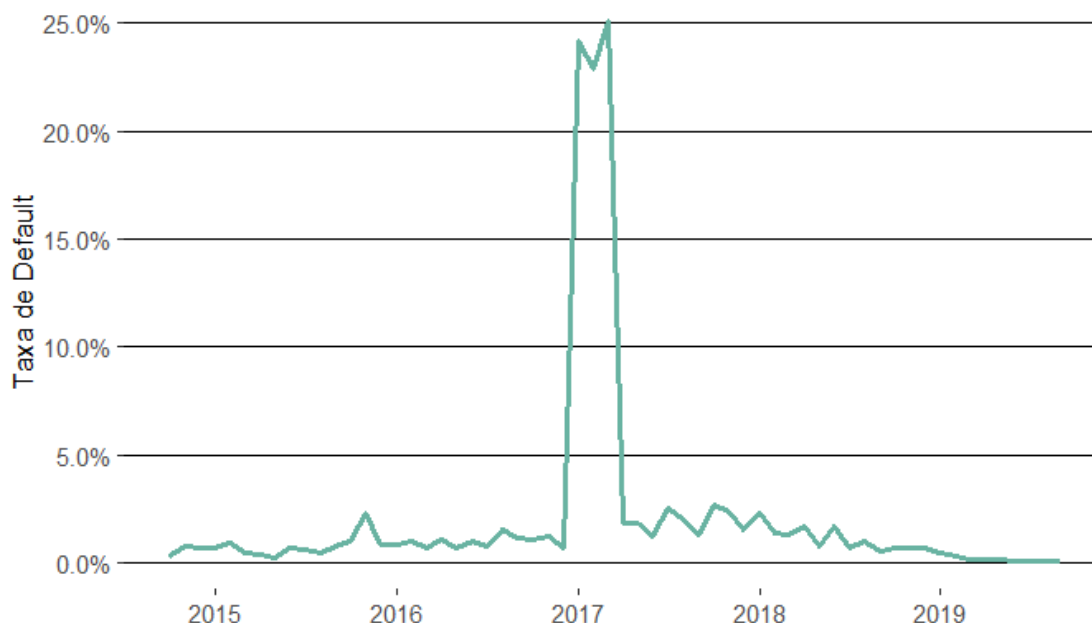
Nota: dados referentes ao período entre out/2014 a set/2019.

Fonte: elaborado pelo autor.

As taxas médias mensais de *default* das linhas de crédito rural são significativamente inferiores à taxa média mensal de *default* de todas as operações de crédito no SFN, demonstrando que a inadimplência deste setor é menor do que a de outras linhas de crédito no Brasil. Entretanto, o coeficiente de variação dessas linhas é no mínimo, 3,9 vezes maior do que o observado em todo o SFN. O coeficiente de variação é uma medida de volatilidade das observações, representado pela divisão do desvio padrão pela média, assim, quanto maior for a dispersão das observações, maior será o coeficiente.

As observações mensais da linha de crédito *Aquisição de Ativos Financeiros*, rótulo “F”, chamam a atenção devido a três valores extremos observados em sequência: os meses de janeiro, fevereiro e março de 2017 apresentaram as taxas de *default* de 24,19%, 22,85% e 25,10%, respectivamente. Estes valores são muito superiores às outras observações mensais desta linha de crédito que oscilam entre 0,06% e 2,68%, conforme apresentado pela Figura 2.

⁴ O SGS faz parte do Portal de Dados Abertos do Banco Central do Brasil, disponível em: <https://www3.bcb.gov.br/sgspub/> (Consulta realizada em 04/03/2022).

Figura 2. Taxa de default mensal - Aquisição de Ativos Financeiros

A evolução da taxa de *default* mensal da linha de crédito *Aquisição de Ativos Financeiros* durante toda a série temporal (out/2014 a set/2019), apresentou valores extremos (acima de 20%), durante três meses consecutivos (janeiro, fevereiro e março de 2017). A análise da série completa indica uma possível inconsistência no número total de operações de crédito desta linha nos meses em questão. Como tratamento, a taxa de *default* dessas observações foi limitada à taxa de *default* média acrescida de três desvios padrões. Todavia, o risco de eventual inconsistência nas análises a serem elaboradas por esta pesquisa ocasionadas pelo valor extremo observado, assim como pelo tratamento proposto é mitigado pela abordagem por cópulas, de natureza quantílica.

Fonte: elaborado pelo autor.

Para identificar o que pode ter causado estas observações de taxa de *default* tão elevadas, verificou-se o número de operações adimplentes e em *default* da linha de crédito em questão durante todo o período. Enquanto o número de operações em *default* sofre pouca alteração, o número de operações adimplentes apresenta uma queda abrupta nos meses de janeiro, fevereiro e março de 2017, e retorna ao nível histórico de normalidade no mês subsequente. Assim, o aumento da taxa de *default* tem como origem a queda de operações adimplentes. Devido à queda abrupta no número de operações adimplentes e relativa manutenção do número de operações em *default*, é possível que, por algum motivo desconhecido, a base de dados não contenha o registro de todas as operações em andamento da linha de crédito “F” nos meses de janeiro, fevereiro e março de 2017.

A fim de evitar uma superestimação da taxa de *default* ocasionada por uma possível inconsistência da base de dados, e ao mesmo tempo respeitar a ocorrência de observações extremas nos meses em questão, optou-se em limitar a taxa de *default* a três desvios padrões além da média das taxas de *default* observadas, excluindo as observações que excediam este patamar. Todavia, o risco de eventual inconsistência nas análises a serem elaboradas por esta pesquisa ocasionadas pelo valor extremo observado, assim como pelo tratamento proposto é mitigado pela abordagem por cópulas, de natureza quantílica. A Tabela 16 apresenta as observações dos meses de taxas de *default* extremas, os três meses que precedem e sucedem estas observações e os impactos na estatística descritiva ocasionado pelo limitador da taxa.

Tabela 16. Observações e Estatísticas Descritivas – Aquisição de Ativos Financeiros

A – Observações da linha de crédito Aquisição de Ativos Financeiros			
Mês (t)	<i>n. default</i>	$n_{(t-3)}$	Taxa <i>Default</i>
10/2016	58	5445	1,07%
11/2016	66	5391	1,22%
12/2016	37	5268	0,70%
01/2017	67	277	24,19%
02/2017	69	302	22,85%
03/2017	61	243	25,10%
04/2017	95	5291	1,80%
05/2017	96	5197	1,85%
06/2017	59	4883	1,21%
B – Estatística descritiva Aquisição de Ativos Financeiros: Amostra			
Média	Desvio Padrão	Coefficiente de Variação	
2,12%	5,12%	2,4171	
C – Estatística descritiva Aquisição de Ativos Financeiros: Amostra excluindo janeiro a março 2017			
Média	Desvio Padrão	Coefficiente de Variação	Média + 3 DP
0,96%	0,66%	0,6882	2,95%
D – Estatística descritiva Aquisição de Ativos Financeiros: Amostra limitada (max = Média + 3 DP)			
Média	Desvio Padrão	Coefficiente de Variação	
1,06%	0,78%	0,7336	

O Painel A detalha as observações do número de operações em *default* e do número de operações ativas utilizadas para o cálculo da taxa de *default* da linha de crédito *Aquisição de Ativos Financeiros*, nos três meses que precedem a superestimação da taxa de *default*, nos três meses de superestimação da taxa de *default* e nos três meses subsequentes à superestimação dos dados. O Painel B apresenta a média, desvio padrão e coeficiente de variação da amostra sem nenhum tratamento. O Painel C apresenta a média, desvio padrão, coeficiente de variação e média acrescida a três desvios padrões da amostra, excluindo as três observações com taxa de *default* superestimada. O Painel D apresenta a média, desvio padrão e coeficiente de variação da amostra completa, com a taxa de *default* limitada à média acrescida a três desvios padrões obtida no Painel C.

Fonte: elaborado pelo autor.

O ajuste na linha de crédito *Aquisição de Ativos Financeiros* reforça a análise preliminar das linhas de crédito rural em comparação à carteira do SFN. Em conjunto, os dados analisados trazem indícios de que, embora a incidência média de *defaults* seja menor nas linhas de crédito rural em relação ao mercado de crédito brasileiro, a sua distribuição ao longo do tempo apresenta maior volatilidade. Parte desta volatilidade pode ser explicada (i) pela ciclicidade da produção rural, impactando o fluxo de caixa dos devedores e influenciando condições contratuais das operações, com acumulação de vencimentos em determinados períodos, e; (ii) pela exposição conjunta dos produtores a riscos de desastres naturais, eventos que não são homoganeamente distribuídos ao longo do tempo, influenciando na distribuição dos *defaults*.

A Tabela 17 apresenta a matriz de correlação dos coeficientes de Pearson, Kendall e Spearman das linhas de crédito rural e da carteira do SFN.

Tabela 17. Matrizes de correlações – Coeficientes de Pearson, Kendall e Spearman

Painel A: Matriz de correlação coeficiente de Pearson								
Linha de Crédito	A	B	C	D	E	F	G	SFN
A	1,0000	0,7496	0,2490	0,0084	0,6946	0,1861	-0,1494	0,0318
B	-	1,0000	0,6347	0,3246	0,6835	0,4054	-0,2175	0,3424
C	-	-	1,0000	0,3175	0,3262	0,2045	-0,1494	0,1461
D	-	-	-	1,0000	0,4063	0,6841	0,0664	0,7738
E	-	-	-	-	1,0000	0,4955	-0,1480	0,5006
F	-	-	-	-	-	1,0000	-0,0531	0,6578
G	-	-	-	-	-	-	1,0000	0,0767
SFN	-	-	-	-	-	-	-	1,0000
Painel B: Matriz de correlação coeficiente de Kendall								
Linha de Crédito	A	B	C	D	E	F	G	SFN
A	1,0000	0,5685	0,1777	-0,0382	0,4721	0,1138	-0,0455	0,0235
B	-	1,0000	0,4820	0,2019	0,4802	0,2554	-0,1238	0,2577
C	-	-	1,0000	0,2056	0,1686	0,1203	-0,0448	0,1843
D	-	-	-	1,0000	0,2422	0,5559	0,0435	0,5343
E	-	-	-	-	1,0000	0,3207	-0,0376	0,3849
F	-	-	-	-	-	1,0000	0,0499	0,4677
G	-	-	-	-	-	-	1,0000	0,0120
SFN	-	-	-	-	-	-	-	1,0000
Painel C: Matriz de correlação coeficiente de Spearman								
Linha de Crédito	A	B	C	D	E	F	G	SFN
A	1,0000	0,7495	0,2356	-0,1044	0,6426	0,1213	-0,0646	0,0365
B	-	1,0000	0,6436	0,2552	0,6333	0,3564	-0,1715	0,3654
C	-	-	1,0000	0,3250	0,2464	0,1979	-0,0717	0,3098
D	-	-	-	1,0000	0,3313	0,7474	0,0683	0,7593
E	-	-	-	-	1,0000	0,4508	-0,0626	0,5285
F	-	-	-	-	-	1,0000	0,0669	0,6889
G	-	-	-	-	-	-	1,0000	0,0292
SFN	-	-	-	-	-	-	-	1,0000

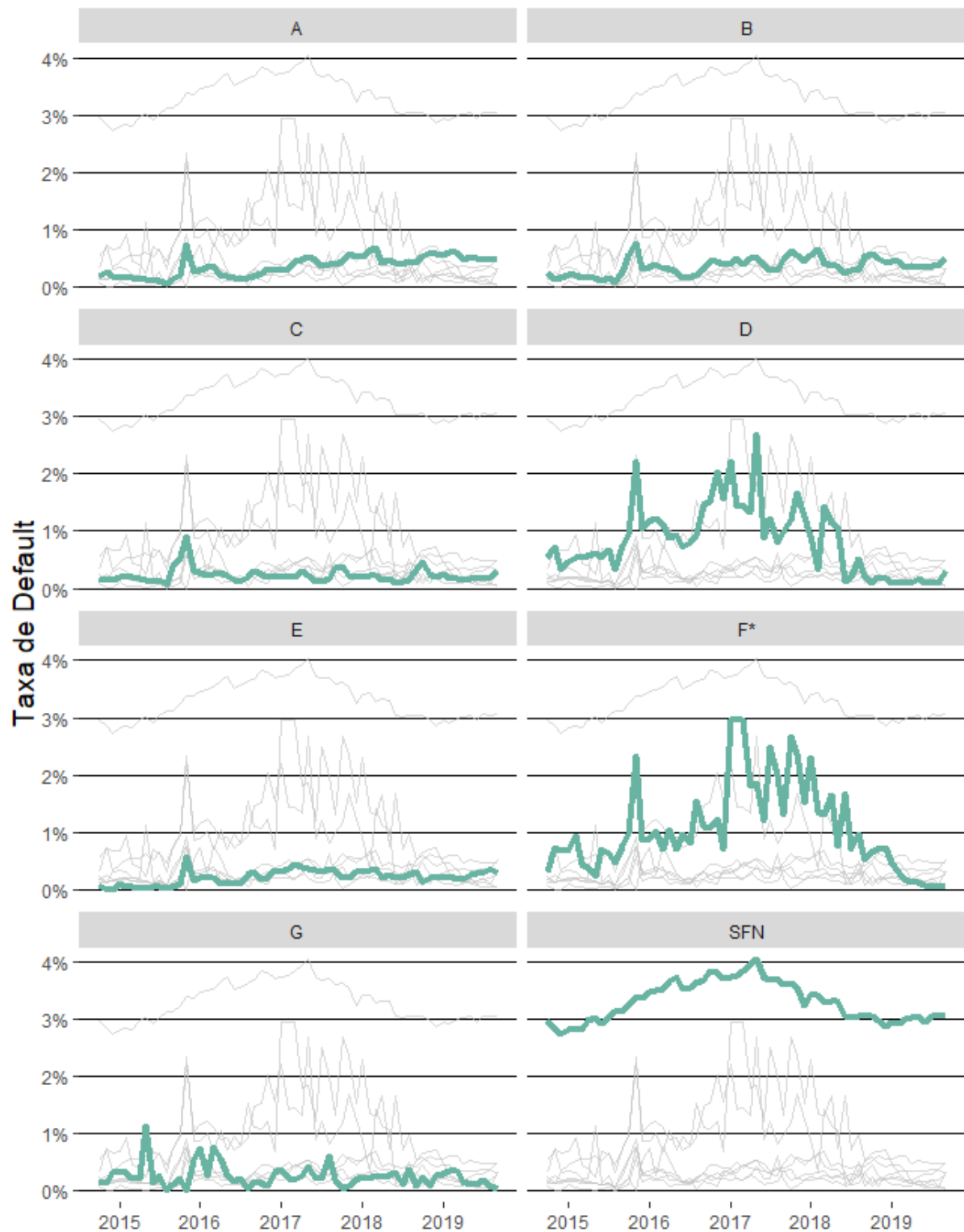
Esta tabela apresenta os coeficientes de correlação de Pearson, Kendall e Spearman, par a par, das sete linhas de crédito rural (de A até G) e o SFN.

Fonte: elaborado pelo autor.

A Tabela 17 é clara em apresentar que diversas linhas apresentam coeficientes de correlação de elevadas magnitudes. Contudo, apenas avaliar a medida-resumo de dependência é insuficiente para generalizar o padrão de ocorrências simultâneas entre as carteiras, que podem apresentar formas funcionais diferentes para cada par de linhas de negócios, ainda que estejam sujeitas a fatores de riscos similares. Especialmente, podem ter padrões conjuntos de ocorrências extremas completamente distintas do nível médio de dependência em quantis centrais.

A verificação empírica sobre os fatores que influenciam a volatilidade observada na taxa de *default* das linhas de crédito rural foge do escopo deste trabalho, mas a constatação desta característica reforça o argumento geral utilizando ao longo deste trabalho de que os modelos de probabilidade de *default* utilizados por outras linhas de crédito, como de varejo ou corporativas, podem não ser adequados ao crédito rural. Assim, faz-se necessário o desenvolvimento de modelos próprios que levem em consideração as características idiossincráticas deste segmento. A Figura 3 apresenta a variação da taxa de *default* observada entre as linhas de crédito rural e todas as operações do SFN. O rótulo da linha de crédito *Aquisição de Ativos Financeiros* foi alterado para “F*”, indicando que se trata da amostra com tratamento de *outliers*.

Figura 3. Taxa de *default* mensal - Linhas de Crédito Rural e SFN



A figura apresenta a variação histórica das observações mensais da taxa de *default* das sete linhas de crédito (de A a G) e todo o SFN durante o período de out/2014 a set/2019. Para melhor identificação visual, a variação histórica foi reproduzida oito vezes, alterando apenas o destaque de uma linha de crédito ou dos dados de todo o SFN. Observa-se que as taxas de *default* das linhas de crédito rural, durante o período observado, são inferiores às taxas de *default* de todo o SFN.

Fonte: elaborado pelo autor.

Os coeficientes de assimetria e de excesso de curtose dão informações sobre o formato das distribuições da taxa de *default*. Uma distribuição apresenta sinais de simetria quando os valores da média e mediana são próximos (no caso da distribuição Normal, essas medidas são coincidentes). O coeficiente de assimetria igual zero implica que a distribuição é simétrica, e, quando negativo, indica que a distribuição tem a média menor do que a mediana, havendo uma concentração de valores do lado direito da distribuição e uma cauda à sua esquerda. Por fim, quando o coeficiente de assimetria é positivo, a média da distribuição é maior do que a sua mediana, havendo uma concentração de valores ao lado esquerdo da distribuição e uma cauda para a direita.

Por sua vez, a curtose mensura a concentração de um conjunto de valores em relação à sua média. Quando positivo, o coeficiente de excesso de curtose indica que a distribuição dos valores é menos concentrada em torno de sua média em comparação a uma distribuição Normal (a curtose de uma Normal padrão é igual a 3), implicando ser uma distribuição com cauda pesada. A Tabela 18 apresenta os coeficientes de assimetria e excesso de curtose das linhas de crédito rural e da carteira do SFN.

Tabela 18. Assimetria e Curtose – taxas de *default* por linha de crédito e SFN

Linha de Crédito	Assimetria	Excesso de Curtose
A	0,0291	1,8769
B	0,2583	2,7096
C	3,1984	16,9133
D	0,8056	3,3585
E	0,1827	2,7771
F*	0,9478	3,1209
G	2,0863	8,9912
SFN	0,2553	1,7522

Esta tabela apresenta os coeficientes de assimetria e curtose das taxas de *default* mensais das sete linhas de crédito (de A a G) e todo o SFN durante o período de out/2014 a set/2019. **Assimetria:** coeficientes menores do que zero indicam distribuições assimétricas à esquerda; coeficientes maiores do que zero indicam distribuições assimétricas à direita; coeficientes iguais à zero indicam distribuições simétricas. **Excesso de Curtose:** coeficientes maiores do que zero indicam distribuições leptocúrticas, de cauda mais pesada do que a distribuição normal; coeficientes menores do que zero indicam distribuições platícúrticas, de caudas mais leves do que a distribuição normal; coeficientes iguais a zero indicam distribuições mesocúrticas, com a mesma estrutura de cauda da distribuição normal.

Fonte: elaborado pelo autor.

Em todas as linhas de crédito rural e na carteira de crédito do SFN, os coeficientes de assimetria são positivos, indicando que as distribuições das taxas de *default* apresentam, em alguma escala, cauda à direita. Sobre os coeficientes de excesso de curtose, nota-se que todos são positivos, indicando que as densidades das caudas são mais elevadas do que em uma distribuição Normal. O excesso de curtose do SFN é o menor da amostra, implicando que a distribuição da taxa de *default* de todas as operações de crédito registradas no Brasil, embora levemente assimétrica, tem maior concentração próximo de sua média.

Embora ambos os coeficientes auxiliem a identificar o formato das distribuições, eles não são suficientes para identificar se a distribuição avaliada segue ou não a normalidade. Para realizar tal inferência, existem testes estatísticos de verificação de normalidade como os testes de Shapiro-Wilk (Shapiro & Wilk, 1965) e de Jarque-Bera (Jarque & Bera, 1987), entre outros.

O teste de Shapiro-Wilk verifica a hipótese de que um conjunto de observações seja proveniente de uma distribuição Normal de média e variância desconhecidas, com base na estatística do teste (W_{cal}). Valores baixos de W_{cal} indicam que a distribuição da variável em estudo não é normal. O teste de Jarque-Bera verifica se a assimetria e curtose da distribuição de um conjunto de dados são adequadas à distribuição Normal. A estatística do teste, JB , é sempre positiva e quanto maior for, menor é a aproximação da distribuição observada de uma curva normal. Ambos os testes têm como hipótese nula H_0 a suposição de que a distribuição dos dados da amostra é aderente à distribuição Normal.

A Tabela 19 apresenta as estatísticas W_{cal} e JB das operações de crédito rural e do SFN e seus respectivos p -valores.

Tabela 19. Testes de Normalidade – taxas de *default* por linha de crédito e SFN

Linha de Crédito	Teste Shapiro-Wilk			Teste Jarque-Bera		
	W_{cal}	p -valor	H_0 é rejeitada?*	JB	p -valor	H_0 é rejeitada?*
A	0,95485	0,02648	Sim	3,1621	0,20580	Não
B	0,98264	0,54990	Não	0,8782	0,64460	Não
C	0,70757	0,00001	Sim	586,25	0,00001	Sim
D	0,92905	0,00181	Sim	6,8117	0,03318	Sim
E	0,96625	0,09544	Não	0,4580	0,79530	Não
F*	0,90143	0,00015	Sim	9,0189	0,01100	Sim
G	0,81802	0,00001	Sim	133,26	0,00001	Sim
SFN	0,93108	0,00220	Sim	4,5439	0,10310	Não

Teste de Shapiro-Wilk: A estatística do teste W_{cal} indica a aderência da distribuição avaliada em relação à curva Normal. Quanto menor for o valor da estatística, menor é a aderência da distribuição à curva Normal. A hipótese nula do teste (H_0) é a de aderência da distribuição avaliada à distribuição normal. Quando H_0 é rejeitada, o resultado do teste indica a não-normalidade da distribuição dos dados. **Jarque-Bera:** o teste verifica se a assimetria e curtose da distribuição de um conjunto de dados são adequadas à distribuição Normal. Quanto maior for o valor da estatística JB , menor é a aproximação da distribuição observada a uma curva Normal. A hipótese nula do teste (H_0) é a de aderência da distribuição avaliada à distribuição normal. Quando H_0 é rejeitada, o resultado do teste indica a não-normalidade da distribuição dos dados.

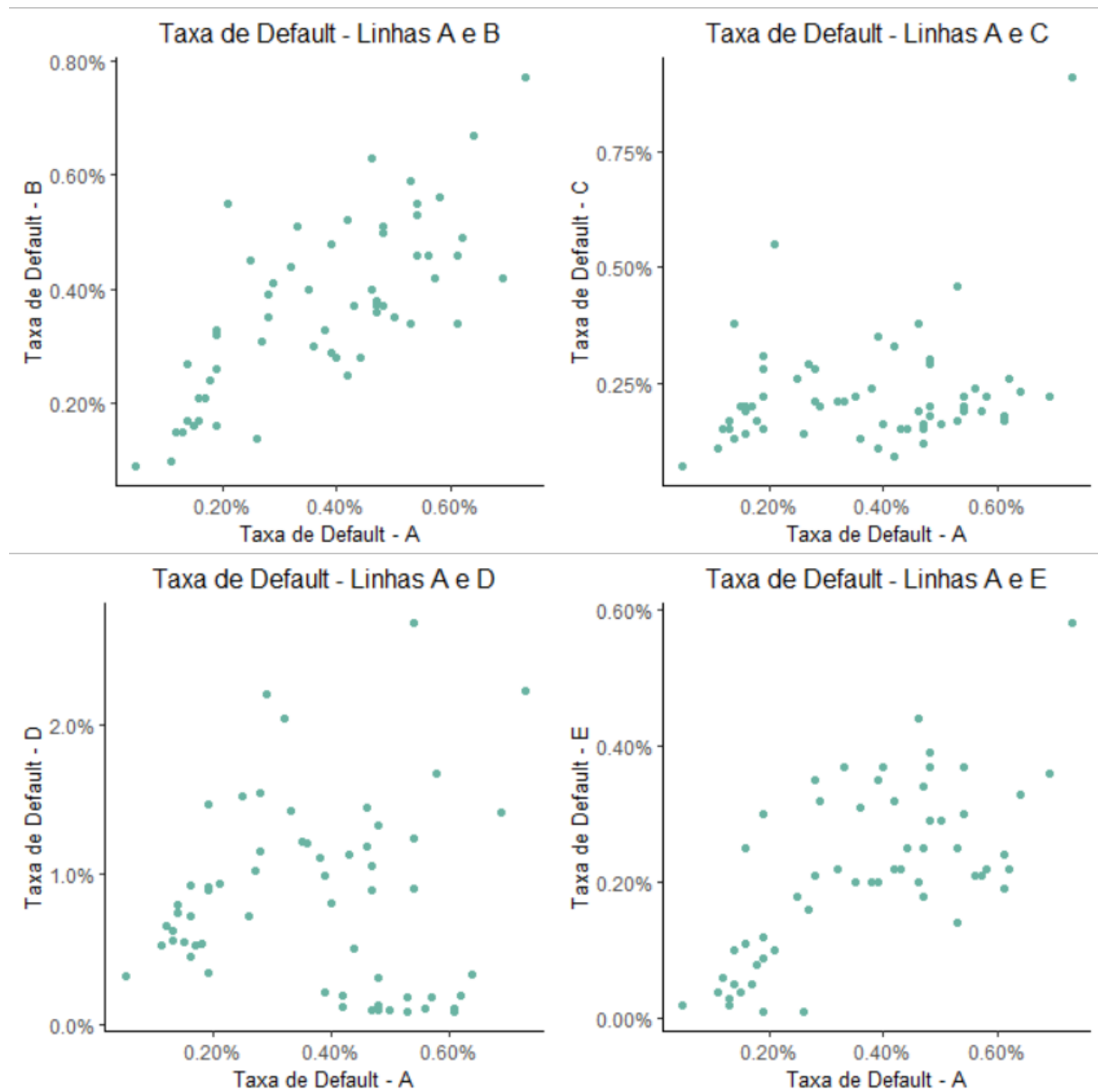
*Rejeita-se H_0 a um nível de significância de 5% (p -valor < 0,05)

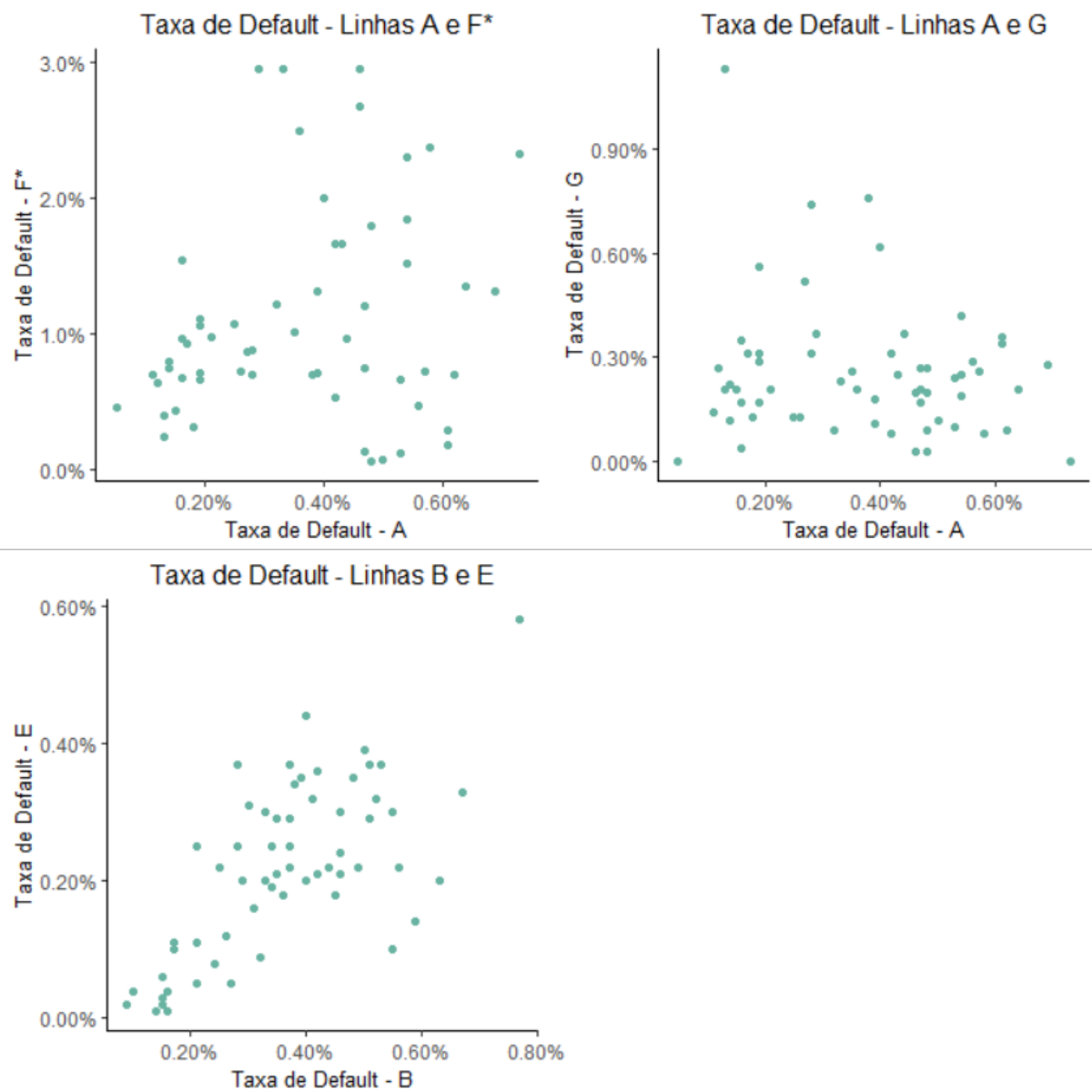
Fonte: elaborado pelo autor.

Os testes indicam que não podemos rejeitar a hipótese de que as linhas de crédito *B* e *E* possuam distribuição Normal em suas distribuições de taxa de *default*. Já as distribuições das taxas de *default* das linhas *C*, *D*, *F** e *G* não são aderentes ao padrão Normal. Por fim, as distribuições das taxas de *default* das linhas de crédito *A* (Pronaf) e da carteira de crédito do SFN apresentam um resultado ambíguo: não têm aderência à distribuição Normal segundo o teste Shapiro-Wilk, mas é possível rejeitar a sua aderência à Normalidade, segundo o teste Jarque-Bera. Embora as taxas de *default* de algumas linhas de crédito possam ser representadas por distribuições Normais, a estimação da taxa de *default* conjunta pode ser mais aderente a outras distribuições, por conta da estrutura de dependência entre as linhas.

Para auxiliar a identificação do comportamento conjunto das taxas de *default* das operações, a Figura 4 apresenta os gráficos de dispersão das observações mensais das taxas de *default* dos seguintes pares: (i) linha de crédito *A* (Pronaf) com todas as demais linhas de crédito rural, e; (ii) das linhas de crédito rural com distribuições marginais de taxa de *default* aderentes à Normal ($B \times E$).

Caso se verifique a existência de observações no quadrante superior direito indica a ocorrência conjunta de elevadas taxas de *default* entre as linhas de crédito, sendo um indício de uma estrutura de dependência de cauda. Em uma análise preliminar, é possível identificar que a linha de crédito *A* (Pronaf) só não apresenta uma tendência de (de)crescimento conjunto de taxas de *default* elevadas com a linha de crédito *G* (Garantia de preço e estocagem).

Figura 4. Representação conjunta das taxas de *default* das linhas de crédito rural avaliadas



Esta figura apresenta a representação conjunta das taxas de *default* dos pares formados pela linha de crédito A (Pronaf) com todas as demais linhas de crédito rural, e; (ii) das linhas de crédito rural com distribuições marginais de taxa de *default* aderentes à Normal (B×E). A incidência de observações no quadrante superior direito indica a ocorrência conjunta de elevadas taxas de *default* entre as linhas de crédito, sendo um indício de uma estrutura de dependência de cauda.

Fonte: elaborado pelo autor.

7. Modelagem e Análises de Resultados

7.1. PD de Portfólio - Estimação dos parâmetros e escolha da cópula

Nesta seção serão estimadas as distribuições probabilidades de *default* conjunta da linha de crédito Pronaf com as demais linhas de crédito rural. O enfoque na linha de crédito Pronaf (A) dá-se pela sua importância no desenvolvimento da agricultura familiar e atingimento dos objetivos da Agenda 2030 (Brasil, 2021; Campbell, 2017; Cervantes-Godoy & Dewbre, 2010; Jong & Vijge, 2021; OECD & FAO, 2021). Também será estimada a estrutura de dependência das linhas Pronamp (B) e Investimento (E) devido ao interesse de se verificar a forma da estrutura da distribuição conjunta, tendo em vista que as distribuições marginais destas linhas de crédito são aderentes à Normal.

O *default* histórico foi assumido como estacionário, i.e., a taxa de *default* observada serve como *proxy* da probabilidade de *default* futura, uma vez que se espera manutenção dos comportamentos da média e da variância, além de estabilidade da estrutura de dependência ao longo do tempo. As distribuições de PD conjunta geradas são em nível de portfólio, uma vez que foram elaboradas com base na taxa de *default* do conjunto de linhas de crédito, agrupando as operações individuais nestas linhas. Dez famílias de cópulas serão testadas para representar a distribuição de probabilidade de *default* conjunta. A Tabela 20 apresenta as classes e famílias de cópulas utilizadas na estimativa e as suas respectivas estruturas de dependência.

Tabela 20. Classes e Famílias de cópulas e suas respectivas estruturas de dependência

Classe	Família	Estrutura de Dependência
Elíptica	Normal (Gaussiana)	Simétrica sem dependência de cauda
Elíptica	t de Student	Simétrica com dependência de cauda à esquerda e à direita
Plackett	Plackett	Simétrica sem dependência de cauda
Arquimediana	Frank	Simétrica sem dependência de cauda
Arquimediana	Clayton	Dependência de cauda à esquerda
Arquimediana	Gumbel	Dependência de cauda à direita
Arquimediana	Joe	Dependência de cauda à direita
Valores Extremos (EV)	Galambos	Dependência de cauda à direita
Valores Extremos (EV)	Hüsler-Reiss	Dependência de cauda à direita
Valores Extremos (EV)	t -EV	Dependência de cauda à direita

Esta tabela apresenta as Classes e Famílias de cópulas e suas respectivas estruturas de dependência aplicadas nesta pesquisa. Ao todo, quatro classes de cópulas, dez famílias e quatro estruturas de dependência foram avaliadas.

Fonte: adaptado de Crook & Moreira (2011).

A cópula que melhor representa cada par de distribuição conjunta será determinada por meio do teste *Goodness-of-Fit* híbrido (GoF_{hb}), proposto por Zhang et al. (2016). Os testes *Goodness-of-Fit* verificam a aderência da cópula elegível ao conjunto de dados disponível. Sob a hipótese nula, assume-se que a cópula proposta é uma boa representação da distribuição conjunta das observações. Assim, dado um conjunto de cópulas que se deseja verificar a melhor aderência à distribuição empírica, elege-se aquela que obteve maior *p-valor* no teste, indicando a não rejeição da hipótese nula de aderência da cópula testada.

O teste GoF_{hb} agrega e pondera o resultado de dois ou mais testes *Goodness-of-Fit*, combinando seus poderes explicativos. A operacionalização do teste será feita usando o software estatístico R. Okhrin et al. (2021) detalham a estrutura teórica dos testes *GoF* passíveis de serem agregados pelo teste híbrido, classificando-os em seis grupos: (i) *Empirical Copula Process*; (ii) *Kendall's process*; (iii) Transformação de *Rosenblatt*; (iv) Transformação de Cópulas Arquimedianas; (v) Estimativa de densidade de Kernel; (vi) Teste White, e; (vii) Teste de Validação Cruzada. A forma funcional dos testes pode restringir a sua aplicação a certas classes de cópulas. A Tabela 21 apresenta os testes *GoF* utilizados na presente análise e as respectivas classes de cópulas as quais eles foram aplicados.

Tabela 21. Testes *Goodness-of-Fit* agregados pelo *GoF* Híbrido para cada classe de cópula

Testes <i>Goodness-of-Fit</i> agregados pelo teste GoF_{hb}			Classes de cópulas elegíveis aos testes			
Teste	Grupo	Referência	Elíptica	Plackett	Arquimediana	EV
<i>CvM</i>	<i>Empirical Copula Process</i>	Genest et al. (2009)	X	X	X	X
<i>KS</i>	<i>Empirical Copula Process</i>	Genest et al. (2009)	X	X	X	X
<i>KendallCvM</i>	<i>Kendall's process</i>	Genest et al. (2006)	X	X	X	X
<i>KendallKS</i>	<i>Kendall's process</i>	Genest et al. (2006)	X	X	X	X
<i>RosenblattChisq</i>	Transformação de Rosenblatt	Breymann et al. (2003)	X	X	X	
<i>RosenblattGamma</i>	Transformação de Rosenblatt	Breymann et al. (2003)	X	X	X	
<i>RosenblattSnB</i>	Transformação de Rosenblatt	Genest et al. (2009)	X	X	X	
<i>RosenblattSnC</i>	Transformação de Rosenblatt	Genest et al. (2009)	X	X	X	
<i>ArchmChisq</i>	Transf. de Cópulas Arquimedianas	Hering & Hofert (2015)			X	
<i>ArchmGamma</i>	Transf. de Cópulas Arquimedianas	Hering & Hofert (2015)			X	
<i>ArchmSnB</i>	Transf. de Cópulas Arquimediana	Hering & Hofert (2015)			X	
<i>ArchmSnC</i>	Transf. de Cópulas Arquimedianas	Hering & Hofert (2015)			X	
<i>Kernel</i>	Estimativa de densidade de Kernel	Scaillet (2007)	X	X	X	X
<i>White</i>	Teste <i>White</i>	Huang & Prokhorov (2014)	X		X	
<i>PIOSRn</i>	Teste de Validação Cruzada	Zhang et al. (2016)	X	X	X	
<i>PIOSTn</i>	Teste de Validação Cruzada	Zhang et al. (2016)	X	X	X	

Esta tabela apresenta a relação dos dezesseis testes *Goodness-of-Fit* passíveis de serem agregados pelo teste *Goodness-of-Fit* híbrido, indicando seus respectivos grupos de teste e referência bibliográfica. Ao lado direito da tabela é marcado com "X" qual teste é aplicável a cada uma das quatro classes de cópulas utilizadas nesta pesquisa.

Fonte: elaborado pelo autor.

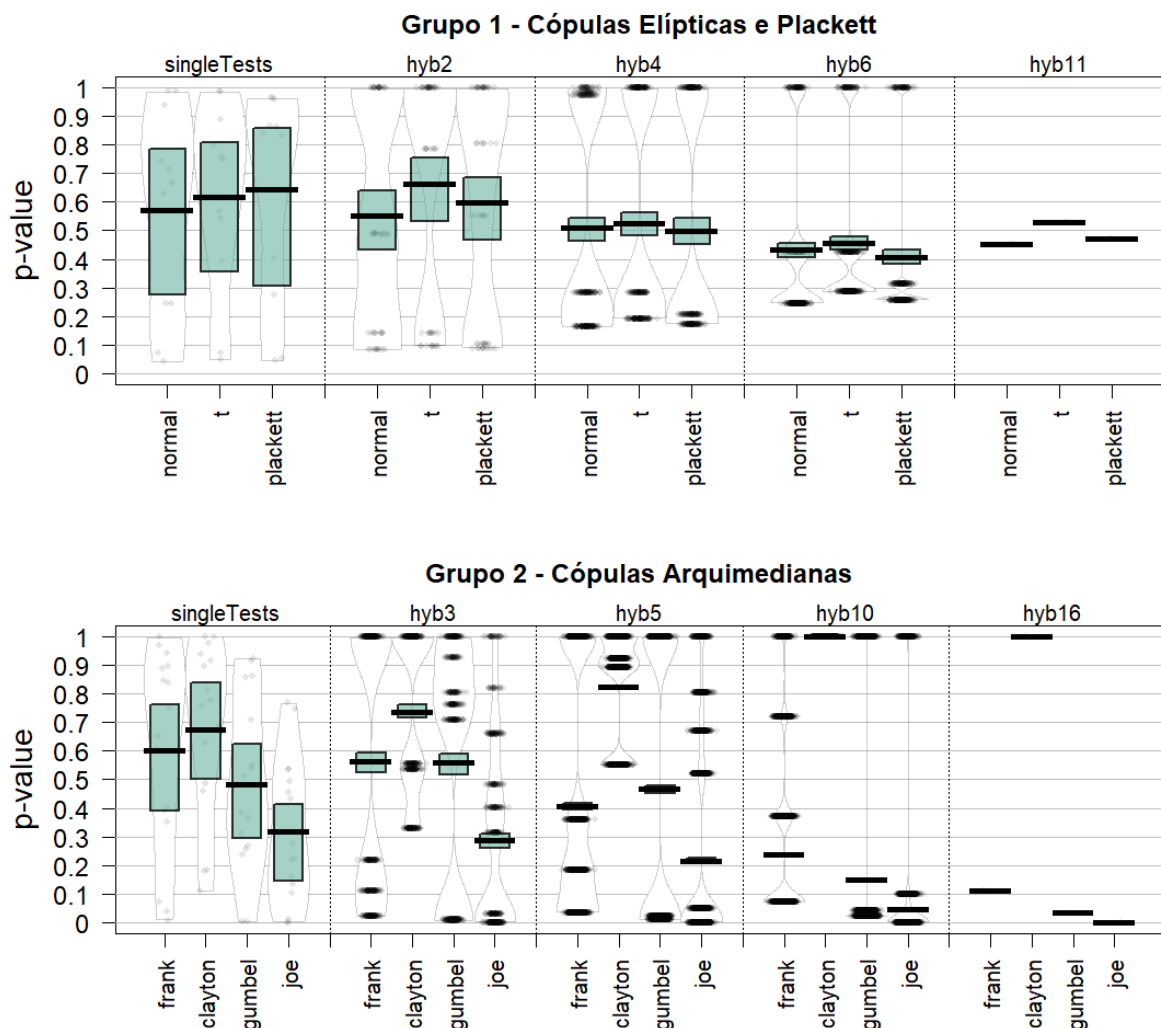
Para efetuar o teste GoF_{hb} , estima-se os parâmetros das cópulas pelo método *Canonical Maximum Likelihood*. Caso não seja possível a estimação por este método, o parâmetro é estimado por meio da inversão do Kendall τ . Sob um processo de *bootstrapping*, todas as cópulas foram testadas mil vezes em todos os testes *GoF* passíveis de serem aplicados. A análise da cópula de melhor aderência à distribuição empírica foi efetuada em duas etapas.

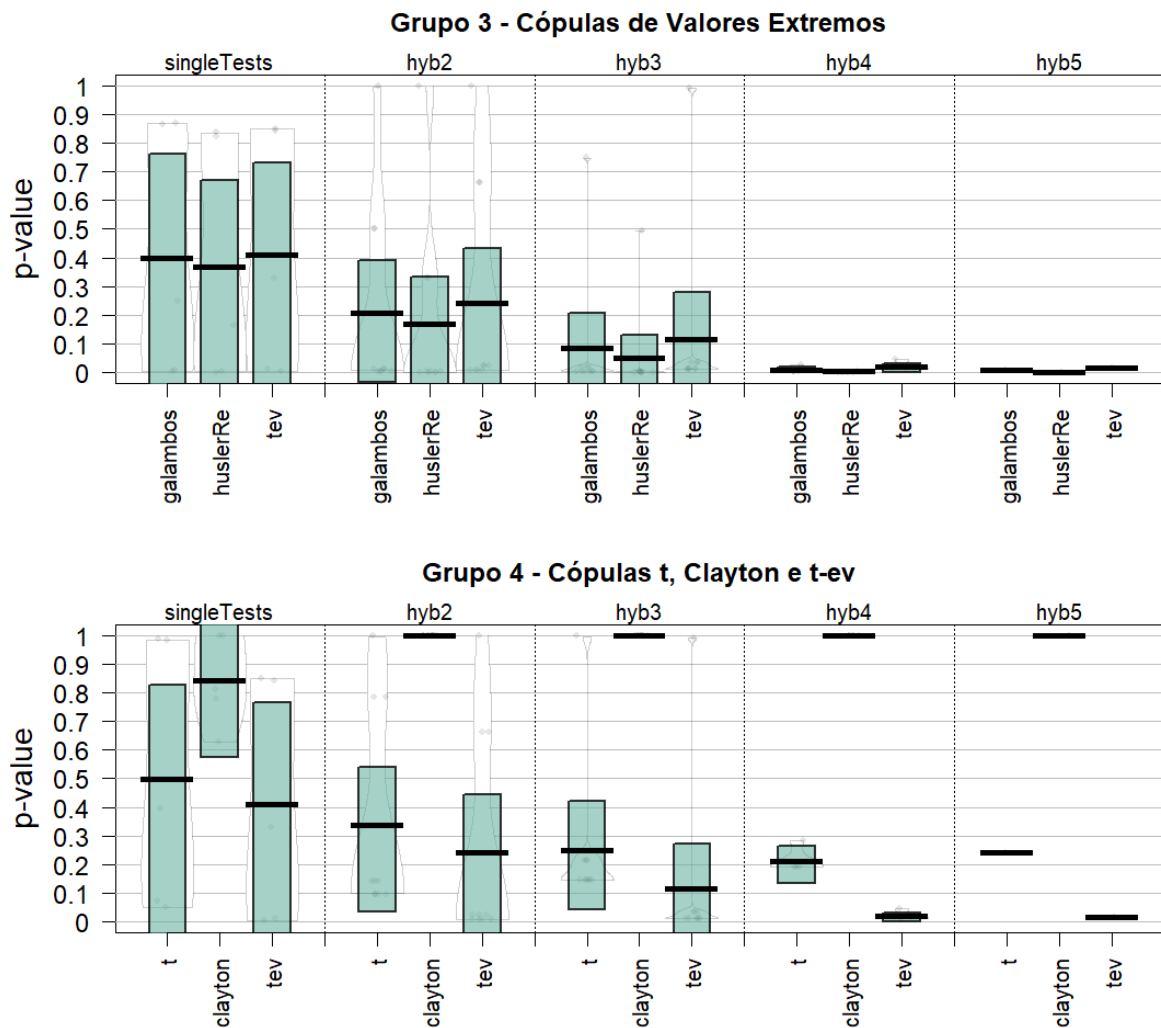
Na primeira, as cópulas foram segregadas em três grupos: (1) elípticas e Plackett; (2) arquimedianas, e; (3) de valores extremos (EV). Cada grupo foi submetido aos testes GoF elegíveis às suas respectivas classes de cópulas, subsidiando a construção do teste GoF_{hb} . Assim, seguindo os critérios do teste híbrido, foi escolhida a cópula com melhor aderência aos dados em cada um dos três grupos.

Na segunda etapa, o Grupo 4 é formado com as três cópulas de melhor aderência nos Grupos de 1 a 3. Se, por algum motivo, alguma cópula for excluída do teste híbrido de seu grupo (o que pode ocorrer pela impossibilidade de execução de algum teste GoF), ela será testada novamente no Grupo 4. Caso haja empate entre as cópulas com maior p -valor nos Grupos de 1 a 3, ambas as cópulas serão classificadas para o teste no Grupo 4. No Grupo 4, as famílias de cópulas foram submetidas aos cinco testes GoF comuns a todas as classes (CvM , KS , $KendallCvM$, $KendalKS$ e $Kernel$), para então se escolher a cópula de maior aderência aos dados, entre todas as cópulas testadas.

A Figura 5 ilustra este processo de forma gráfica, para a escolha da cópula mais adequada à distribuição de PD conjunta das linhas de crédito rural A e B : $PD(A \times B)$.

Figura 5. Teste GoF híbrido $PD(A \times B)$





A Figura 5 ilustra o processo utilizado para a escolha da cópula de melhor aderência à distribuição de probabilidade de *default* conjunta PD(AxB), com base no teste *GoF* híbrido. Na primeira etapa, as cópulas foram segregadas em três grupos: (1) elípticas e Plackett; (2) arquimedianas, e; (3) de valores extremos (EV). Cada grupo foi submetido aos testes *GoF* elegíveis às suas respectivas classes de cópulas. Seguindo os critérios do teste híbrido, foi escolhida a cópula com melhor aderência aos dados em cada um dos três grupos, sendo representada por aquela com maior *p*-valor. Na segunda etapa, o Grupo 4 foi formado com as três cópulas de melhor aderência nos Grupos de 1 a 3 (*t*, Clayton e *t*-EV) e outro teste *GoF* híbrido foi processado, tendo como a cópula mais aderente aos dados, indicada pelo maior *p*-valor, a cópula de Clayton. Assim, após as duas etapas e análise dos quatro grupos, dentre as dez famílias de cópulas testadas, a cópula de Clayton é a de maior adesão à distribuição de probabilidade de *default* conjunta PD(AxB).

Fonte: elaborado pelo autor.

Cada gráfico da Figura 5 é formado por cinco eixos verticais, responsáveis por dividir a quantidade de testes *GoF* utilizados para compor o *p*-valor do teste híbrido. O primeiro eixo apresenta a distribuição do *p*-valor de todos os testes *GoF* elegíveis às cópulas testadas, aplicados separadamente. O último eixo apresenta *p*-valor do teste híbrido que incorporou todos os testes *GoF* possíveis. Os pontos cinzas são as observações brutas, resultantes das 1000 repetições do processo de *bootstrapping*; a linha cinza indica a distribuição do *p*-valor; a barra horizontal preta indica o valor médio do *p*-valor, e; a barra vertical verde indica o intervalo de confiança de 95% da distribuição de *p*-valores.

O gráfico do Grupo 1 apresenta o teste *GoF* híbrido das cópulas elípticas e de Plackett. Observa-se que no primeiro eixo, representando a média dos testes *GoF* individuais, o intervalo de confiança do *p*-valor é elevado para as três famílias, dificultando a conclusão de qual cópula é mais aderente aos dados (maior *p*-valor, afastando a possibilidade de rejeição da hipótese nula). À medida que o número de testes *GoF* são agregados pelo teste *GoF*_{hb}, a variabilidade

da distribuição do p -valor decai para todas as cópulas testadas. O gráfico apresenta essa progressão a partir do segundo eixo, *hyb2*, que traz todas as combinações possíveis de dois testes *GoF*, seguindo o terceiro (*hyb4*) e quarto eixo (*hyb6*), representando todas as combinações de 4 e 6 testes respectivamente, até chegar no último eixo, *hyb11*, que representa o teste híbrido formado por todos os 11 testes *GoF* disponibilizados às cópulas elípticas e de Plackett.

O gráfico do Grupo 1 indica a melhor performance da cópula t de Student, seguido pelos gráficos dos Grupos 2 e 3, os quais indicam respectivamente, a melhor performance da cópula de Clayton e de t -EV. O gráfico do Grupo 4 apresenta o teste final, ao comparar as três cópulas mais aderentes em seus grupos, informando o resultado de que, entre as 10 famílias de cópulas avaliadas, os dados da $PD(A \times B)$ são melhores representados pela cópula de Clayton.

Para validar o resultado do teste GoF_{hb} , a aderência das cópulas avaliadas aos dados reais também foi testada pelo método dos momentos (MME), o qual indica que a cópula mais aderente à distribuição empírica é aquela de maior estatística de teste.

Este processo foi repetido para estimar a cópula mais aderente para todos os 7 pares de probabilidades de *default* conjunta das linhas de crédito rural avaliadas. O Anexo A traz para todos os pares de linhas de crédito (i) a análise gráfica do teste GoF_{hb} ; (ii) a informação do p -valor estimado nos testes individuais e híbridos, e; (iii) a estatística dos testes MME.

A Tabela 22 apresenta o resultado do teste GoF_{hb} de duas etapas e a validação do resultado por meio dos testes MME baseados nas medidas de concordância *Kendall Tau* (*MME Kendall Tau*) e no *Spearman Rho* (*MME Spearman Rho*).

Tabela 22. PD Conjunta – Cópulas de maior aderência segundo testes GoF_{hb} e MME

PD Conjunta	GoF_{hb}	MME Kendall Tau	MME Spearman Rho
AxB	Clayton	Clayton	Clayton
AxC	Clayton	Clayton	Clayton
AxD	Clayton	t -EV	t -EV
AxE	Clayton	Clayton	Clayton
AxF	t -EV	Normal	Normal
AxG	t -EV*	Gumbel*	Gumbel*
BxE	Clayton	Clayton	Clayton

Esta tabela apresenta as cópulas de maior aderência aos sete pares de distribuições bivariadas avaliados, segundo os testes GoF_{hb} , MME Kendall Tau e MME Spearman Rho.

* Os testes indicaram mais de uma família de cópula com potencial de maior aderência.

Fonte: elaborado pelo autor.

A cópula de Clayton apresentou melhor aderência na maior parte dos testes executados, sendo escolhida em aproximadamente 62% dos testes. Esta cópula tem como característica a representação da dependência de cauda à esquerda, ou seja, nos quadrantes valores inferiores das distribuições conjuntas. Tal resultado é coerente à uma análise de distribuição de probabilidade de *default* (cujas evidências são apresentadas na Figura 3), uma vez que, por ser o *default* um evento tipicamente de baixa frequência (algo a ser evitado pelos bancos), há elevada concentração de observações deste tipo de dado nos quantis inferiores da distribuição conjunta, sendo tal característica observada em Andrievskaya & Penikas (2012). Fenech et al. (2015) também verificaram a melhor aderência da cópula de Clayton a um conjunto de distribuições de probabilidade de *default* de linhas de crédito de varejo. Como veremos adiante, mesmo nos casos em que há maior dependência na cauda inferior, eventos extremos de *default* são perceptíveis e modelados pela cópula de Clayton.

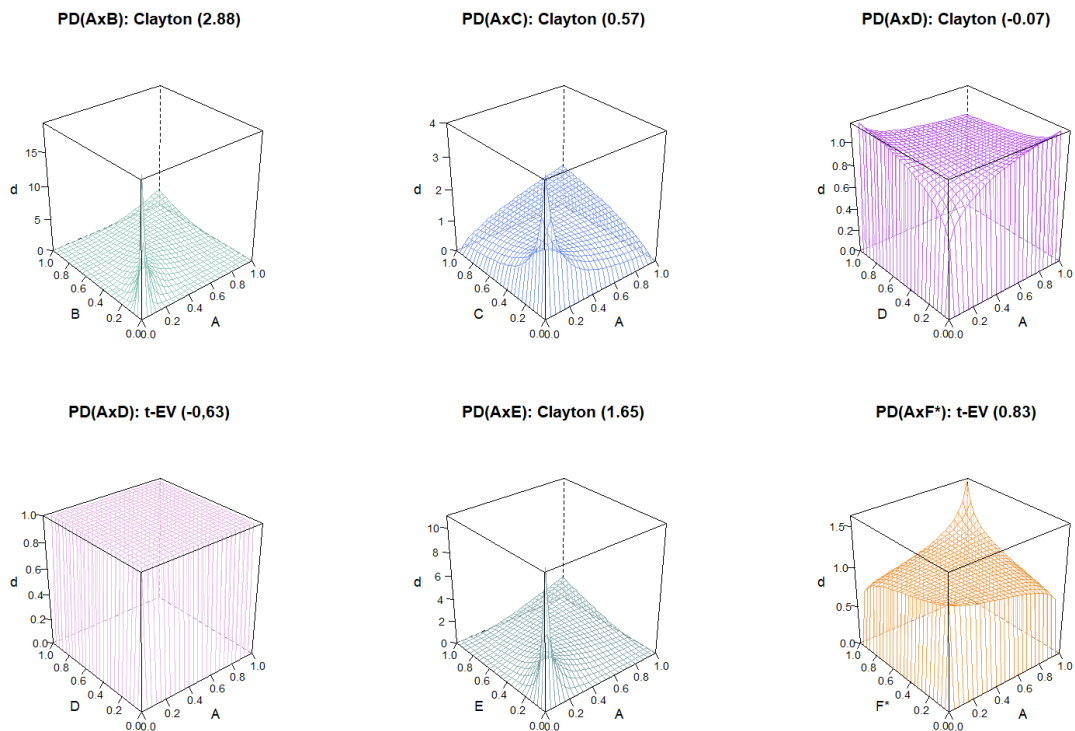
Ao menos uma cópula de valor extremo foi a mais aderente aos pares de PD conjunta $A \times D$, $A \times F$ e $A \times G$. No conjunto de testes, estas cópulas foram a melhor escolha em

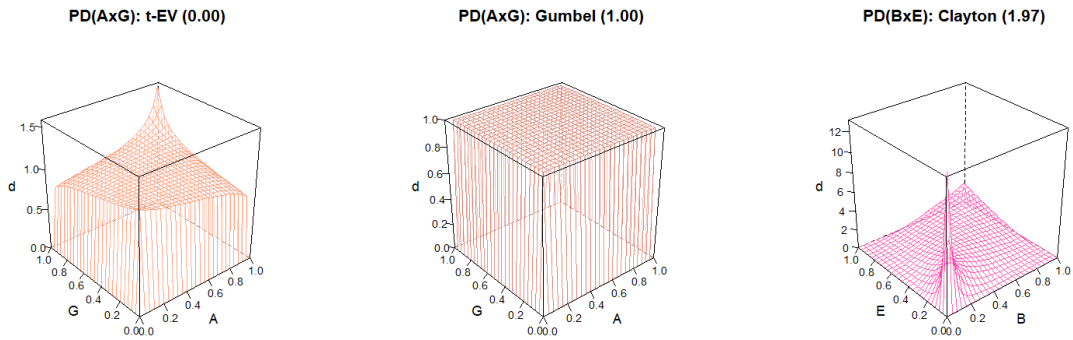
aproximadamente 19% das vezes, indicado uma concentração elevada de eventos extremos conjuntos, capaz de sobrepor a natural concentração de eventos nos quantis inferiores.

Por fim, a baixa aderência da distribuição normal (aproximadamente em 9,5% dos testes), reforça o argumento geral desta pesquisa de que os modelos de risco de crédito de um portfólio de crédito rural devem levar em conta os eventos de cauda, não devendo assumir diretamente a premissa dos modelos tradicionais e regulatório (Basileia) de Normalidade da distribuição conjunta dos dados. Tampouco o coeficiente de correlação linear de Pearson como a medida mais apropriada para a modelagem de capital regulatório de portfólios de crédito.

Finalizando a análise das distribuições conjuntas de probabilidade de *default*, a Figura 6 traz a representação gráfica da densidade de probabilidade conjunta das cópulas de maior aderência segundo os testes GoF_{hb} e MME para todos os pares avaliados. Nos subtítulos, indicamos qual é a cópula e o parâmetro representado (em parênteses).

Figura 6. Densidade de probabilidade conjunta das cópulas de maior aderência aos pares de PD





Fonte: elaborado pelo autor.

7.2. PD de Portfólio – Simulações de Normalidade e de Cópulas

Seguindo o trabalho de Crook & Moreira (2011), as taxas de *default* conjuntas efetivamente observadas serão comparadas com valores simulados em dois cenários (i) assumindo a normalidade da distribuição, e; (ii) a partir da(s) cópula(s) de maior aderência. A Tabela 23 apresenta os parâmetros da distribuição Normal bivariada e das cópulas de maior aderência, obtidos pelos testes GoF_{hb} e MME .

Tabela 23. PD Conjunta – Parâmetros da distribuição Normal Bivariada e das cópulas de maior aderência

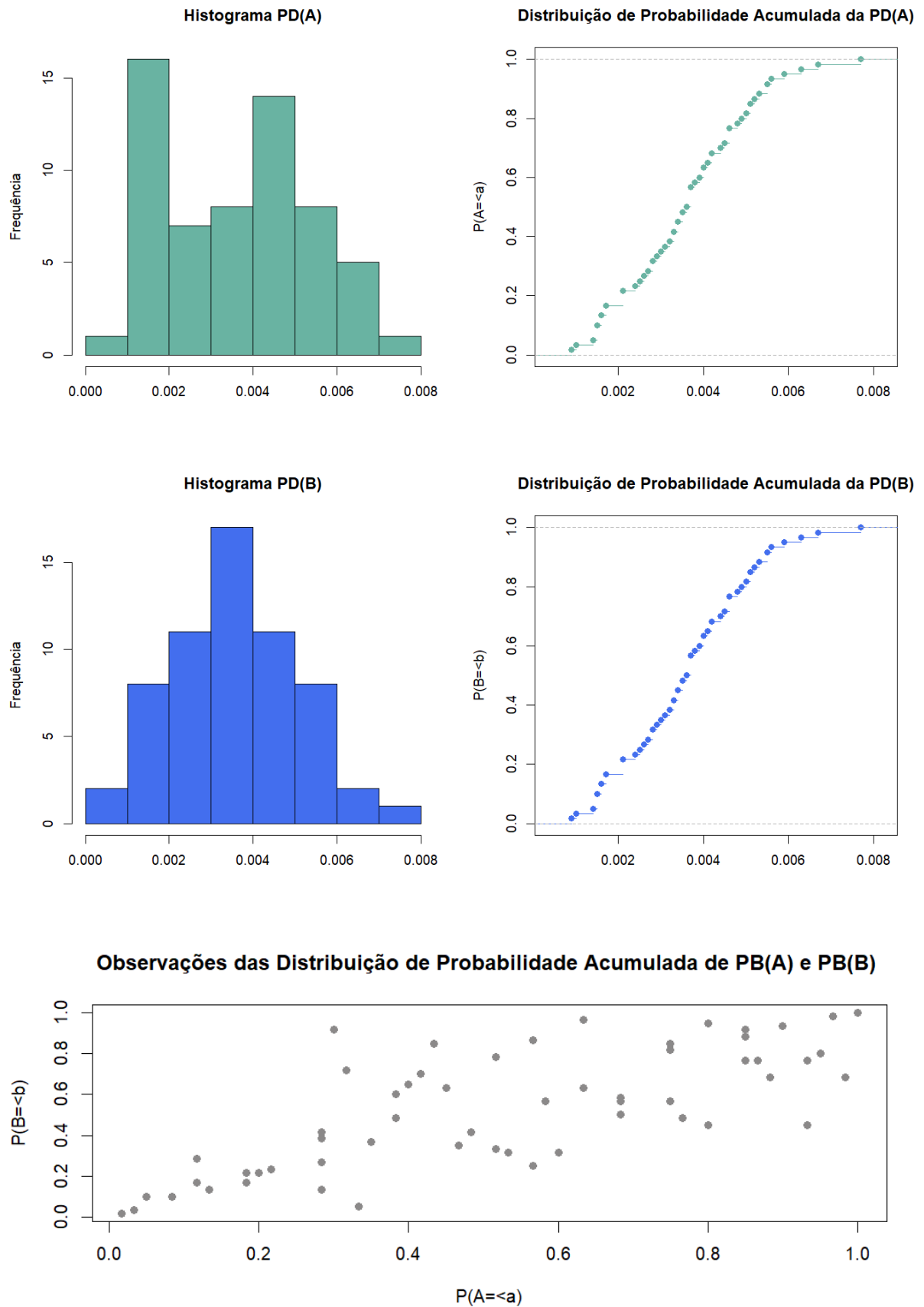
PD Conjunta	Normal (ρ)	GoF_{hb} (θ)	MME (θ)
AxB	0,7974	2,8873	-
AxC	0,3581	0,5776	-
AxD	-0,0537	-0,0737	0,6396
AxE	0,6751	1,6564	-
AxF^*	0,1365	0,8352	-
AxG	-0,0879	0,000	1,000
BxE	0,6842	1,9753	-

Esta tabela apresenta os parâmetros dos sete pares de distribuições de probabilidade conjunta, assumindo o pressuposto de normalidade e das cópulas de maior aderência à distribuição empírica.

Fonte: elaborado pelo autor.

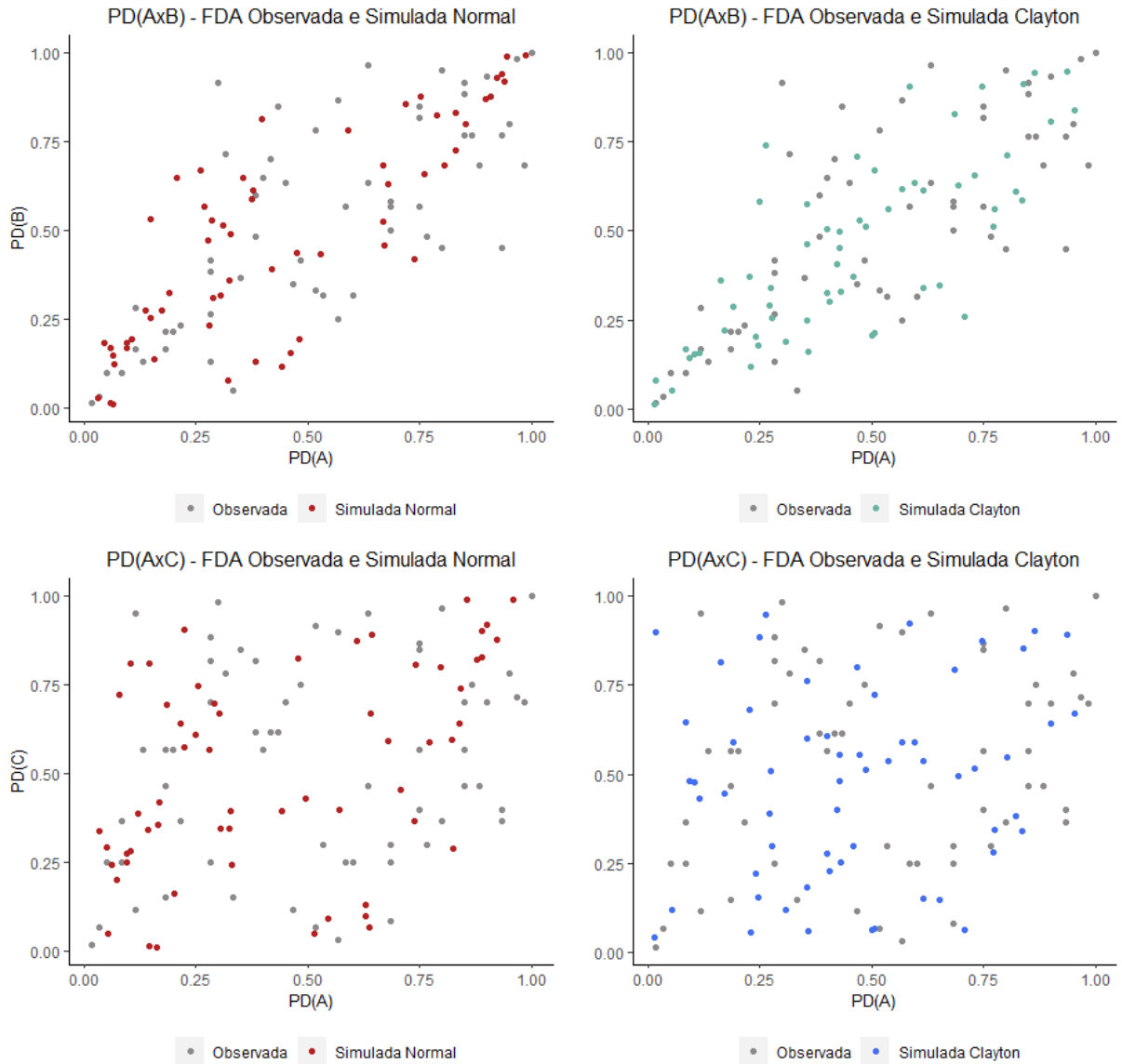
A função Cópula fornece a distribuição de probabilidade acumulada das probabilidades de *default* (PD). A Figura 7 auxilia esta interpretação, ao ilustrar a distribuição de frequência e de probabilidade acumulada das distribuições marginais $PD(A)$ e $PD(B)$ e da distribuição conjunta $PD(A \times B)$.

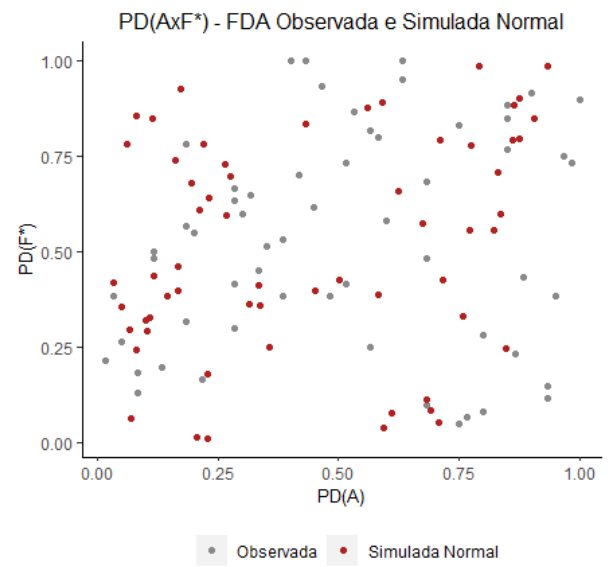
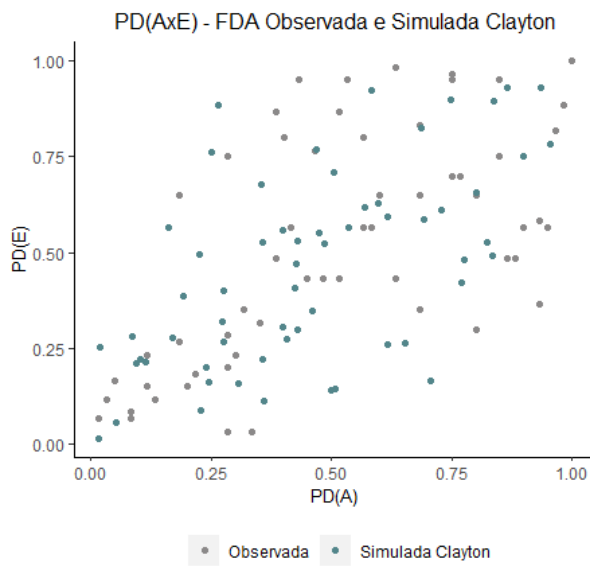
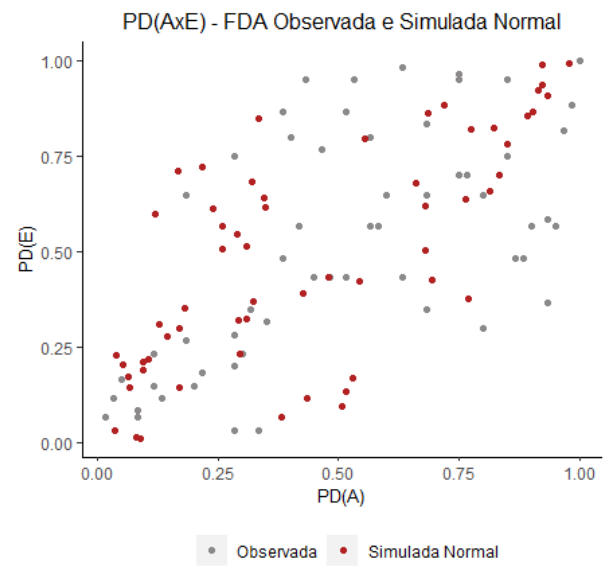
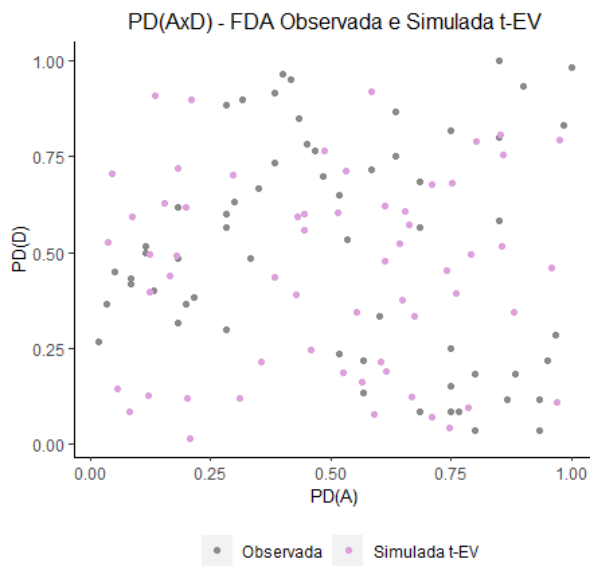
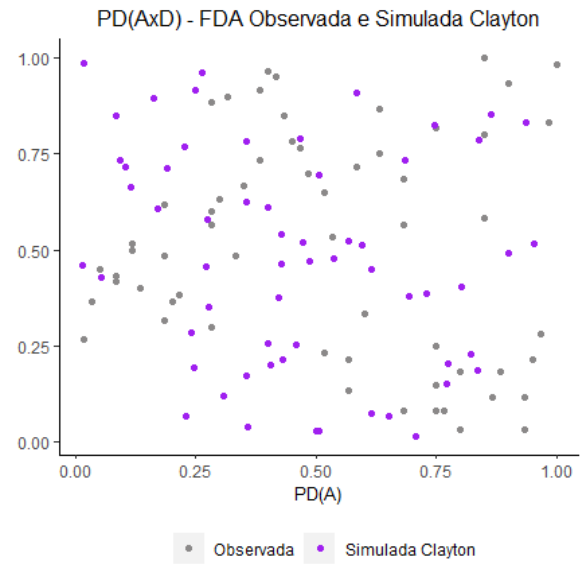
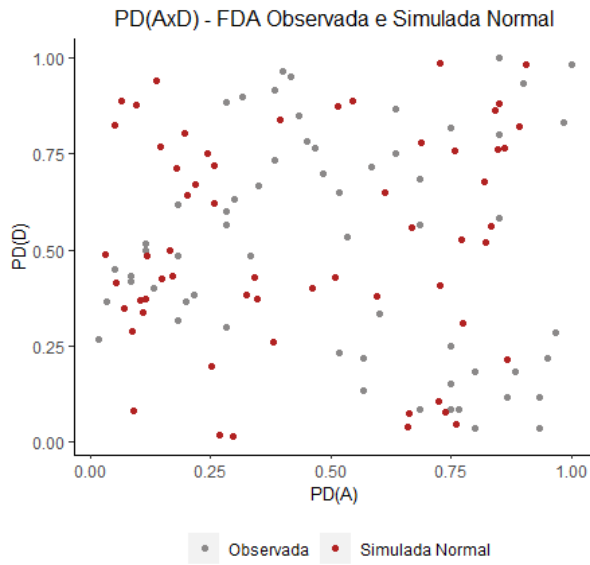
Figura 7. Distribuições de frequência e de probabilidade acumulada PD(A), PD(B)

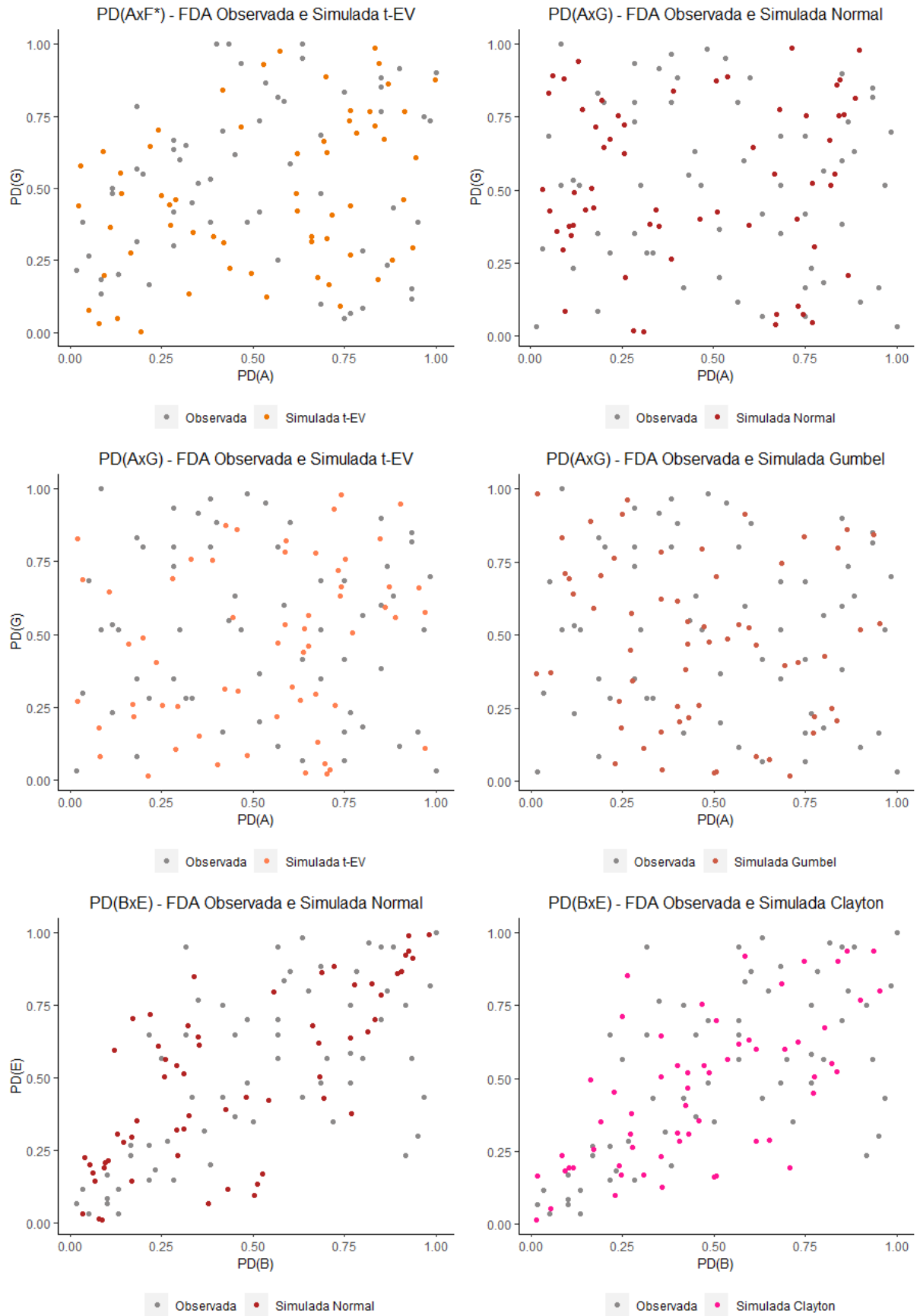


A Figura 8 permite a comparação das observações das distribuições de probabilidade acumulada conjunta para cada par de distribuição de PD (60 observações por par), com a simulação de 60 observações assumindo (i) a normalidade da distribuição conjunta, e; (ii) estrutura de dependência da(s) cópula(s) de maior aderência.

Figura 8. Função de Distribuição Acumulada – Observações Empíricas e Dados Simulados sob a Distribuição Normal e Cópuas







Esta figura apresenta as distribuições de probabilidade de *default* acumulada empírica e simuladas (via pressuposto da normalidade e cópulas de maior adesão) para cada par de distribuições de probabilidade de *default* conjunta analisado. A maior proximidade entre os pontos das observações e das simulações indicam maior poder explicativo da simulação. Fonte: elaborado pelo autor.

Os gráficos contidos na Figura 8 ilustram a incidência de probabilidades de *default* extremas conjuntas em todos os pares de linhas de crédito, sendo tal característica também capturada nas simulações. Todavia, não é possível inferir de forma visual, qual simulação é mais aderente às observações.

Para uma análise mais aprofundada, cada distribuição conjunta (assumindo Normalidade e via cópulas) foi novamente simulada, agora a um nível robusto (100.000 vezes), para então ser feita a comparação entre a raiz do erro quadrático (REQ) da distância entre a proporção em cada percentil de observações da probabilidade de *default* conjunta simulada e a proporção em cada percentil de observações empíricas. Visando verificar o efeito nas PD extremas, também foi observado o comportamento nos percentis extremos (acima de 85%). Como cada conjunto de PD contém 60 observações, os percentis extremos avaliados foram: 98,33% (59/60), 96,67% (58/60), 95,00% (57/60), 93,33% (56/60), 91,67% (55/60), 90,00% (54/60), 88,33% (53/60), 86,67% (52/60) e 85,00% (51/60). A Tabela 24 apresenta os resultados desta análise.

Tabela 24. Erro Quadrático Médio – Simulações de normalidade e via Cópulas

Painel A: AxB – Cópula Clayton							
Percentis Extremos	Dados	Normal	Clayton	REQ Normal	REQ Clayton		
98,33%	0,9500	0,9665	0,9685	0,0165	0,0185		
96,67%	0,9500	0,9333	0,9383	0,0167	0,0117		
95,00%	0,9333	0,8992	0,9104	0,0342	0,0229		
93,33%	0,8500	0,8650	0,8827	0,0150	0,0327		
91,67%	0,8500	0,8316	0,8562	0,0184	0,0062		
90,00%	0,8166	0,7993	0,8317	0,0174	0,0151		
88,33%	0,7833	0,7666	0,8079	0,0168	0,0245		
86,67%	0,7833	0,7332	0,7850	0,0502	0,0016		
85,00%	0,7500	0,7001	0,7638	0,0499	0,0138		
EQM Percentis Extremos		0,0260	0,0163				
EQM Percentis 1 a 99		0,1587	0,0244				
Painel B: AxC – Cópula Clayton							
Percentis Extremos	Dados	Normal	Clayton	REQ Normal	REQ Cópula		
98,33%	0,9500	0,9693	0,9674	0,0193	0,0174		
96,67%	0,9500	0,9401	0,9360	0,0099	0,0140		
95,00%	0,9167	0,9114	0,9049	0,0053	0,0118		
93,33%	0,8333	0,8826	0,8747	0,0493	0,0414		
91,67%	0,8333	0,8546	0,8449	0,0212	0,0115		
90,00%	0,8167	0,8277	0,8160	0,0110	0,0006		
88,33%	0,7500	0,8001	0,7888	0,0501	0,0388		
86,67%	0,7500	0,7741	0,7612	0,0241	0,0112		
85,00%	0,7167	0,7475	0,7346	0,0308	0,0179		
EQM Percentis Extremos		0,0245	0,0182				
EQM Percentis 1 a 99		0,0488	0,0475				
Painel C: AxD – Cópulas Clayton e t-EV							
Percentis Extremos	Dados	Normal	Clayton	t-EV	REQ Normal	REQ Clayton	REQ t-EV
98,33%	0,9500	0,9674	0,9668	0,9683	0,0174	0,0168	0,0183
96,67%	0,9500	0,9364	0,9354	0,9368	0,0136	0,0146	0,0132
95,00%	0,9167	0,9044	0,9038	0,9063	0,0123	0,0128	0,0104
93,33%	0,8333	0,8723	0,8718	0,8747	0,0389	0,0385	0,0413
91,67%	0,8333	0,8408	0,8400	0,8447	0,0075	0,0066	0,0114
90,00%	0,8167	0,8105	0,8105	0,8155	0,0062	0,0062	0,0012
88,33%	0,7667	0,7798	0,7808	0,7859	0,0131	0,0142	0,0192
86,67%	0,7667	0,7507	0,7519	0,7571	0,0160	0,0148	0,0096
85,00%	0,7333	0,7209	0,7228	0,7291	0,0124	0,0106	0,0042
EQM Percentis Extremos		0,0152	0,0150	0,0143			
EQM Percentis 1 a 99		0,0264	0,0263	0,0363			

Painel D: AxE – Cópula Clayton							
Percentis Extremos	Dados	Normal	Clayton	REQ Normal	REQ Cópula		
98,33%	0,9500	0,9722	0,9678	0,0222	0,0178		
96,67%	0,9500	0,9466	0,9373	0,0034	0,0127		
95,00%	0,9167	0,9212	0,9074	0,0045	0,0093		
93,33%	0,8000	0,8957	0,8787	0,0957	0,0787		
91,67%	0,8000	0,8720	0,8504	0,0720	0,0504		
90,00%	0,8000	0,8479	0,8242	0,0479	0,0242		
88,33%	0,7667	0,8238	0,7988	0,0571	0,0321		
86,67%	0,7667	0,8013	0,7743	0,0346	0,0076		
85,00%	0,7167	0,7784	0,7501	0,0617	0,0335		
EQM Percentis Extremos		0,0443	0,0295				
EQM Percentis 1 a 99		0,0768	0,0787				
Painel E: AxF* – Cópula t-EV							
Percentis Extremos	Dados	Normal	t-EV	REQ Normal	REQ t-EV		
98,33%	0,9167	0,9677	0,9696	0,0510	0,0529		
96,67%	0,9167	0,9377	0,9375	0,0211	0,0208		
95,00%	0,9000	0,9075	0,9066	0,0075	0,0066		
93,33%	0,8167	0,8761	0,8759	0,0595	0,0592		
91,67%	0,8167	0,8458	0,8461	0,0292	0,0294		
90,00%	0,8000	0,8172	0,8168	0,0172	0,0168		
88,33%	0,7667	0,7878	0,7882	0,0211	0,0215		
86,67%	0,7667	0,7595	0,7600	0,0071	0,0067		
85,00%	0,7333	0,7314	0,7328	0,0019	0,0006		
EQM Percentis Extremos		0,0239	0,0238				
EQM Percentis 1 a 99		0,0159	0,0150				
Painel F: AxG – Cópulas t-EV e Gumbel							
Percentis Extremos	Dados	Normal	t-EV	Gumbel	REQ Normal	REQ t-EV	REQ Gumbel
98,33%	0,9167	0,9672	0,9686	0,9668	0,0506	0,0519	0,0502
96,67%	0,9167	0,9362	0,9375	0,9356	0,0195	0,0209	0,0189
95,00%	0,9000	0,9039	0,9067	0,9041	0,0039	0,0067	0,0041
93,33%	0,8333	0,8716	0,8767	0,8722	0,0383	0,0433	0,0388
91,67%	0,8333	0,8400	0,8462	0,8406	0,0067	0,0129	0,0073
90,00%	0,8333	0,8095	0,8169	0,8111	0,0238	0,0165	0,0222
88,33%	0,8000	0,7785	0,7887	0,7819	0,0215	0,0113	0,0181
86,67%	0,8000	0,7493	0,7609	0,7533	0,0507	0,0391	0,0467
85,00%	0,7667	0,7193	0,7329	0,7241	0,0473	0,0338	0,0426
EQM Percentis Extremos		0,0291	0,0262	0,0276			
EQM Percentis 1 a 99		0,0222	0,0355	0,0285			
Painel G: BxE – Cópula Clayton							
Percentis Extremos	Dados	Normal	Clayton	REQ Normal	REQ Cópula		
98,33%	0,9500	0,9722	0,9680	0,0222	0,0180		
96,67%	0,9500	0,9469	0,9376	0,0031	0,0124		
95,00%	0,9167	0,9216	0,9083	0,0050	0,0084		
93,33%	0,8167	0,8963	0,8797	0,0796	0,0630		
91,67%	0,8167	0,8726	0,8520	0,0560	0,0353		
90,00%	0,7833	0,8487	0,8264	0,0653	0,0431		
88,33%	0,7667	0,8246	0,8014	0,0579	0,0347		
86,67%	0,7667	0,8023	0,7772	0,0356	0,0106		
85,00%	0,7167	0,7794	0,7537	0,0628	0,0370		
EQM Percentis Extremos		0,0430	0,0291				
EQM Percentis 1 a 99		0,0772	0,0725				

Esta tabela apresenta, para cada par de distribuição de probabilidade de *default* conjunta avaliado, o grau de concentração dos dados empíricos e simulados assumindo o pressuposto da normalidade e pela cópula de maior aderência, nos percentis extremos das distribuições de probabilidade conjunta acumulada. O erro quadrático médio é calculado pela média da raiz das diferenças entre a concentração dos dados empíricos e das simulações em todos os percentis avaliados. Por fim, é apresentado o erro quadrático médio dos percentis extremos e de toda a distribuição (percentis 1 a 99) para as simulações assumindo o pressuposto da normalidade e pela cópula de maior aderência.

Fonte: elaborado pelo autor.

A Tabela 24 resume a abordagem mais aderente aos percentis extremos (acima de 85%) e a toda a distribuição de probabilidade acumulada da probabilidade de *default* dos sete pares testados.

Tabela 25. Simulações de maior aderência aos dados observados.

EQM	AxB	AxC	AxD	AxE	AxF*	AxG	BxE
Percentis Extremos	Clayton	Clayton	<i>t</i> -EV	Clayton	<i>t</i> -EV	<i>t</i> -EV	Clayton
Percentis 1 a 99	Clayton	Clayton	Clayton	Normal	<i>t</i> -EV	Normal	Clayton

Esta tabela resume os resultados das simulações de maior aderência apresentados pela Tabela 24.

Fonte: elaborado pelo autor.

A análise da aderência das simulações aos percentis extremos indica a supremacia das cópulas em detrimento do pressuposto da Normalidade da distribuição conjunta. Dos sete pares testados, quatro foram mais aderentes à cópula de Clayton e três à cópula *t*-EV, não havendo em nenhum caso uma melhor modelagem via distribuição normal bivariada. Há um equilíbrio entre a performance da cópula de Clayton e *t*-EV, indicando a existência de eventos extremos conjuntos.

Sob o prisma da aderência das simulações entre os percentis 1 a 99, novamente o resultado sob a hipótese da Normalidade foi inferior ao das cópulas, sendo indicada para apenas dois dos sete pares. Já no conjunto das cópulas, a performance da cópula de Clayton foi superior à da cópula *t*-EV, reforçando a característica desta cópula de enfoque nas caudas inferiores das distribuições de probabilidade conjunta, havendo alinhamento com as características idiossincrática da probabilidade de *default*.

O resultado da $PD(B \times E)$ merece destaque. Conforme apresentado pela Tabela 19, os testes Shapiro Wilk e Jarque-Bera não rejeitaram a hipótese de que as distribuições marginais $PD(B)$ e $PD(E)$ têm o seu formato aproximado à uma distribuição Normal. Entretanto, a distribuição conjunta $PD(B \times E)$ é mais aderente ao formato regido pela cópula de Clayton, evidenciando que distribuições marginais Normais não implicam necessariamente em uma distribuição conjunta Normal, sendo possível que a interação entre elas resulte em outro formato de distribuição conjunta. Este resultado também foi obtido por Crook & Moreira (2011).

Em relação à assertividade dos testes de aderência das cópulas aos dados, os testes GoF_{hb} e MME divergiram em dois casos, para a $PD(A \times D)$, o teste GoF_{hb} indicou a cópula de Clayton e os testes MME indicaram a cópula *t*-EV. Já para a $PD(A \times F^*)$, os testes GoF_{hb} e MME indicaram, respectivamente, as cópulas *t*-EV e normal. Considerando as simulações para os percentis extremos e de 1 a 99, a cópula indicada pelo teste GoF_{hb} foi a mais aderente em três das quatro simulações, deixando para a cópula indicada pelos testes MME , a melhor aderência em apenas uma das quatro simulações. Embora o número de testes que comparam os dois testes seja pequeno para qualquer conclusão, o resultado é um indício da qualidade do teste GoF_{hb} , mas que não permite descartar uma análise via MME .

Em síntese, os resultados supramencionados corroboram com o argumento desta pesquisa sobre a falta de aderência da estimação do risco de portfólio do crédito rural sob o pressuposto de uma estrutura de dependência de normalidade entre as operações. Isto reforça a necessidade de criação e desenvolvimento de modelos internos de capital, além de chamar a atenção para modelos específicos de risco de crédito que levem em conta as características idiossincráticas deste segmento.

7.3. PD de Portfólio - Diretrizes do Acordo de Basileia

Embora o risco de crédito rural seja dotado de características próprias, o Acordo de Basileia não dá nenhum tratamento específico para este segmento em sua regra de composição

de capital regulatório. Sob o escopo do tratado, os empréstimos rurais são classificados como dívida corporativa ou de varejo (Bandyopadhyay, 2007; Kim, 2005).

Na abordagem por modelos internos (IRB), o cálculo da perda total de um portfólio assume que a estrutura de dependência de operações de varejo e corporativas é Normalmente distribuída (conforme Equação 4), tendo como parâmetro ρ para o caso de operações de varejo:

$$\rho_{\text{varejo}} = 0,15 \quad (39)$$

Nesta seção, as simulações de normalidade e de cópulas serão refeitas, adotando as diretrizes do Acordo de Basileia para operações de varejo. Para isto, a base de dados original foi filtrada, considerando somente as operações contratadas por pessoas físicas, resultando em 73.438.558 observações, ante as 74.964.743 da base completa. Após o filtro, a probabilidade de *default* de cada linha de crédito rural foi recalculada para cada grupo.

A linha de crédito Aquisição de Ativos Financeiros (F), recebeu o mesmo tratamento da base de dados completa (operações de pessoas físicas e jurídicas), apresentado pela Tabela 16. A Tabela 26 apresenta o resultado das simulações aplicando o parâmetro ρ_{varejo} para o caso de Normalidade.

Tabela 26. Simulações de normalidade e via Cópulas – Diretrizes de Basileia

Painel A: AxB – Cópula Clayton							
EQM	Normal	Normal Basileia	Clayton				
Percentis Extremos	0,0260	0,0147	0,0163				
Percentis 1 a 99	0,1587	0,0692	0,0244				
Painel B: AxC – Cópula Clayton							
EQM	Normal	Normal Basileia	Clayton				
Percentis Extremos	0,0245	0,0178	0,0182				
Percentis 1 a 99	0,0487	0,0312	0,0474				
Painel C: AxD – Cópulas Clayton e t-EV							
EQM	Normal	Normal Basileia	Clayton	t-EV			
Percentis Extremos	0,0152	0,0139	0,0150	0,0143			
Percentis 1 a 99	0,0264	0,0437	0,0263	0,0363			
Painel D: AxE – Cópula Clayton							
EQM	Normal	Normal Basileia	Clayton				
Percentis Extremos	0,0425	0,0231	0,0277				
Percentis 1 a 99	0,0764	0,1429	0,0783				
Painel E: AxF* – Cópula t-EV							
EQM	Normal	Normal Basileia	t-EV				
Percentis Extremos	0,0239	0,0240	0,0238				
Percentis 1 a 99	0,0144	0,0149	0,0138				
Painel F: AxG – Cópulas t-EV e Gumbel							
EQM	Normal	Normal Basileia	t-EV	Gumbel			
Percentis Extremos	0,0309	0,02935	0,02930	0,0295			
Percentis 1 a 99	0,0156	0,0311	0,0259	0,0202			
Painel G: BxE – Cópula Clayton							
EQM	Normal	Normal Basileia	Clayton				
Percentis Extremos	0,0467	0,0268	0,0328				
Percentis 1 a 99	0,0772	0,1451	0,0725				
Painel H: Simulações de maior aderência aos dados observados							
EQM	AxB	AxC	AxD	AxE	AxF*	AxG	BxE
Percentis Extremos	Normal Basileia	Normal Basileia	Normal Basileia	Normal Basileia	t-EV	t-EV	Normal Basileia
Percentis 1 a 99	Clayton	Normal Basileia	Clayton	Normal	t-EV	Normal	Clayton

Esta tabela apresenta, para cada um dos sete pares de distribuição de probabilidade de *default* conjunta avaliados, o erro quadrático médio nos percentis extremos e de toda a distribuição (1 a 99) dos dados simulados assumindo os pressupostos de (i) normalidade com parâmetro empírico; (ii) normalidade com parâmetro determinado por Basileia e; (iii) pela cópula de maior aderência segundo teste GoF_{hb} .

Fonte: elaborado pelo autor.

Diferente do que se esperava, a simulação da distribuição de probabilidade de *default* conjunta sob a premissa do Acordo de Basileia apresentou melhor aderência à distribuição dos dados nos percentis extremos em comparação às simulações por cópulas. Já para os percentis 1 a 99, a simulação por cópulas apresentou a melhor aderência na maior parte dos casos. Em resumo, os resultados continuam indicando que a alternativa de se modelar estruturas de dependência entre operações de crédito rural por cópulas estatísticas apresenta boa assertividade e deve ser considerada. Entretanto, em linha com Koziol et al. (2015), não há supremacia dos modelos de cópulas em relação aos modelos tradicionais.

Para verificar o impacto do resultado obtido, sob a hipótese de que ele pode ter sido influenciado pela linha de crédito Pronaf, presente em todas as distribuições conjuntas testadas, refizemos os testes considerando o modelo normal de Basileia e a cópula de melhor aderência. Devido ao elevado custo computacional do teste GoF_{hb} (aproximadamente 8 horas para cada distribuição bivariada testada), optou-se por reestimar os parâmetros das cópulas pelo método MME e testar a aderência das cópulas pelo teste GoF_{vc} . A Tabela 27 apresenta o resultado desta modelagem.

Tabela 27. Simulações de normalidade e via Cópulas – Todas as linhas de crédito

Painel A: Cópula de melhor aderência (triangular superior); Parâmetro estimado (triangular inferior)							
Linha	A	B	C	D	E	F	G
A	-	Clayton	Clayton	<i>t</i> -Student	Clayton	<i>t</i> -Student	Gumbel
B	2.6277	-	Normal	Joe	Clayton	Normal	Clayton
C	0.4168	0.6913	-	Frank	Clayton	Frank	Clayton
D	-0.0600	1.4477	2.0215	-	HR	Frank	Plackett
E	1.7992	1.8663	0.3909	0.9745	-	Galambos	Clayton
F	0.1974	0.4013	1.1395	6.6168	0.7738	-	Clayton
G	1.0116	-0.0956	0.0955	1.8259	0.0911	0.2341	-

Painel B: Menor EQM Percentis Extremos (triangular superior); Menor EQM Percentis 1 a 99 (triangular inferior)							
Linha	A	B	C	D	E	F	G
A	-	Normal BS	Clayton	Normal BS	Normal BS	Normal BS	Gumbel
B	Clayton	-	Normal	Normal BS	Normal BS	Normal	Clayton
C	Normal BS	Normal	-	Normal BS	Normal BS	Frank	Clayton
D	<i>t</i> -Student	Normal BS	Normal BS	-	Normal BS	Normal BS	Plackett
E	Clayton	Clayton	Normal BS	Normal BS	-	Normal BS	Clayton
F	Normal BS	Normal	Normal BS	Frank	Galambos	-	Clayton
G	Gumbel	Clayton	Clayton	Plackett	Clayton	Clayton	-

O Painel A apresenta cópula de melhor aderência e seu respectivo parâmetro para todos os subconjuntos formados pelos pares das sete linhas de crédito avaliadas. O Painel B apresenta, para cada par de probabilidade de *default* conjunta formado pelas sete linhas de crédito avaliadas, dentre as três simulações elaboradas: (i) normalidade com parâmetro empírico; (ii) normalidade com parâmetro determinado por Basileia e; (iii) pela cópula de maior aderência segundo teste GoF_{hb} , aquela com menor erro quadrático médio nos percentis extremos e em toda a distribuição (percentis 1 a 99).

Fonte: elaborado pelo autor.

A estrutura de dependência proposto pelo Acordo de Basileia apresentou melhor aderência aos percentis extremos dos dados em 11 dos 21 pares de distribuições, corroborando com o entendimento de que o modelo Basileia possui poder representativo da probabilidade de *default* conjunta em valores extremos. Para o risco entre os percentis 1 a 99, o modelo de

Basileia foi menos eficiente, sendo a escolha em 7 dos 21 pares analisados, se igualando à cópula de Clayton em termos de poder representativo.

Em síntese, os testes efetuados auxiliam na percepção contraintuitiva de boa adequação do modelo de Basileia às operações de crédito rural. Porém, os resultados sugerem que as cópulas, sobretudo a de Clayton, se consideradas, podem trazer ganhos de precisão na estimação de reais estruturas de dependência, afetando o dimensionamento de capital regulatório.

8. Considerações Finais

O crédito rural é fundamental ao desenvolvimento da agricultura, sendo um instrumento catalisador do objetivo humanitário de erradicação da fome e da miséria. Em países com vocação para o agronegócio, como o Brasil, o financiamento dos produtores rurais é estratégico e é sempre tema de políticas públicas. A melhora na qualidade da concessão de crédito rural e no gerenciamento de risco proporciona ganhos sociais e econômicos.

A mensuração do risco de crédito é basilar em todo este processo. Estudos em *agricultural economics* colocam os riscos naturais como um dos principais riscos dos produtores rurais. Por sua vez, estudos em *environmental economics* indentificaram que a incidência e a severidade de riscos naturais catastróficos são aderentes a distribuições de probabilidade com dependência de cauda.

Com base nas evidências existentes, formulou-se a hipótese de que o risco de crédito rural possa apresentar aumento de dependência em cenários extremos, como reflexo da influência de eventos severos oriundos de desastres naturais, uma vez que estes podem afetar um conjunto de produtores, reduzindo ou anulando a sua capacidade de pagamento. Ressalta-se que hipóteses associadas diretamente aos eventos geradores de situações extremas não foram testadas nesta pesquisa, embora o racional tenha servido como motivação para a abordagem de um segundo problema relacionado ao risco de crédito rural.

O segundo problema é a escassez de modelos de risco desenvolvidos especificamente ao crédito rural, cabendo a este segmento a adaptação de modelos de dívida de outros mercados, como o corporativo e o de varejo.

Diante destes três fatores: (i) relevância social e econômica (para o Brasil) do crédito rural (ii) escassez de modelos de risco de crédito elaborados especificamente para este segmento, e; (iii) da exposição compartilhada de produtores rurais (tomadores de crédito) aos riscos naturais, o objetivo deste trabalho foi verificar a melhor forma de se representar a estrutura de dependência das distribuições de probabilidade de *default* de diferentes carteiras de crédito rural. Para viabilizar o estudo, a principal instituição financeira (pelo critério de *market share*) neste segmento no Brasil gentilmente cedeu os dados. Dentre as diferentes carteiras, foi dada maior ênfase à linha de crédito Pronaf, não apenas pelo seu elevado volume de operações, mas principalmente porque colabora com o desenvolvimento da agricultura familiar.

Como método, utilizou-se uma abordagem por funções cópulas. As cópulas são funções que generalizam o conceito de dependência, para além das tradicionais medidas-resumo, fundamentais em ciências sociais aplicadas. A utilização desta classe de funções é justificada não só pela sua atuação na modelagem preditiva de desastres naturais, como também, e principalmente, pela sua vasta aplicação em modelos de risco de crédito de outros segmentos. Como argumentamos ao longo deste trabalho, até o melhor conhecimento dos autores, somos pioneiros em utilizar esta abordagem para a modelagem do risco de crédito rural.

Sob a perspectiva do risco de crédito em nível de portfólio, efetuou-se a estimação de distribuições de probabilidade de *default* conjunta de linhas de crédito rural utilizando dez famílias de cópulas. A escolha da cópula que melhor se ajustava aos dados reais foi feita via teste *Goodness-of-Fit* híbrido. Para verificarmos se a estimativa via cópulas tem bom poder explicativo, comparamos o erro quadrático médio dos quantis estimados empiricamente por meio do erro quadrático médio dos ajustes da cópula mais aderente, comparando com a cópula Normal, cuja característica principal é a simetria da distribuição conjunta. A premissa de simetria é bastante usual tanto na literatura de modelos de risco de crédito, como na abordagem

de Basileia, que indiretamente supõe distribuição Normal ao exigir que os bancos utilizem o coeficiente de correlação linear de Pearson no cálculo do capital mínimo requerido.

Os resultados indicam que, de forma geral, as distribuições geradas por meio de cópulas representam adequadamente os dados reais, em comparação à representação gerada por uma Normal bivariada. O destaque é a cópula de Clayton – capturando a elevada dependência em quantis inferiores das distribuições –, que foi a que mais se ajustou aos dados reais. De forma inesperada e contraintuitiva, a performance das cópulas foi melhor quando avaliamos a aderência considerando todo o domínio da distribuição empírica do que quando observamos somente os quantis superiores, região onde as cópulas de valores extremos teriam maior potencial de identificação de dependência.

Os experimentos sugerem não haver plena supremacia de uma abordagem (cópulas ou simetria geral ou de Basileia) em detrimento de qualquer outra. Assim, pelo menos para o portfólio de crédito rural avaliado (que é bem representativo do país, considerando o elevado número de operações e grandes volumes financeiros transacionados), é interessante que ambas as abordagens sejam de algum modo consideradas e testadas, refinando a abordagem tradicional dos modelos de risco de crédito.

Os principais agentes econômicos a se beneficiarem deste refinamento são: (i) *credores rurais*, devido a uma maior eficiência no processo de concessão de crédito; (ii) *seguradoras*, dada a característica idiossincrática do risco de crédito rural de exposição a desastres naturais serem em boa parte absorvidos por essas companhias; (iii) *reguladores* (Banco Central e Susep), pela redução potencial do risco sistêmico oriunda de modelos de risco mais precisos de seus regulados, e; (iv) ao *Setor Público*, pela melhor eficiência no processo de concessão de crédito rural por ele subsidiado, além do auxílio na promoção dos objetivos humanitários relacionados à agricultura.

Além dos resultados principais, outras contribuições merecem destaque, diante do pioneirismo deste trabalho: (i) a fundamentação da hipótese de incorporação da dependência de cauda de riscos de desastres naturais no risco de crédito rural; (ii) revisão de literatura sobre modelos de *credit scoring* aplicados ao crédito rural, abarcando modelos paramétricos e não paramétricos em trabalhos entre 1973 a 2021; (iii) aplicação do teste *Goodness-of-Fit* híbrido para a escolha de cópulas elaboradas com dados reais de crédito, e; (iv) estudo feito com microdados, o que possibilitou o cálculo das medidas de concordância das cópulas. Trabalhos com objetivos similares (Moreira, 2010, 2015) precisaram definir as medidas de concordância de forma arbitrária, devido à ausência desta informação.

Embora os resultados obtidos sugiram o questionamento da premissa da normalidade das distribuições de probabilidade de *default* conjunta de linhas de crédito rural, é possível que a presente análise (focada no *default* observado pela instituição financeira) esteja subestimando a estrutura de dependência de cauda do risco das operações de crédito rural. Diversas linhas de crédito rural, como a Pronaf que teve maior enfoque nesta pesquisa, têm como condicionante à sua liberação: (i) a vinculação do bem a ser produzido como garantia, e; (ii) a contratação de um seguro obrigatório para os casos de perda ou prejuízo desta produção a fim de não prejudicar a execução da garantia. Assim, a materialidade do risco de desastres naturais e a sua dependência de cauda entre os produtores individuais, será mais perceptível na sinistralidade observada pela seguradora do que no *default* observado pelo credor. A análise da estrutura de dependência do risco de operações de crédito rural sob a perspectiva do segurador tem um grande potencial de obter resultados ainda mais expressivos do que os encontrados nesta pesquisa, sendo um bom direcionador para trabalhos futuros.

Além disso, trabalhos futuros podem abordar outros dos elementos que aqui não pudemos acessar. Uma possibilidade é testar outras metodologias que revelem empiricamente

a estrutura de dependência causal (e.g., Redes Bayesianas). Outra possibilidade é incorporar a dimensão temporal para a estrutura de dependência (e.g., *Vine Copulas*) ou utilizar uma metodologia que incorpore nos modelos os eventos exógenos geradores que impactem a probabilidade de *default* (e.g., *Hidden Markov Models*).

Referências

- Acclassato Houensou, D., Goudjo, G. G., & Senou, M. M. (2021). Access to finance and difference in family farm productivity in Benin: Evidence from small farms. *Scientific African*, 13. <https://doi.org/10.1016/j.sciaf.2021.e00940>
- Accornero, M., Cascarino, G., Felici, R., Parlapiano, F., & Sorrentino, A. M. (2018). Credit risk in banks' exposures to non-financial firms. *European Financial Management*, 24(5), 775–791. <https://doi.org/10.1111/eufm.12138>
- Altman, E. I., & Saunders, A. (1998). Credit risk measurement: Developments over the last 20 years. *Journal of Banking and Finance*, 21(11–12), 1721–1742. [https://doi.org/10.1016/S0378-4266\(97\)00036-8](https://doi.org/10.1016/S0378-4266(97)00036-8)
- Altman, E., Saunders, A., & Bharath, S. T. (2002). Credit Ratings and the Bis Reform Agenda. *Journal of Banking & Finance* 26, 26, 909–921.
- Andrievskaya, I. K., & Penikas, H. I. (2012). Copula-application to modelling Russian banking system capital adequacy according to Basel II IRB-approach. *Model Assisted Statistics and Applications*, 7(4), 267–280. <https://doi.org/10.3233/MAS-2012-0239>
- Anjos, U. U., Ferreira, F. H., Kolev, N., & Mendes, B. V. M. (2004). *Modelando Dependências via Cópulas*. ABE.
- Antle, J. M. (2019). Data, Economics and Computational Agricultural Science. *American Journal of Agricultural Economics*, 101(2), 365–382. <https://doi.org/10.1093/ajae/aay103>
- Araujo, T. D. de, & Carvalho, J. V. de F. (2021). Independence or ... copulas! Capturing the dependence among large losses using extreme-value copulas. *21ª USP International Conference in Accounting*. www.congressosp.fipecafi.org
- Arora, N., Bohn, J. R., & Zhu, F. (2005). Reduced Form vs . Structural Models of Credit Risk : A Case Study of Three Models. *Moody's KMV*.
- Avramova, S., & le Leslé, V. (2012). Revisiting Risk-Weighted Assets. *IMF Working Papers*, 12(90), 1. <https://doi.org/10.5089/9781475502657.001>
- Babbel, D. F. (1989). Insuring Banks against Systematic Credit Risk. *The Journal of Futures Markets*, 9(6), 487–505. <https://doi.org/https://doi.org/10.1002/fut.3990090603>
- Baghdasaryan, V., Davtyan, H., Grigoryan, A., & Khachatryan, K. (2021). Comparison of econometric and deep learning approaches for credit default classification. *Strategic Change*, 30(3), 257–268. <https://doi.org/10.1002/jsc.2408>
- Bai, C., Shi, B., Liu, F., & Sarkis, J. (2019). Banking credit worthiness: Evaluating the complex relationships. *Omega (United Kingdom)*, 83, 26–38. <https://doi.org/10.1016/j.omega.2018.02.001>
- Balthazar, L. (2006). From Basel 1 to Basel 3. In *From Basel 1 to Basel 3*. Palgrave Macmillan. https://doi.org/10.1057/9780230501171_15
- Bandyopadhyay, A. (2007). Munich Personal RePEc Archive Credit Risk Models for Managing Bank ' s Agricultural Loan Portfolio. *Munich Personal RePEc Archive*.
- Barry, P. J. (2001). Modern capital management by financial institutions: Implications for agricultural lenders. *Agricultural Finance Review*, 61(2), 103–122. <https://doi.org/10.1108/00214780180001119>
- Barry, P. J., & Robison, L. J. (2001). Agricultural Finance: Credit, Credit Constraints, and Consequences. In B. L. Gardner & G. C. Rausser (Eds.), *Handbook of agricultural*

- economics: Vol. Volume IA*. Elsevier Science B.V. [https://doi.org/10.1016/S1574-0072\(06\)03049-0](https://doi.org/10.1016/S1574-0072(06)03049-0)
- Barry, P. J., Sherrick, B. J., & Ellinger, P. N. (2000). Valuation of credit risk in agricultural mortgages. *American Journal of Agricultural Economics*, 82(February), 71–81. <https://doi.org/10.1111/0002-9092.00007>
- Bauer, M., Chytilová, J., & Morduch, J. (2012). Behavioral Foundations of Microcredit: Experimental and Survey Evidence from Rural India. *American Economic Review*, 102(2), 1118–1139. <https://doi.org/10.1257/aer.102.2.1118>
- Bazrafshan, O., Zamani, H., & Shekari, M. (2020). A copula-based index for drought analysis in arid and semi-arid regions of Iran. *Natural Resource Modeling*, 33(1), 1–19. <https://doi.org/10.1111/nrm.12237>
- BCBS. (1975). Report to the Governor's on the Supervision of bank's Foreign Establishments. In *Basel, September* (p. 5).
- BCBS. (1988). *International Convergence of Capital Measurement and Capital Standards*. July, 28.
- BCBS. (1995). *An Internal Model-Based Approach to Market Risk Capital Requirements*. April, 19.
- BCBS. (1996). *Amendment to the Capital Accord to Incorporate Basle Committee on Banking Supervision*. January. <http://www.bis.org/publ/bcbs24.pdf>
- BCBS. (2000). *Principles for the Management of Credit Risk*.
- BCBS. (2011). Basel III: A global regulatory framework for more resilient banks and banking systems. In *Basel Committee on Banking Supervision Basel* (Issue June).
- BCBS. (2015). Basel Committee on Banking Supervision: A brief history of the Basel Committee. In *Bank for International Settlement* (Issue October).
- BCBS. (2016). *Regulatory consistency assessment programme (RCAP) – Analysis of risk-weighted assets for credit risk in the banking book* (Issue April). <https://doi.org/10.1080/09291016.2016.1197475>
- BCBS. (2019). *The BIS - Promoting global monetary and financial stability through international cooperation*.
- BCBS. (2021a). *Basel Committee on Banking Supervision Basel III Monitoring Report - September 2021*. <https://www.bis.org/bcbs/publ/d524.htm>
- BCBS. (2021b). *Progress report on adoption of the Basel regulatory framework*. www.bis.org
- Bégin, J. F., Boudreault, M., Doljanu, D. A., & Gauthier, G. (2019). Credit and Systemic Risks in the Financial Services Sector: Evidence From the 2008 Global Crisis. *Journal of Risk and Insurance*, 86(2), 263–296. <https://doi.org/10.1111/jori.12210>
- Beltratti, A., & Paladino, G. (2016). Basel II and regulatory arbitrage. Evidence from financial crises. *Journal of Empirical Finance*, 39, 180–196. <https://doi.org/10.1016/j.jempfin.2016.02.006>
- Bennouna, G., & Tkiouat, M. (2019). Scoring in microfinance: Credit risk management tool - Case of Morocco. *Procedia Computer Science*, 148(ICDS 2018), 522–531. <https://doi.org/10.1016/j.procs.2019.01.025>

- Berg, T., & Koziol, P. (2017). An analysis of the consistency of banks' internal ratings. *Journal of Banking and Finance*, 78, 27–41. <https://doi.org/10.1016/j.jbankfin.2017.01.013>
- Bessis, J. (2015). *Risk Management in Banking* (Fourth Edi). Wiley Finance.
- Black, F., & Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. *The Journal of Political Economy*, 81(3), 637–654.
- Bliss, R. (2002). Comments on “Credit ratings and the BIS capital adequacy reform agenda.” *Journal of Banking and Finance*, 26. [https://doi.org/10.1016/S0378-4266\(01\)00269-2](https://doi.org/10.1016/S0378-4266(01)00269-2)
- Bluhm, C., Overbeck, L., & Wagner, C. (2010). *Introduction to Credit Risk Modeling* (Second). Taylor & Francis Group.
- Bohn, J. R., & Stein, R. M. (2009). *Active Credit Portfolio Management in Practice*. John Wiley & Sons, Inc.
- Borgo, M. D. (2020). Internal models for deposits : Effects on banks ' capital and interest rate risk of assets. *Journal of Banking and Finance*. <https://doi.org/10.1016/j.jbankfin.2020.105940>
- Brasil. (2021). *Plano Safra 2021/2022*. <https://www.gov.br/agricultura/pt-br/assuntos/politica-agricola/todas-publicacoes-de-politica-agricola/plano-agricola-pecuario/plano-safra-2021-2022/@@download/file/Plano%20Agri%CC%81cola%20e%20Pecua%CC%81rio%2021%20-%202022.pdf>
- Breymann, W., Dias, A., & Embrechts, P. (2003). Dependence structures for multivariate high-frequency data in finance. *Quantitative Finance*, 3(1), 1–14. <https://doi.org/10.1080/713666155>
- Brezigar-Masten, A., Masten, I., & Volk, M. (2021). Modelin-g credit risk with a Tobit model of days past due. *Journal of Banking and Finance*, 122. <https://doi.org/10.1016/j.jbankfin.2020.105984>
- Bülbül, D., Hakenes, H., & Lambert, C. (2019). What influences banks' choice of credit risk management practices? Theory and evidence. *Journal of Financial Stability*, 40, 1–14. <https://doi.org/10.1016/j.jfs.2018.11.002>
- Calabrese, R., Osmetti, S. A., & Zanin, L. (2019). A joint scoring model for peer-to-peer and traditional lending: a bivariate model with copula dependence. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 182(4), 1163–1188. <https://doi.org/10.1111/rssa.12523>
- Campbell, D. A. (2017). An Update on the United Nations Millennium Development Goals. *JOGNN - Journal of Obstetric, Gynecologic, and Neonatal Nursing*, 46(3), e48–e55. <https://doi.org/10.1016/j.jogn.2016.11.010>
- Carvalho, J. V. de F. (2011). *Redes Bayesianas: um método para avaliação de interdependência e contágio em séries temporais multivariadas*.
- Carvalho, J. V. de F., & Chiann, C. (2013). Redes Bayesianas: Um Método para Avaliação de Interdependência e Contágio em Séries Temporais Multivariadas. *Revista Brasileira de Economia*. <http://go.worldbank.org/>
- Cervantes-Godoy, D., & Dewbre, J. (2010). Economic Importance of Agriculture for Poverty Reduction. *OECD Food, Agriculture and Fisheries Papers*, 23. <https://doi.org/http://dx.doi.org/10.1787/5kmmv9s20944-en>

- Chandio, A. A., Jiang, Y., Gessesse, A. T., & Dunya, R. (2019). The Nexus of Agricultural Credit, Farm Size and Technical Efficiency in Sindh, Pakistan: A Stochastic Production Frontier Approach. *Journal of the Saudi Society of Agricultural Sciences*, 18(3), 348–354. <https://doi.org/10.1016/j.jssas.2017.11.001>
- Chaudhary, M. A., & Ishfaq, M. (2003). Credit worthiness of rural borrowers of Pakistan. *Journal of Socio-Economics*, 32(6), 675–684. <https://doi.org/10.1016/j.socec.2003.10.005>
- Chavira, D. A. G., Lopez, J. C. L., Noriega, J. J. S., Valenzuela, O. A., & Carrillo, P. A. A. (2017). A credit ranking model for a parafinancial company based on the ELECTRE-III method and a multiobjective evolutionary algorithm. *Applied Soft Computing Journal*, 60, 190–201. <https://doi.org/10.1016/j.asoc.2017.06.021>
- Cheng, L., Hoerling, M., Liu, Z., & Eischeid, J. (2019). Physical understanding of human-induced changes in U.S. hot droughts using equilibrium climate simulations. *Journal of Climate*, 32(14), 4431–4443. <https://doi.org/10.1175/JCLI-D-18-0611.1>
- Cherubini, U., Luciano, E., & Vecchiato, W. (2004). *Copula Methods in Finance*.
- Chhikara, R. K. (1989). The State of the Art in Credit Evaluation. *American Journal of Agricultural Economics*, 71(5), 1138–1144. <https://doi.org/10.2307/1243096>
- Chorafas, D. N. (2004). Economic Capital Allocation with Basel II: Cost, Benefit and Implementation Procedures. In *Elsevier Butterworth-Heinemann*. <https://doi.org/10.1016/B978-0-7506-6182-9.X5000-9>
- Clemente, A. di. (2020). Modeling Portfolio Credit Risk Taking into Account the Default Correlations Using a Copula Approach: Implementation to an Italian Loan Portfolio. *Journal of Risk and Financial Management*, 13(6), 129. <https://doi.org/10.3390/jrfm13060129>
- Clemente, A. di, & Romano, C. (2004). *Measuring and Optimizing Portfolio Credit Risk: A Copula-based Approach*.
- Condori-Alejo, H. I., Aceituno-Rojo, M. R., & Alzamora, G. S. (2021). Rural Micro Credit Assessment using Machine Learning in a Peruvian microfinance institution. *Procedia Computer Science*, 187, 408–413. <https://doi.org/10.1016/j.procs.2021.04.117>
- Conte, M. N., & Kelly, D. L. (2018). An imperfect storm: Fat-tailed tropical cyclone damages, insurance, and climate policy. *Journal of Environmental Economics and Management*, 92, 677–706. <https://doi.org/10.1016/j.jeem.2017.08.010>
- Conte, M. N., & Kelly, D. L. (2021). Understanding the Improbable: A Survey of Fat Tails in Environmental Economics. *Annual Review of Resource Economics*, 13(1), 289–310. <https://doi.org/10.1146/annurev-resource-102020-094143>
- Cooper, M., Technow, F., Messina, C., Gho, C., & Radu Totir, L. (2016). Use of crop growth models with whole-genome prediction: Application to a maize multienvironment trial. *Crop Science*, 56(5), 2141–2156. <https://doi.org/10.2135/cropsci2015.08.0512>
- Cossette, H., Marceau, E., Mtalai, I., & Veilleux, D. (2018). Dependent risk models with Archimedean copulas: A computational strategy based on common mixtures and applications. *Insurance: Mathematics and Economics*, 78, 53–71. <https://doi.org/10.1016/j.insmatheco.2017.11.002>
- Crook, J., & Moreira, F. (2011). Checking for asymmetric default dependence in a credit card portfolio: A copula approach. *Journal of Empirical Finance*, 18(4), 728–742. <https://doi.org/10.1016/j.jempfin.2011.05.005>

- Crouhy, M., Galai, D., & Mark, R. (2014). The Essentials of Risk Management. In M. G.-H. Education (Ed.), *Journal of Chemical Information and Modeling* (Second, Vol. 53, Issue 9).
- Darrell, D., & Schaefer, S. (2003). *Credit Risk : Pricing , Measurement , and Management*. Princeton University Press.
- Daul, S., Giorgi, E. de, Lindskog, F., & Mcneil, Z. A. (2003). The Grouped t-copula with an Application to Credit Risk Stéphane Daul Quantitative Risk Management Methods Swiss Re. *Risk*.
- de Castro, E. R., & Teixeira, E. C. (2012). Rural credit and agricultural supply in Brazil. *Agricultural Economics*, 43(3), 293–302. <https://doi.org/10.1111/j.1574-0862.2012.00583.x>
- Dillon, W. R. (1979). The Performance of the Linear Discriminant Function in Nonoptimal Situations and the Estimation of Classification Error Rates: A Review of Recent Findings. *Source: Journal of Marketing Research*, 16(3), 370–381. <https://doi.org/https://doi.org/10.2307/3150711>
- Dinh, T. H. T., & Kleimeier, S. (2007). A credit scoring model for Vietnam's retail banking market. *International Review of Financial Analysis*, 16(5), 471–495. <https://doi.org/10.1016/j.irfa.2007.06.001>
- Durguner, S., Barry, P. J., Katchova, A. L., & Drive, W. G. (2006). Credit Scoring Models : A Comparison between Crop and Livestock Farms. *American Agricultural Economics Association Meeting, May*, 1–31.
- Durrleman, V., Nikeghbali, A., & Roncalli, T. (2000). *Which copula is the right one?* <http://ssrn.com/abstract=1032545>
- EBA. (2013). *Interim results of the EBA review of the consistency of risk-weighted assets of the banking book* (Issue February).
- Eisenbeis, R. A. (1978). Problems In Applying Discriminant Analysis In Credit Scoring Models. *Journal of Banking and Finance*, 2, 205–219. [https://doi.org/https://doi.org/10.1016/0378-4266\(78\)90012-2](https://doi.org/https://doi.org/10.1016/0378-4266(78)90012-2)
- Elizalde, A., & Repullo, R. (2007). Economic and Regulatory Capital in Banking: What Is the Difference. *International Journal of Central Banking*, 3(3), 87–117.
- Embrechts, P., McNeil, A. J., & Straumann, D. (2002). Correlation and Dependence in Risk Management: Properties and Pitfalls. In M. A. H. Dempster (Ed.), *Risk Management: Value At Risk And Beyond* (pp. 176–223). Cambridge University Press.
- Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R., & Beleraj, A. (2014). An academic response to Basel 3.5. *Risks*, 2(1), 25–48. <https://doi.org/10.3390/risks2010025>
- Erhardt, T. M., Czado, C., & Schepsmeier, U. (2015). R-vine models for spatial time series with an application to daily mean temperature. *Biometrics*, 71(2), 323–332. <https://doi.org/10.1111/biom.12279>
- Fang, K. W., Kotz, S., & Ng, K. W. (1990). Symmetric Multivariate and Related Distributions. In *Statistical Methods J.S. Maritz* (CRC Press, Issue 2).
- Featherstone, A. M., Roessler, L. M., & Barry, P. J. (2006). Determining the probability of default and risk-rating class for loans in the Seventh Farm Credit District Portfolio. *Review of Agricultural Economics*, 28(1), 4–23. <https://doi.org/10.1111/j.1467-9353.2006.00270.x>

- Federico, G. (2005). *Feeding the World: An Economic History of Agriculture, 1800 - 2000* (J. Mokyr, Ed.). Princeton University Press.
- Fenech, J. P., Vosgha, H., & Shafik, S. (2015). Loan default correlation using an Archimedean copula approach: A case for recalibration. *Economic Modelling*, 47, 340–354. <https://doi.org/10.1016/j.econmod.2015.03.001>
- Figueira, J., Greco, S., & Ehrogott, M. (2005). *Multiple Criteria Decision Analysis: State of the Art Surveys* (J. Figueira, S. Greco, & M. Ehrogott, Eds.; 1st ed., Vol. 78). Springer New York. <https://doi.org/10.1007/b100605>
- Figueira, S. R. F., & Rolim, G. de S. (2020). Economic and agrometeorological modeling of sugarcane productivity in São Paulo state, Brazil. *Agronomy Journal*, 112(6), 4836–4848. <https://doi.org/10.1002/agj2.20367>
- Fisher, N. I. (2006). *Encyclopedia Of Statistical Sciences: Vol. Volume 2* (S. Kotz, N. Balakrishnan, C. B. Read, B. Vidakovic, & N. L. Johnson, Eds.; Second Edition, pp. 1363–1367). John Wiley & Sons, Inc.
- Fisher, N. I., & Sen, P. K. (1994). *The Collected Works of Wassily Hoeffding*. Springer. <https://link.springer.com/book/10.1007/978-1-4612-0865-5>
- Fraisse, H., & Laporte, M. (2022). Return on investment on artificial intelligence: The case of bank capital requirement. *Journal of Banking & Finance*, 106401. <https://doi.org/10.1016/j.jbankfin.2022.106401>
- Fréchet, M. (1951). Sur les tableaux dont les marges et des bornes sont données. In *Review of the International Statistical Institute* (Vol. 28, Issue 1).
- Frey, R., & McNeil, A. J. (2003). *Dependent Defaults in Models of Portfolio Credit Risk* (Vol. 11).
- Genest, C., Nešlehová, J., & Ghorbal, N. ben. (2011). Estimators Based On Kendall's Tau In Multivariate Copula Models. *Australian and New Zealand Journal of Statistics*, 53(2), 157–177. <https://doi.org/10.1111/j.1467-842X.2011.00622.x>
- Genest, C., Quessy, J. F., & Rémillard, B. (2006). Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scandinavian Journal of Statistics*, 33(2), 337–366. <https://doi.org/10.1111/j.1467-9469.2006.00470.x>
- Genest, C., Rémillard, B., & Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44(2), 199–213. <https://doi.org/10.1016/j.insmatheco.2007.10.005>
- Geske, R. (1977). The Valuation of Corporate Liabilities as Compound Options. *The Journal of Financial and Quantitative Analysis*, 12(4), 541–552.
- Glasserman, P., Kang, W., & Shahabuddin, P. (2007). Large Deviations in Multifactor Portfolio Credit Risk. *Mathematical Finance*, 17(3), 345–379.
- Gonzales, R. B. (2012). *Uma avaliação do capital regulatório no Sistema Bancário*.
- Goodwin, B. K. (2001). Problems with Market Insurance in Agriculture. *American Journal of Agricultural Economics*, 83(3), 643–649.
- Grass, G. (2012). Does conglomeration really reduce credit risk? *Accounting and Finance*, 52(3), 831–848. <https://doi.org/10.1111/j.1467-629X.2012.00501.x>
- Gudendorf, G., & Segers, J. (2010). Extreme-Value Copulas. In *Copula Theory and Its Applications* (pp. 127–145). https://doi.org/10.1007/978-3-642-12465-5_6

- Guirkinger, C., & Boucher, S. R. (2008). Credit constraints and productivity in Peruvian agriculture. *Agricultural Economics*, 39(3), 295–308. <https://doi.org/10.1111/j.1574-0862.2008.00334.x>
- Gupta, A. K., Varga, T., & Bodnar Taras. (2013). *Elliptically Contoured Models in Statistics and Portfolio Theory Second Edition* (2^o). Springer.
- Haldane, A. G. (2011). Capital Discipline Based on a Speech Given at the American Economic Association. In *Bank of International Settlements* (Issue January). <http://www.bis.org/review/r110325a.pdf>
- Hamilton, D., James, J., & Webber, N. (2001). *Copula Methods and the Analysis of Credit Risk*. <http://ssrn.com/abstract=1014407>
- Hardy, W. E., & Adrian, J. L. (1985). A linear programming alternative to discriminant analysis in credit scoring. *Agribusiness*, 1(4), 285–292. [https://doi.org/10.1002/1520-6297\(198524\)1:4<285::AID-AGR2720010406>3.0.CO;2-M](https://doi.org/10.1002/1520-6297(198524)1:4<285::AID-AGR2720010406>3.0.CO;2-M)
- Hayden, E., & Porath, D. (2006). Statistical Methods to Develop Rating Models. In B. Engemann & R. Rauhmeier (Eds.), *The Basel II Risk Parameters – Estimation, Validation, and Stress Testing* (p. 376). Springer Verlag.
- Hering, C., & Hofert, M. (2015). Goodness-of-fit tests for archimedean copulas in high dimensions. *Springer Proceedings in Mathematics and Statistics*, 99, 357–373. https://doi.org/10.1007/978-3-319-09114-3_21
- Hering, I., & Musshoff, O. (2017). Progressive Lending in Microfinance—What about the Farmers? *Review of Development Economics*, 21(3), 803–828. <https://doi.org/10.1111/rode.12273>
- Hoeffding, W. (1940). Scale-invariant correlation theory. *Schriften Des Mathematischen Instituts Und Des Instituts Für Angewandte Mathematik Der Universitiit Berlin*, 181–233.
- Hoeffding, W. (1941). Scale-invariant correlations for discontinuous distributions. *Archiv Fur Mathematische Wirtschafts - Und Sozialforschung*, 49–70.
- Hoeffding, W. (1942). Stochastic dependence and functional relationships. *Skandinavisk Aktuarietidskrift*, 200–227.
- Hohl, R. M. (2019). *Agricultural Risk Transfer: From Insurance to Reinsurance to Capital Markets*. Wiley Finance.
- Hu, Y., Liu, C., & Peng, J. (2021). Financial inclusion and agricultural total factor productivity growth in China. *Economic Modelling*, 96, 68–82. <https://doi.org/10.1016/j.econmod.2020.12.021>
- Huang, W., & Prokhorov, A. (2014). A Goodness-of-fit Test for Copulas. *Econometric Reviews*, 33(7), 751–771. <https://doi.org/10.1080/07474938.2012.690692>
- Hull, J. (1989). Assessing Credit Risk in a Financial Institution's Off-Balance Sheet Commitments. *The Journal of Financial and Quantitative Analysis*, 24(4), 489–501.
- Hünseler, M. (2013). *Credit Portfolio Management* (1st ed.). Palgrave Macmillan. <https://doi.org/10.1057/9780230391505>
- Ikeo, K. (1992). BIS Capital Requirements As An Incomplete Form Of Risk-Based Capital Regulations. *Kyoto University Economic Review*, 62(1 (132)), 37–49.

- Ipea. (2018). *ODS - Metas Nacionais dos Objetivos de Desenvolvimento Sustentável*. https://www.ipea.gov.br/portal/images/stories/PDFs/livros/livros/180801_ods_metas_na_c_dos_obj_de_desenv_susten_propos_de_adequa.pdf
- Jacobson, T., & Roszbach, K. (2003). Bank lending policy, credit scoring and value-at-risk. *Journal of Banking and Finance*, 27(4), 615–633. [https://doi.org/10.1016/S0378-4266\(01\)00254-0](https://doi.org/10.1016/S0378-4266(01)00254-0)
- Jarque, C. M., & Bera, A. K. (1987). A Test for Normality of Observations and Regression Residuals. *International Statistical Review*, 55(2), 163–172. <https://doi.org/https://doi.org/1403192>
- Jarrow, R. A., & Turnbull, S. M. (1995). Pricing Derivatives on Financial Securities Subject to Credit Risk. *The Journal of Finance*.
- Jeanblanc, M., & Rutkowski, M. (1999). Modelling of Default Risk: An Overview. *Mathematical Finance: Theory and Practice*.
- Jimi, N. A., Nikolov, P. v., Malek, M. A., & Kumbhakar, S. (2019). The effects of access to credit on productivity: separating technological changes from changes in technical efficiency. *Journal of Productivity Analysis*, 52(1–3), 37–55. <https://doi.org/10.1007/s11123-019-00555-8>
- Johnson, R. B., & Hagan, A. R. (1973). Agricultural loan evaluation with discriminant analysis. *Southern Journal of Agricultural Economics*, 05(02), 57–62.
- Jong, E. de, & Vijge, M. J. (2021). From Millennium to Sustainable Development Goals: Evolving discourses and their reflection in policy coherence for development. *Earth System Governance*, 7, 100087. <https://doi.org/10.1016/j.esg.2020.100087>
- Jorion, P. (2007). *Value at Risk - The New Benchmark for Managing Financial Risk* (Third). McGraw-Hill.
- Just, D. R., Khantachavana, S. v., & Just, R. E. (2010). Empirical Challenges for Risk Preferences and Production. *Annual Review of Resource Economics*, 2, 13–31. <https://doi.org/10.1146/annurev.resource.012809.103902>
- Kassouri, Y., & Kacou, K. Y. T. (2022). Does the structure of credit markets affect agricultural development in West African countries? *Economic Analysis and Policy*, 73, 588–601. <https://doi.org/10.1016/j.eap.2021.12.015>
- Katchova, A. L., & Barry, P. J. (2005). Credit Risk Models and Agricultural Lending. *American Journal of Agricultural Economics*, 87(1), 194–205. <https://doi.org/http://www.jstor.org/stable/3698001>
- Khoju, M. R., & Barry, P. J. (1993). *Business Performance Based Credit Scoring Models: A New Approach to Credit Evaluation*.
- Kim, J. (2005). *A Credit Risk Model for Agricultural Loan Portfolios Under the New Basel Capital Accord* (Issue May). Texas A&M University.
- Klieštk, T., & Cúg, J. (2015). Comparison of Selected Models of Credit Risk. *Procedia Economics and Finance*, 23(2015), 356–361. [https://doi.org/10.1016/s2212-5671\(15\)00452-9](https://doi.org/10.1016/s2212-5671(15)00452-9)
- Kole, E., Koedijk, K., & Verbeek, M. (2007). Selecting copulas for risk management. *Journal of Banking and Finance*, 31(8), 2405–2423. <https://doi.org/10.1016/j.jbankfin.2006.09.010>

- Kousky, C., & Cooke, R. M. (2009). The Unholy Trinity: Fat Tails, Tail Dependence, and Micro-Correlations. *Resources for the Future*. <https://doi.org/10.1108/eb002035>
- Koziol, P., Schell, C., Eckhardt, M., & Zentrale ÖB 200-3 Katalogisierung. (2015). *Credit risk stress testing and copulas - is the Gaussian copula better than its reputation?*
- Kumar, C. S., Turvey, C. G., & Kropp, J. D. (2013). The Impact of Credit Constraints on Farm Households: Survey Results from India and China. *Applied Economic Perspectives and Policy*, 35(3), 508–527. <https://doi.org/10.1093/aep/ppt002>
- Laurent, J.-P. (2008). Credit Risk Models. In E. L. Melnick & B. S. Everitt (Eds.), *Encyclopedia of Quantitative Risk Analysis and Assessment*. John Wiley & Sons. <https://doi.org/https://doi.org/10.1002/9780470061596.risk0343>
- Lewis, M. K. (1988). Off-balance sheet activities and financial innovation in banking. In *Psl Quarterly Review* (Vol. 41, Issue 167).
- Li, L., Khan, S. U., Xia, X., Zhang, H., & Guo, C. (2020). Screening of agricultural land productivity and returning farmland to forest area for sensitivity to rural labor outward migration in the ecologically fragile Loess Plateau region. *Environmental Science and Pollution Research*. <https://doi.org/10.1007/s11356-020-09022-6>/Published
- Linh, T. N., Long, H. T., Chi, L. van, Tam, L. T., & Lebailly, P. (2019). Access to rural credit markets in developing countries, the case of Vietnam: A literature review. *Sustainability*, 11(5). <https://doi.org/10.3390/su11051468>
- Lopez, J. A., & Saidenberg, M. R. (2000). Evaluating credit risk models. *Journal of Banking & Finance*, 24. https://doi.org/10.1007/3-540-26993-2_11
- Lu, M. J., Chen, C. Y. H., & Härdle, W. K. (2017). Copula-based factor model for credit risk analysis. *Review of Quantitative Finance and Accounting*, 49(4), 949–971. <https://doi.org/10.1007/s11156-016-0613-x>
- Lufburrow, J., Barry, P. J., & Dixon, B. L. (1984). Credit Scoring for Farm Loan Pricing. *Agricultural Finance Review*, 44.
- Luo, S., Kong, X., & Nie, T. (2016). Spline based survival model for credit risk modeling. *European Journal of Operational Research*, 253(3), 869–879. <https://doi.org/10.1016/j.ejor.2016.02.050>
- Malamud, B. D., Millington, J. D. A., & Perry, G. L. W. (2005). Characterizing wildfire regimes in the United States. *Proceedings of the National Academy of Sciences of the United States of America*, 102(13), 4694–4699. <https://doi.org/10.1073/pnas.0500880102>
- Malamud, B. D., & Turcotte, D. L. (2006). The applicability of power-law frequency statistics to floods. *Journal of Hydrology*, 322(1–4), 168–180. <https://doi.org/10.1016/j.jhydrol.2005.02.032>
- Mandala, I. G. N., Nawangpalupi, C. B., & Praktikto, F. R. (2012). Assessing Credit Risk: An Application of Data Mining in a Rural Bank. *Procedia Economics and Finance*, 4, 406–412. [https://doi.org/10.1016/s2212-5671\(12\)00355-3](https://doi.org/10.1016/s2212-5671(12)00355-3)
- Manoharan, N., & Varkey, R. S. (2021). Agricultural credit and agricultural productivity across Indian states: An analysis. *Journal of Public Affairs*. <https://doi.org/10.1002/pa.2597>
- Mariathanan, M., & Merrouche, O. (2014). The manipulation of basel risk-weights. *Journal of Financial Intermediation*, 23(3), 300–321. <https://doi.org/10.1016/j.jfi.2014.04.004>

- McNamara, C., Piontek, T., & Metrick, A. (2015). Basel III A: Regulatory History. *Yale Program On Financial Stability*. <https://doi.org/10.2139/ssrn.2576906>
- McNeil, A., Frey, R., & Embrechts, P. (2005). *Quantitative Risk Management*. Princeton University Press.
- Medina-Olivares, V., Calabrese, R., Dong, Y., & Shi, B. (2021). Spatial dependence in microfinance credit default. *International Journal of Forecasting*. <https://doi.org/10.1016/j.ijforecast.2021.05.009>
- Mendes-Da-Silva, W., Lucas, E. C., & Carvalho, J. V. de F. (2021). Flood insurance: The propensity and attitudes of informed people with disabilities towards risk. *Journal of Environmental Management*, 294. <https://doi.org/10.1016/j.jenvman.2021.113032>
- Merton, R. C. (1973). Theory of rational option pricing. *The Bell Journal of Economics and Management Science*, 4(1), 141–183. <http://www.jstor.org/stable/3003143><http://www.jstor.org/page/info/about/policies/terms.jsp><http://www.jstor.org>
- Mesbahzadeh, T., Mirakbari, M., Saravi, M. M., Sardoo, F. S., & Miglietta, M. M. (2020). Meteorological drought analysis using copula theory and drought indicators under climate change scenarios (RCP). *Meteorological Applications*, 27(1), 1–20. <https://doi.org/10.1002/met.1856>
- Mester, L. J. (1997). What's the Point of Credit Scoring? *Business Review*, 3(February 1997), 3–16.
- Miller, L. H., & LaDue, E. L. (1988). *Credit Assessment Models for Farm Borrowers: A Logit Analysis*. <https://doi.org/10.22004/ag.econ.178692>
- Moreira, F. F. (2010). Copula-Based formulas to estimate unexpected credit losses (The future of basel accords?). *Financial Markets, Institutions and Instruments*, 19(5), 381–404. <https://doi.org/10.1111/j.1468-0416.2010.00162.x>
- Moreira, F. F. (2015). Estimating Portfolio Credit Losses in Downturns. *Financial Markets, Institutions & Instruments*, 24(5), 391–414. <https://doi.org/10.1111/fmii.12033>
- Nelsen, R. B. (2006). *An Introduction to Copulas* (Second Edition). Springer Science+Business Media, Inc.
- Novak, M. P., & LaDue, E. (1999). Application of Recursive Partitioning to Agricultural Credit Scoring. *Journal of Agricultural and Applied Economics*, 31(1), 109–122. <https://doi.org/10.1017/s0081305200028818>
- OECD, & FAO. (2021). *OECD-FAO Agricultural Outlook 2021-2030*. OECD Publisher. <https://doi.org/10.1787/19428846-en>
- Okhrin, O., Trimborn, S., & Waltz, M. (2021). gofCopula: Goodness-of-Fit Tests for Copulae. *The R Journal*, 13, 467–498. <https://doi.org/https://doi.org/10.32614/RJ-2021-060>
- Oliveira, L. H. A. de, & Carvalho, J. V. de F. (2020). *We are Living on the Edge! Gerenciando Sinistros de Extrema Severidade com o Apoio da Teoria de Valores Extremos*. www.congressosp.fipecafi.org
- Oyinbo, O., Chamberlin, J., Abdoulaye, T., & Maertens, M. (2021). Digital Extension, Price Risk, and Farm Performance: Experimental Evidence from Nigeria. *American Journal of Agricultural Economics*, 00(00), 1–20. <https://doi.org/10.1111/ajae.12242>

- Penikas, H. (2020). The review of the open challenges in the IRB loan portfolio credit risk modeling. *Model Assisted Statistics and Applications*, 15(4), 371–388. <https://doi.org/10.3233/MAS-200508>
- Penson, J. B., Capps, O., Rosson, C. P., & Woodward, R. T. (2017). *Introduction to Agricultural Economics* (Seventh ed). Pearson.
- Perotti, E., Lev, R., & Vlahu Razvan. (2011). Capital Regulation and Tail Risk. *International Journal of Central Banking*, 7(4), 165–170.
- Pertaia, G., Prokhorov, A., & Uryasev, S. (2021). A new approach to credit ratings. *Journal of Banking and Finance*. <https://doi.org/10.1016/j.jbankfin.2021.106097>
- Rao, C., Liu, M., Goh, M., & Wen, J. (2020). 2-stage modified random forest model for credit risk assessment of P2P network lending to “Three Rurals” borrowers. *Applied Soft Computing Journal*, 95, 106570. <https://doi.org/10.1016/j.asoc.2020.106570>
- Rosen, D., & Saunders, D. (2009). Analytical methods for hedging systematic credit risk with linear factor portfolios. *Journal of Economic Dynamics and Control*, 33(1), 37–52. <https://doi.org/10.1016/j.jedc.2008.03.010>
- Ryser, M., & Denzler, S. (2009). Selecting credit rating models: A cross-validation-based comparison of discriminatory power. *Financial Markets and Portfolio Management*, 23(2), 187–203. <https://doi.org/10.1007/s11408-009-0101-6>
- Scaillet, O. (2007). Kernel-based goodness-of-fit tests for copulas with fixed smoothing parameters. *Journal of Multivariate Analysis*, 98(3), 533–543. <https://doi.org/10.1016/j.jmva.2006.05.006>
- Schäfer, R., & Koivusalo, A. F. R. (2013). Dependence of defaults and recoveries in structural credit risk models. *Economic Modelling*, 30(1), 1–9. <https://doi.org/10.1016/j.econmod.2012.08.033>
- Schönbucher, P. J., & Schubert, D. (2001). *Copula-Dependent Default Risk in Intensity Models*. <http://www.finasto.uni-bonn.de/schonbucher/JELClassification:G13http://www.finasto.uni-bonn.de/~schonbuc/>
- Schweizer, B. (1991). Thirty Years of Copulas. In G. Dall’Aglia, S. Kotz, & G. Salinetti (Eds.), *Advances in Probability Distributions with Given Marginals* (pp. 13–50). Kluwer Academic Publishers.
- Shapiro, S. S., & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52(3), 591–611. <https://doi.org/https://doi.org/2333709>
- Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. . *Publications de l’Institut de Statistique de L’Université de Paris*, 8, 229–231.
- Tauger, M. B. (2011). *Agriculture in World - History Themes in World History* (1st ed.). Routledge.
- Technow, F., Messina, C. D., Totir, L. R., & Cooper, M. (2015). Integrating crop growth models with whole genome prediction through approximate Bayesian computation. *PLoS ONE*, 10(6). <https://doi.org/10.1371/journal.pone.0130855>
- Tente, N., Westernhagen, N. von, & Slopek, U. (2019). M-PRESS-CreditRisk: Microprudential and Macroprudential Capital Requirements for Credit Risk under Systemic Stress. *Journal of Money, Credit and Banking*, 51(7), 1923–1961. <https://doi.org/10.1111/jmcb.12636>

- Thomas, L. C. (2009). Modelling the credit risk for portfolios of consumer loans: Analogies with corporate loan models. *Mathematics and Computers in Simulation*, 79(8), 2525–2534. <https://doi.org/10.1016/j.matcom.2008.12.006>
- Tiesset, M., & Troussard, P. (2005). Regulatory capital and economic capital. *Financial Stability Review*, November.
- Turvey, C. G., & Weersink, A. (1997). Credit risk and the demand for agricultural loans. *Canadian Journal of Agricultural Economics*, 45(3), 201–217. <https://doi.org/10.1111/j.1744-7976.1997.tb00204.x>
- Twumasi, M. A., Jiang, Y., & Acheampong, M. O. (2020). Capital and credit constraints in the engagement of youth in Ghanaian agriculture. *Agricultural Finance Review*, 80(1), 22–37. <https://doi.org/10.1108/AFR-11-2018-0100>
- Twumasi, M. A., Jiang, Y., Ntiamoah, E. B., Akaba, S., Darfor, K. N., & Boateng, L. K. (2021). Access to credit and farmland abandonment nexus: The case of rural Ghana. *Natural Resources Forum*. <https://doi.org/10.1111/1477-8947.12233>
- Vasicek, O. A. (1984). Credit Valuation. *KMV Corporation*. www.kmv.com
- Vasicek, O. A. (1987). Probability of Loss on Loan. *Kmv Corporation*.
- Weed, J. B., & Hardy, W. E. (1980). *Objective Credit Scoring of Alabama Borrowers*. Auburn University. Auburn, Alabama.
- Weisdorf, J. L. (2005). From Foraging To Farming: the So-Called “Neolithic Revolution.” *Journal of Economic Surveys*, 19(4).
- Wei, G. (2011). Copula parameter estimation by maximum-likelihood and minimum-distance estimators: A simulation study. *Computational Statistics*, 26(1), 31–54. <https://doi.org/10.1007/s00180-010-0203-7>
- Weitzman, M. L. (2009). On Modeling and Interpreting the Economics of Catastrophic Climate Change. *Review of Economics and Statistics*, 91(1), 1–19. <https://doi.org/10.1162/rest.91.1.1>
- Wong, F., & Collins, J. J. (2020). Evidence that coronavirus superspreading is fat-tailed. *Proceedings of the National Academy of Sciences of the United States of America*, 117(47), 29416–29418. <https://doi.org/10.1073/pnas.2018490117>
- Wu, H., Su, X., Singh, V. P., Feng, K., & Niu, J. (2021). Agricultural Drought Prediction Based on Conditional Distributions of Vine Copulas. *Water Resources Research*, 57(8), 1–23. <https://doi.org/10.1029/2021WR029562>
- Wu, X. Z. (2015). Modelling dependence structures of soil shear strength data with bivariate copulas and applications to geotechnical reliability analysis. *Soils and Foundations*, 55(5), 1243–1258. <https://doi.org/10.1016/j.sandf.2015.09.023>
- Xia, Y., Li, Y., He, L., Xu, Y., & Meng, Y. (2021). Incorporating multilevel macroeconomic variables into credit scoring for online consumer lending. *Electronic Commerce Research and Applications*, 49. <https://doi.org/10.1016/j.elerap.2021.101095>
- Yanaka, G. M. (2009). *Modelo Interno de Risco de Crdito de Basileia II: Possveis impactos no capital mnimo exigido dos bancos*. Fundao Getlio Vargas.
- Zech, L., & Pederson, G. (2004). Application of credit risk models to agricultural lending. *Agricultural Finance Review*, 64(2), 91–106. <https://doi.org/10.1108/00214660480001156>

- Zhang, S., Okhrin, O., Zhou, Q. M., & Song, P. X. K. (2016). Goodness-of-fit test for specification of semiparametric copula dependence models. *Journal of Econometrics*, *193*(1), 215–233. <https://doi.org/10.1016/j.jeconom.2016.02.017>
- Zhou, L., Fujita, H., Ding, H., & Ma, R. (2021). Credit risk modeling on data with two timestamps in peer-to-peer lending by gradient boosting. *Applied Soft Computing*, *110*. <https://doi.org/10.1016/j.asoc.2021.107672>
- Zilberman, D. (2019). Agricultural economics as a poster child of applied economics: Big data & big issues. *American Journal of Agricultural Economics*, *101*(2), 353–364. <https://doi.org/10.1093/ajae/aay101>

Anexo A – Estimação do ajuste das cópulas aos dados

I - Testes *GoF* híbrido e MME - $PD(A \times B)$

Os gráficos que ilustram o teste *GoF* híbrido foram apresentados na **Figura 4**.

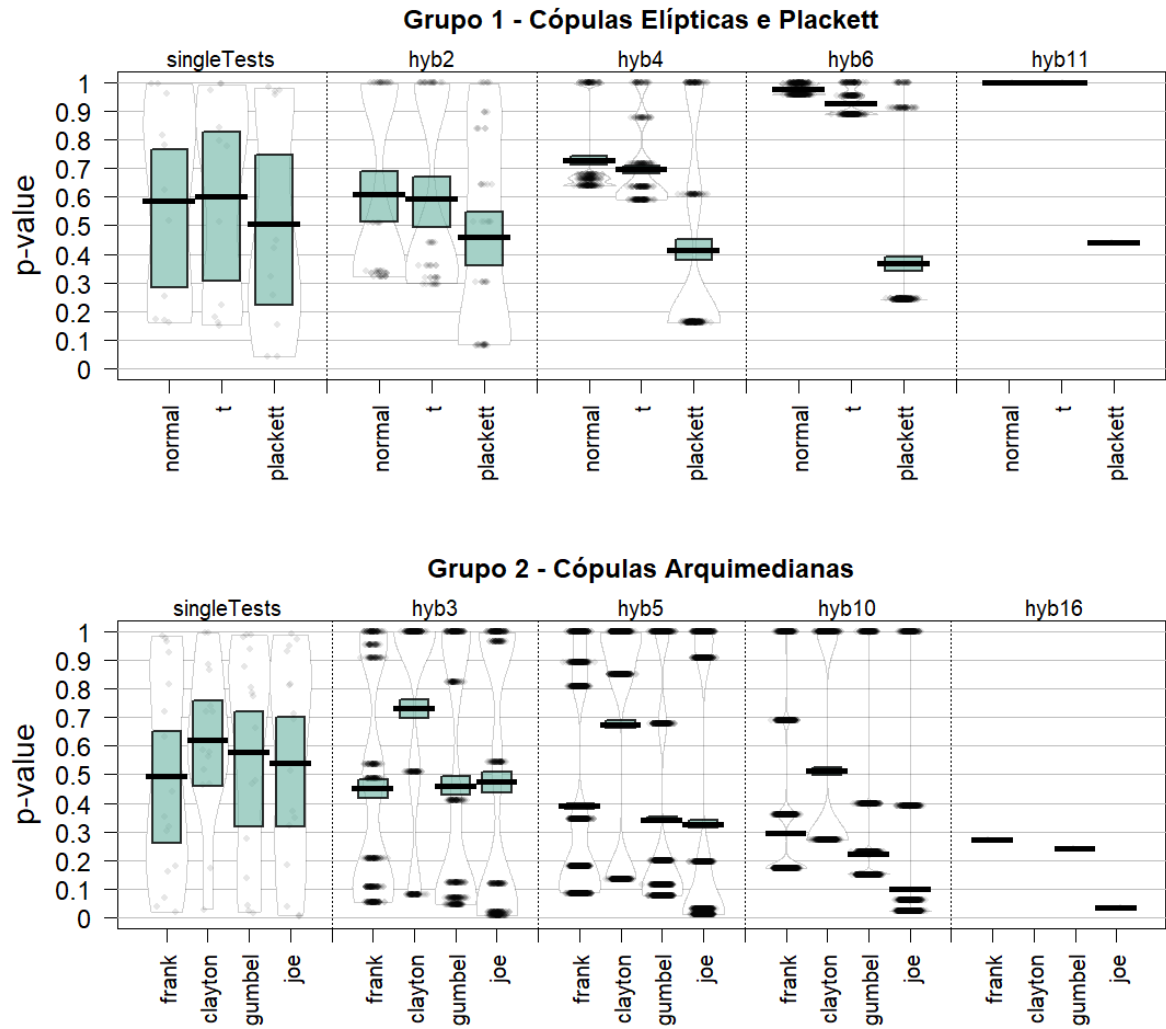
Tabela 28. *p*-valor dos testes *GoF* e estatística dos testes MME - $PD(A \times B)$

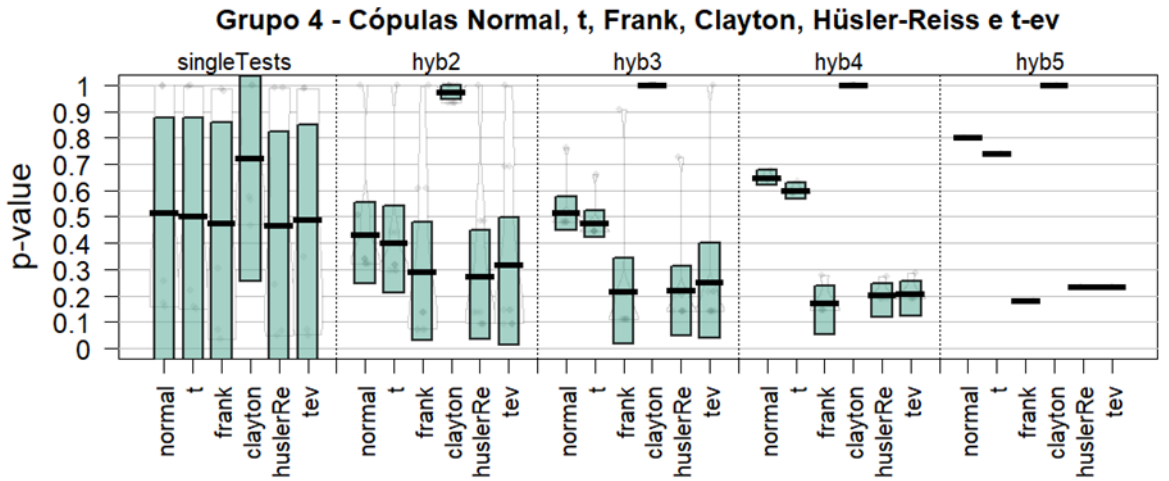
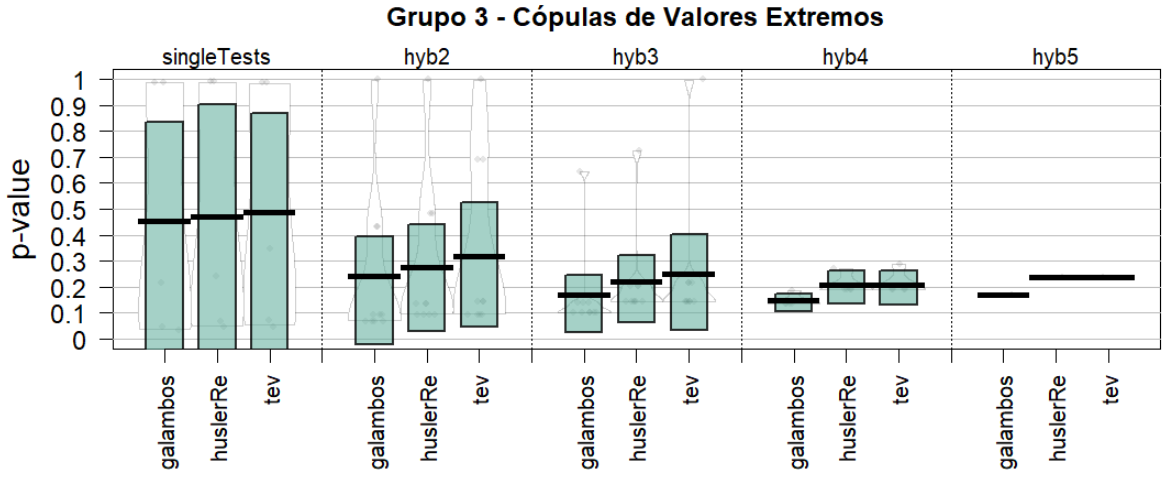
Testes Grupo 1	Normal	<i>t</i> -Student	Plackett	
<i>GoF CvM</i>	0,988	0,985	0,966	
<i>GoF KendallCvM</i>	0,041	0,071	0,043	
<i>GoF KendallKS</i>	0,071	0,048	0,052	
<i>GoF Kernel</i>	0,628	0,394	0,402	
<i>GoF KS</i>	0,988	0,988	0,965	
<i>GoF PIOSRn</i>	0,244	0,566	0,959	
<i>GoF PIOSTn</i>	0,246	0,891	0,862	
<i>GoF RosenblattChisq</i>	0,743	0,545	0,275	
<i>GoF RosenblattGamma</i>	0,938	0,760	0,834	
<i>GoF RosenblattSnB</i>	0,714	0,797	0,866	
<i>GoF RosenblattSnC</i>	0,668	0,747	0,841	
<i>GoF_{hb}</i>	0,451	0,528	0,473	
<i>MME Kendall Tau</i>	26,375	27,415	24,213	
<i>MME Spearman Rho</i>	26,484	-	24,286	
Testes Grupo 2	Frank	Clayton	Gumbel	Joe
<i>GoF ArchmChisq</i>	0,072	0,179	0,710	0,537
<i>GoF ArchmGamma</i>	0,610	0,760	0,540	0,536
<i>GoF ArchmSnB</i>	0,354	0,686	0,255	0,104
<i>GoF ArchmSnC</i>	0,401	0,915	0,310	0,134
<i>GoF CvM</i>	0,944	1,000	0,862	0,431
<i>GoF KendallCvM</i>	0,007	0,779	0,002	0,000
<i>GoF KendallKS</i>	0,037	0,630	0,004	0,000
<i>GoF Kernel</i>	0,650	0,814	0,268	0,010
<i>GoF KS</i>	0,971	1,000	0,858	0,455
<i>GoF PIOSRn</i>	0,846	0,185	0,384	0,274
<i>GoF PIOSTn</i>	1,000	0,110	0,237	0,161
<i>GoF RosenblattChisq</i>	0,389	0,461	0,916	0,770
<i>GoF RosenblattGamma</i>	0,898	0,941	0,926	0,746
<i>GoF RosenblattSnB</i>	0,888	0,896	0,553	0,494
<i>GoF RosenblattSnC</i>	0,752	0,980	0,515	0,220
<i>GoF White</i>	0,838	0,488	0,365	0,221
<i>GoF_{hb}</i>	0,112	1,000	0,032	0,000
<i>MME Kendall Tau</i>	23,069	31,973	22,174	8,147
<i>MME Spearman Rho</i>	23,294	32,104	22,447	-
Testes Grupo 3	Galambos	Hüsler-Reiss	t-EV	
<i>GoF CvM</i>	0,872	0,839	0,851	
<i>GoF KendallCvM</i>	0,001	0,000	0,003	
<i>GoF KendallKS</i>	0,006	0,001	0,011	
<i>GoF Kernel</i>	0,250	0,165	0,331	
<i>GoF KS</i>	0,868	0,826	0,844	
<i>GoF_{hb}</i>	0,005	0,000	0,015	
<i>MME Kendall Tal</i>	21,769	19,883	22,241	
<i>MME Spearman Rho</i>	22,077	20,554	22,430	

Fonte: elaborado pelo autor.

II- Testes *GoF* híbrido e MME - $PD(A \times C)$

Figura 9. Teste *GoF* híbrido $PD(A \times C)$





Fonte: elaborado pelo autor.

Tabela 29. *p*-valor dos testes *GoF* e estatística dos testes MME - $PD(A \times C)$

Testes Grupo 1	Normal	<i>t</i>-Student	Plackett	
<i>GoF CvM</i>	0,996	0,997	0,976	
<i>GoF KendallCvM</i>	0,160	0,159	0,040	
<i>GoF KendallKS</i>	0,254	0,220	0,041	
<i>GoF Kernel</i>	0,170	0,148	0,420	
<i>GoF KS</i>	0,999	0,998	0,984	
<i>GoF PIOSRn</i>	0,626	0,843	0,256	
<i>GoF PIOSTn</i>	0,817	0,780	0,322	
<i>GoF RosenblattChisq</i>	0,166	0,179	0,152	
<i>GoF RosenblattGamma</i>	0,516	0,512	0,449	
<i>GoF RosenblattSnB</i>	0,783	0,796	0,959	
<i>GoF RosenblattSnC</i>	0,964	0,974	0,971	
<i>GoF_{hb}</i>	1,000	1,000	0,440	
<i>MME Kendall Tau</i>	2,418	1,218	0,464	
<i>MME Spearman Rho</i>	2,293	-	0,590	
Testes Grupo 2	Frank	Clayton	Gumbel	Joe
<i>GoF ArchmChisq</i>	0,017	0,027	0,023	0,039
<i>GoF ArchmGamma</i>	0,815	0,865	0,786	0,715
<i>GoF ArchmSnB</i>	0,631	0,463	0,664	0,695
<i>GoF ArchmSnC</i>	0,720	0,518	0,776	0,812
<i>GoF CvM</i>	0,977	0,998	0,982	0,931
<i>GoF KendallCvM</i>	0,036	0,564	0,015	0,002
<i>GoF KendallKS</i>	0,069	0,468	0,040	0,006
<i>GoF Kernel</i>	0,303	0,578	0,136	0,182
<i>GoF KS</i>	0,987	0,999	0,988	0,950
<i>GoF PIOSRn</i>	0,179	0,588	0,469	0,515
<i>GoF PIOSTn</i>	0,318	0,741	0,877	0,350
<i>GoF RosenblattChisq</i>	0,162	0,170	0,275	0,322
<i>GoF RosenblattGamma</i>	0,441	0,720	0,479	0,370
<i>GoF RosenblattSnB</i>	0,930	0,719	0,938	0,976
<i>GoF RosenblattSnC</i>	0,965	0,885	0,990	0,994
<i>GoF White</i>	0,351	-	0,806	0,816
<i>GoF_{hb}</i>	0,272	-	0,240	0,032
<i>MME Kendall Tau</i>	0,664	3,478	1,870	0,390
<i>MME Spearman Rho</i>	0,745	3,328	1,949	-
Testes Grupo 3	Galambos	Hüsler-Reiss		<i>t</i>-EV
<i>GoF CvM</i>	0,987	0,993		0,99
<i>GoF KendallCvM</i>	0,034	0,047		0,047
<i>GoF KendallKS</i>	0,046	0,068		0,072
<i>GoF Kernel</i>	0,215	0,242		0,346
<i>GoF KS</i>	0,990	0,991		0,988
<i>GoF_{hb}</i>	0,170	0,235		0,235
<i>MME Kendall Tau</i>	2,189	2,284		1,963
<i>MME Spearman Rho</i>	2,203	2,279		2,016

Fonte: elaborado pelo autor.

III - Testes GoF híbrido e MME - PD(AxD)

Figura 10. Teste *GoF* híbrido $PD(A \times D)$

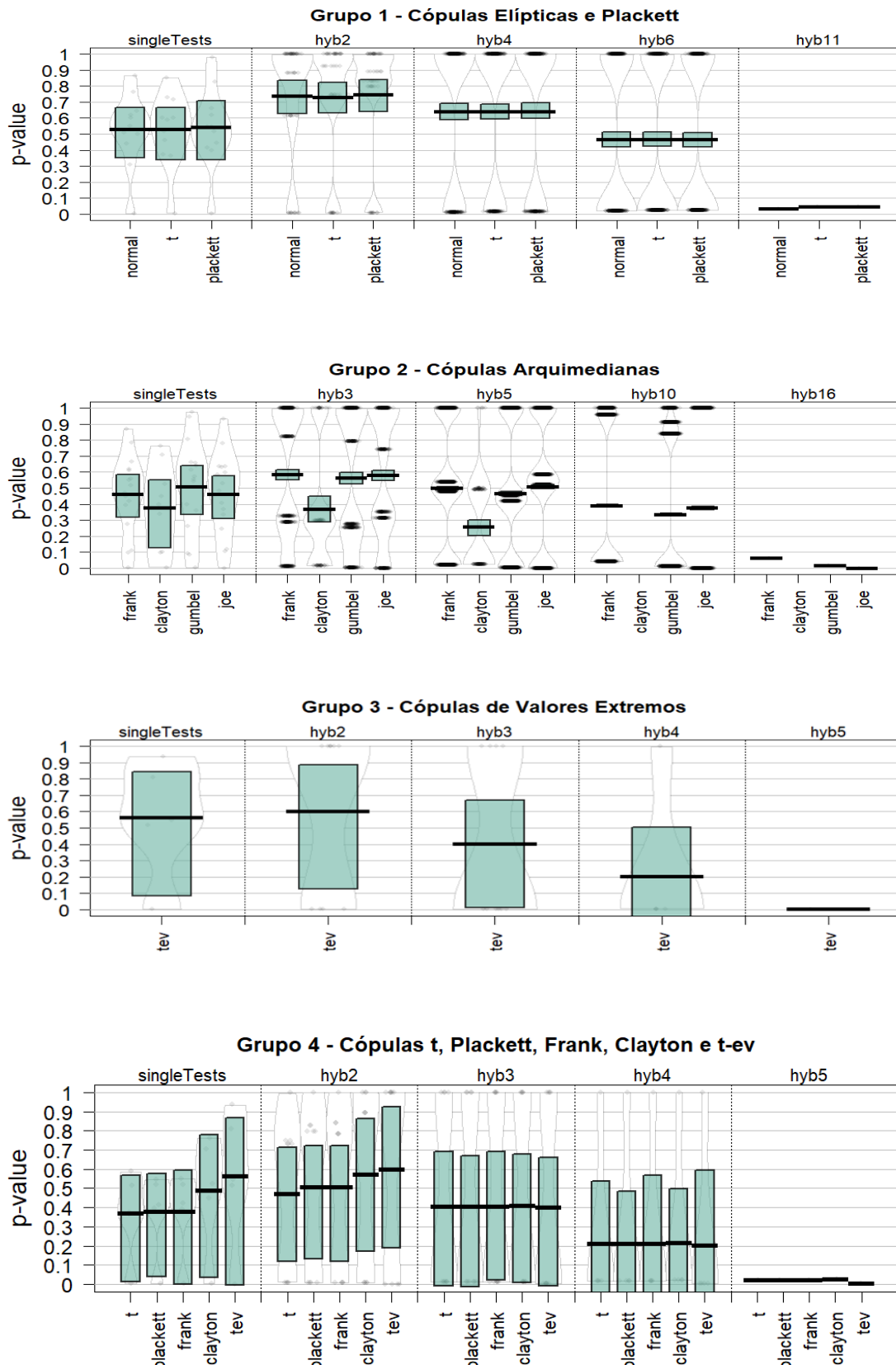


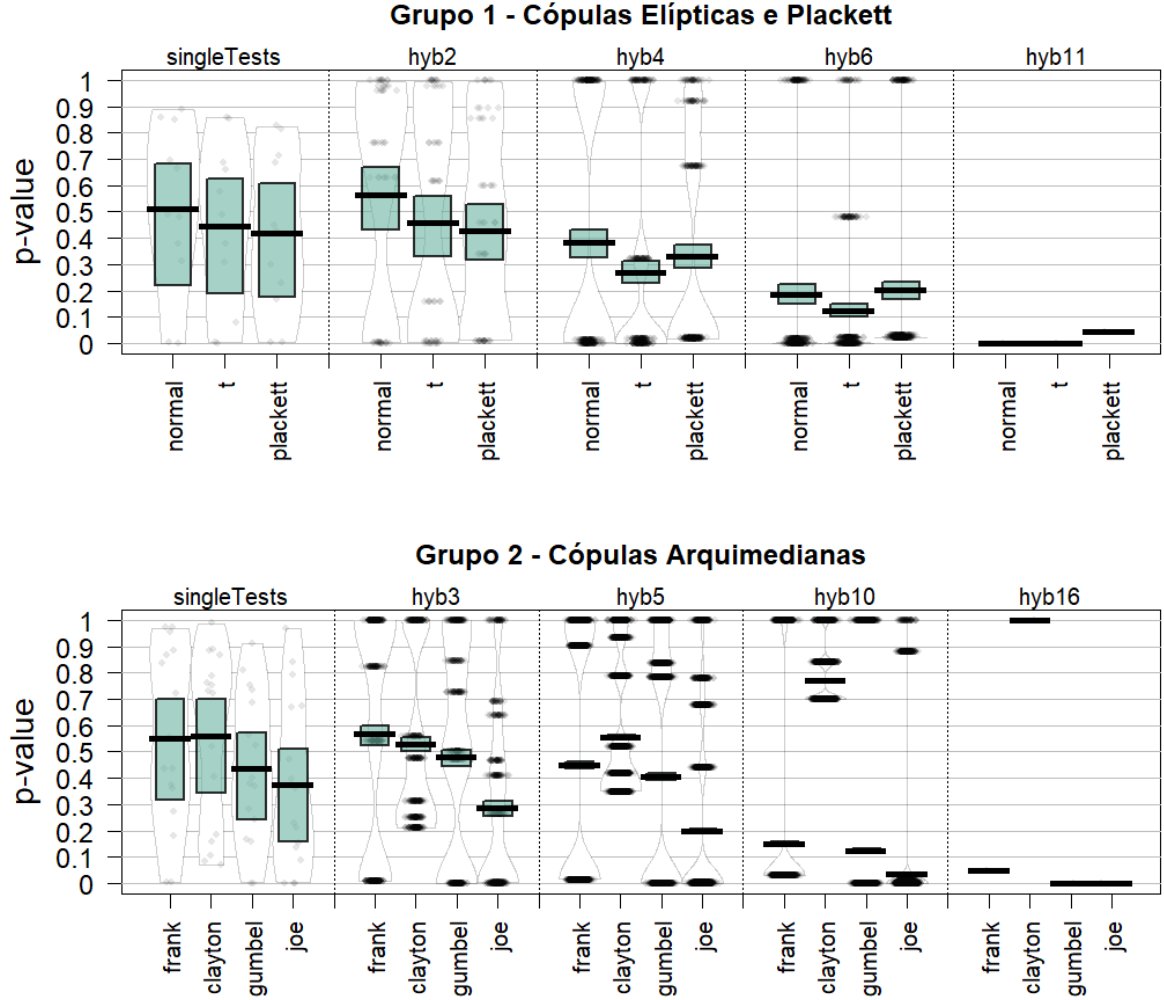
Tabela 30. *p*-valor dos testes *GoF* e estatística dos testes MME - $PD(A \times D)$

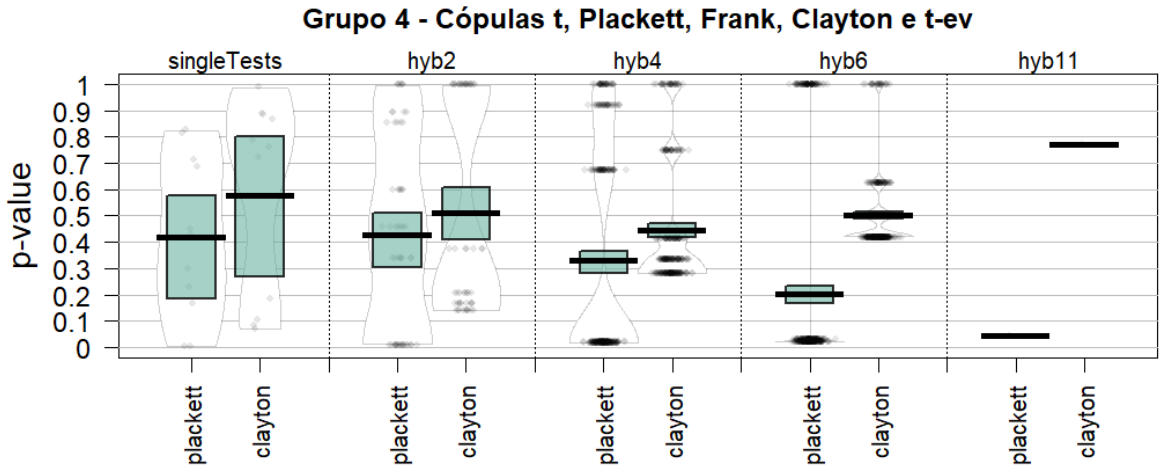
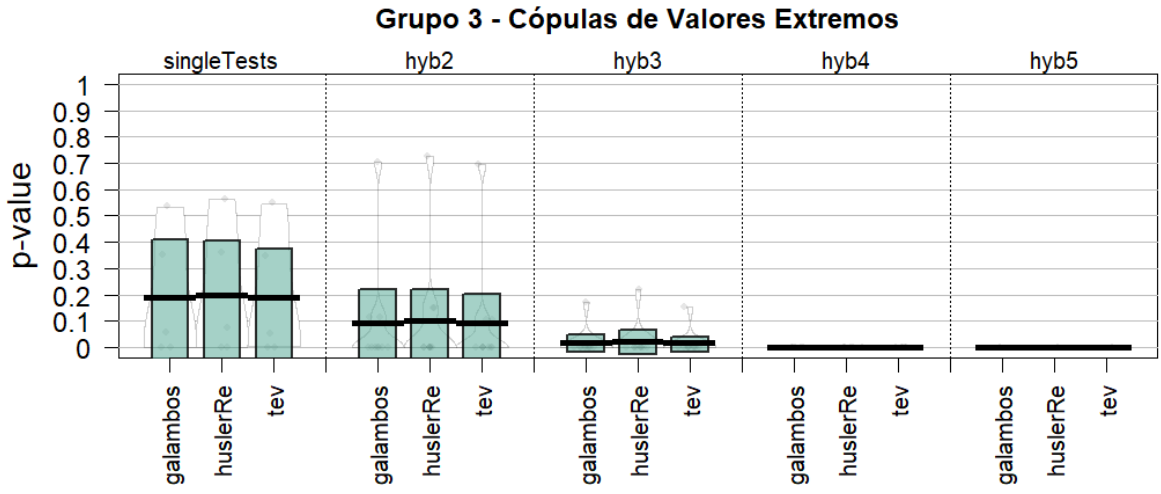
Testes Grupo 1	Normal	<i>t</i>-Student	Plackett	
<i>GoF CvM</i>	0,441	0,366	0,398	
<i>GoF KendallCvM</i>	0,763	0,589	0,548	
<i>GoF KendallKS</i>	0,865	0,516	0,519	
<i>GoF Kernel</i>	0,003	0,004	0,004	
<i>GoF KS</i>	0,509	0,374	0,415	
<i>GoF PIOSRn</i>	0,307	0,73	0,982	
<i>GoF PIOSTn</i>	0,644	0,854	0,829	
<i>GoF RosenblattChisq</i>	0,603	0,600	0,548	
<i>GoF RosenblattGamma</i>	0,617	0,715	0,663	
<i>GoF RosenblattSnB</i>	0,553	0,600	0,620	
<i>GoF RosenblattSnC</i>	0,501	0,462	0,446	
<i>GoF_{hb}</i>	0,033	0,044	0,044	
<i>MME Kendall Tau</i>	-1,203	-1,622	-0,843	
<i>MME Spearman Rho</i>	-1,412	-	-0,781	
Testes Grupo 2	Frank	Clayton	Gumbel	Joe
<i>GoF ArchmChisq</i>	0,275	0,337	0,398	0,370
<i>GoF ArchmGamma</i>	0,469	0,399	0,264	0,247
<i>GoF ArchmSnB</i>	0,108	0,100	0,091	0,104
<i>GoF ArchmSnC</i>	0,096	0,098	0,084	0,117
<i>GoF CvM</i>	0,393	0,448	0,571	0,597
<i>GoF KendallCvM</i>	0,551	0,707	0,805	0,779
<i>GoF KendallKS</i>	0,519	0,764	0,945	0,933
<i>GoF Kernel</i>	0,004	0,005	0,001	0
<i>GoF KS</i>	0,421	0,525	0,555	0,528
<i>GoF PIOSRn</i>	0,868	-	0,613	0,485
<i>GoF PIOSTn</i>	0,609	-	0,976	0,405
<i>GoF RosenblattChisq</i>	0,560	-	0,658	0,635
<i>GoF RosenblattGamma</i>	0,666	-	0,642	0,635
<i>GoF RosenblattSnB</i>	0,619	-	0,359	0,336
<i>GoF RosenblattSnC</i>	0,449	-	0,542	0,586
<i>GoF White</i>	0,786	-	0,661	0,631
<i>GoF_{hb}</i>	0,064	-	0,016	0
<i>MME Kendall Tau</i>	-0,859	-0,366	0	0
<i>MME Spearman Rho</i>	-0,810	-0,130	0,022	-
Testes Grupo 3	Galambos	Hüsler-Reiss	<i>t</i>-EV	
<i>GoF CvM</i>	-	-	0,548	
<i>GoF KendallCvM</i>	-	-	0,812	
<i>GoF KendallKS</i>	-	-	0,937	
<i>GoF Kernel</i>	-	-	0,001	
<i>GoF KS</i>	-	-	0,516	
<i>GoF_{hb}</i>	-	-	0,005	
<i>MME Kendall Tau</i>	0	0	0,366	
<i>MME Spearman Rho</i>	0	0	0,366	

Fonte: elaborado pelo autor.

IV - Testes GoF híbrido e MME - PD(A × E)

Figura 11. Teste GoF híbrido PD(A × E)





Fonte: elaborado pelo autor.

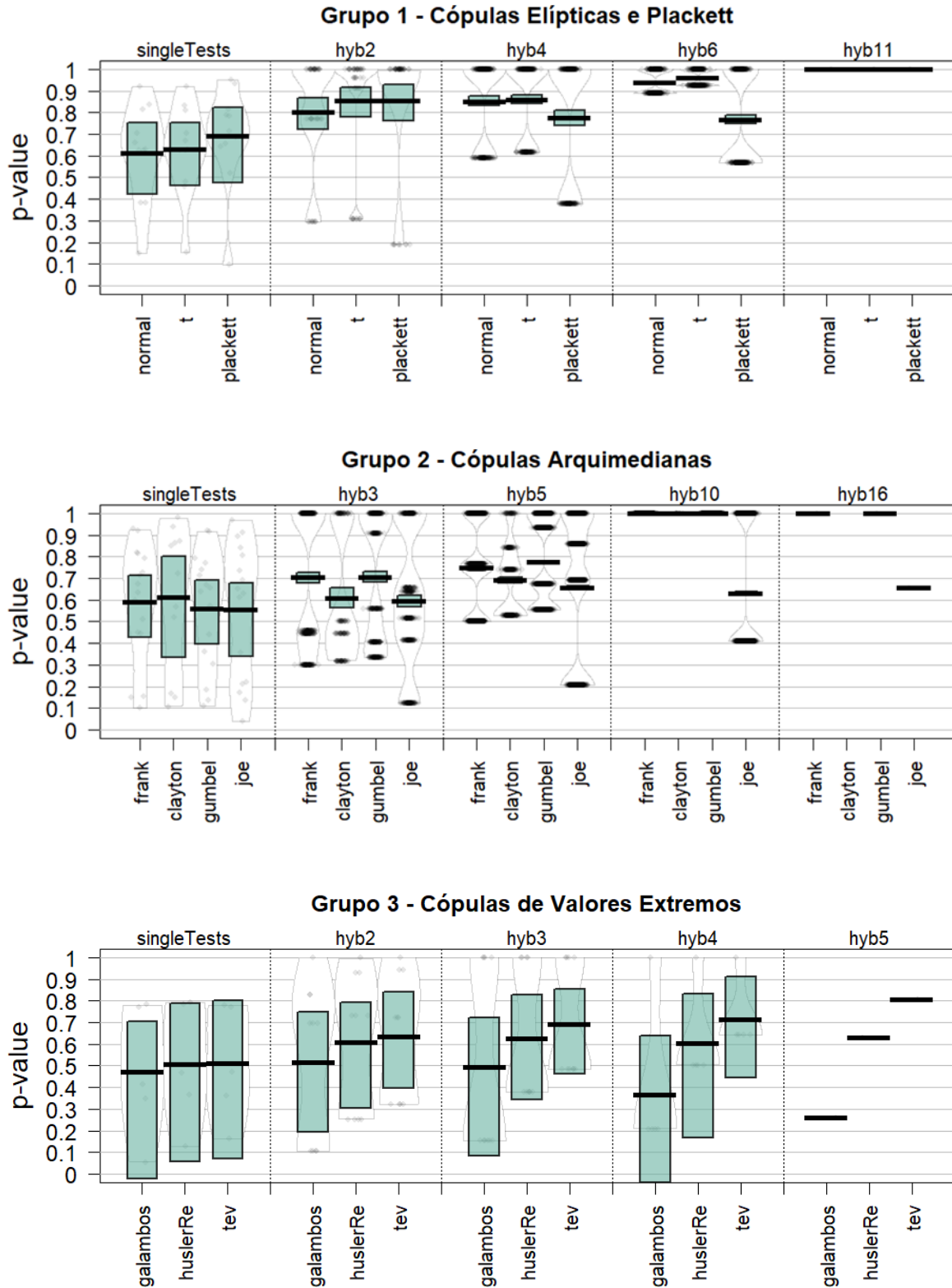
Tabela 31. *p*-valor dos testes *GoF* e estatística dos testes MME - $PD(A \times E)$

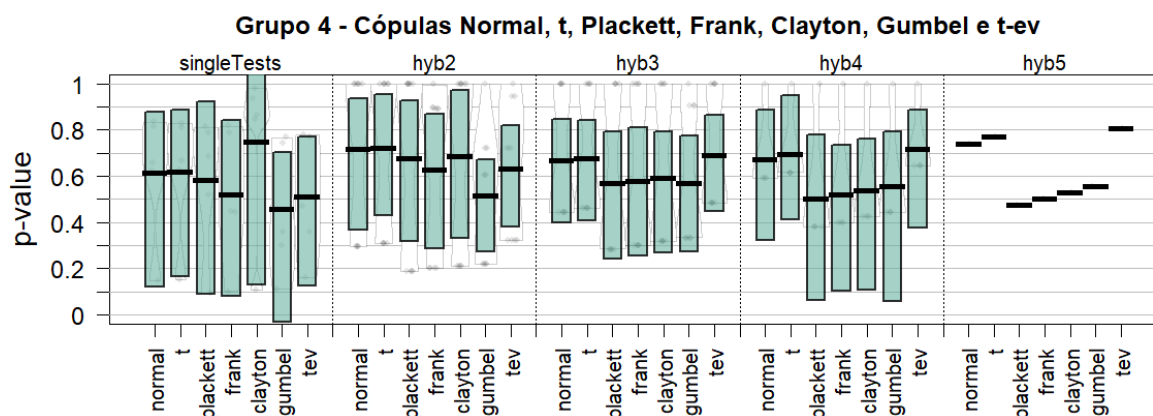
Testes Grupo 1	Normal	<i>t</i>-Student	Plackett	
<i>GoF CvM</i>	0,860	0,860	0,826	
<i>GoF KendallCvM</i>	0	0	0,005	
<i>GoF KendallKS</i>	0,003	0,004	0,004	
<i>GoF Kernel</i>	0,314	0,309	0,300	
<i>GoF KS</i>	0,663	0,662	0,687	
<i>GoF PIOSRn</i>	0,480	0,080	0,230	
<i>GoF PIOSTn</i>	0,892	0,577	0,169	
<i>GoF RosenblattChisq</i>	0,488	0,488	0,448	
<i>GoF RosenblattGamma</i>	0,849	0,853	0,813	
<i>GoF RosenblattSnB</i>	0,381	0,381	0,427	
<i>GoF RosenblattSnC</i>	0,696	0,687	0,713	
<i>GoF_{hb}</i>	0	0	0,044	
<i>MME Kendall Tau</i>	14,866	14,866	13,776	
<i>MME Spearman Rho</i>	15,136	-	13,916	
Testes Grupo 2	Frank	Clayton	Gumbel	Joe
<i>GoF ArchmChisq</i>	0,181	0,731	0,400	0,473
<i>GoF ArchmGamma</i>	0,958	0,752	0,810	0,672
<i>GoF ArchmSnB</i>	0,274	0,406	0,167	0,088
<i>GoF ArchmSnC</i>	0,436	0,521	0,282	0,213
<i>GoF CvM</i>	0,839	0,989	0,564	0,23
<i>GoF KendallCvM</i>	0,003	0,104	0	0
<i>GoF KendallKS</i>	0,003	0,07	0	0
<i>GoF Kernel</i>	0,361	0,723	0,157	0,001
<i>GoF KS</i>	0,722	0,889	0,377	0,136
<i>GoF PIOSRn</i>	0,973	0,187	0,754	0,841
<i>GoF PIOSTn</i>	0,886	0,084	0,914	0,971
<i>GoF RosenblattChisq</i>	0,376	0,79	0,524	0,670
<i>GoF RosenblattGamma</i>	0,868	0,867	0,734	0,361
<i>GoF RosenblattSnB</i>	0,434	0,762	0,371	0,395
<i>GoF RosenblattSnC</i>	0,554	0,886	0,242	0,156
<i>GoF White</i>	0,973	0,158	0,687	0,792
<i>GoF_{hb}</i>	0,048	1	0	0
<i>MME Kendall Tau</i>	14,142	15,819	9,065	-3,040
<i>MME Spearman Rho</i>	14,389	16,209	9,423	-
Testes Grupo 3	Galambos	Hüsler-Reiss	<i>t</i>-EV	
<i>GoF CvM</i>	0,537	0,565	0,551	
<i>GoF KendallCvM</i>	0	0	0	
<i>GoF KendallKS</i>	0	0	0	
<i>GoF Kernel</i>	0,058	0,074	0,052	
<i>GoF KS</i>	0,353	0,363	0,348	
<i>GoF_{hb}</i>	0	0	0	
<i>MME Kendall Tau</i>	9,055	9,090	8,615	
<i>MME Spearman Rho</i>	9,482	9,711	8,912	

Fonte: elaborado pelo autor.

V - Testes *GoF* híbrido e *MME - PD(A × F*)*

Figura 12. Teste *GoF* híbrido *PD(A × F*)*





Fonte: elaborado pelo autor.

Tabela 32. *p*-valor dos testes *GoF* e estatística dos testes MME - $PD(A \times F^*)$

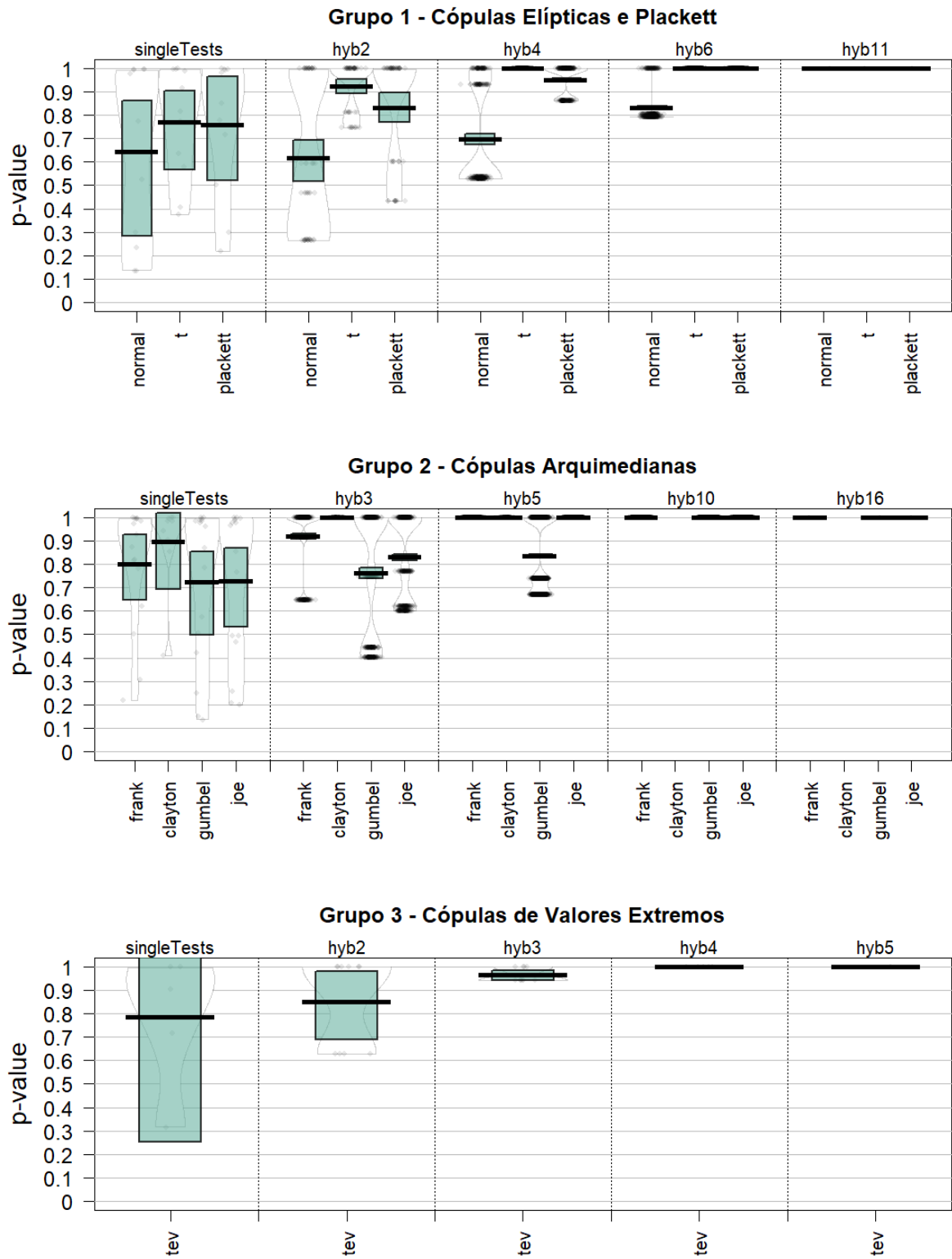
Testes Grupo 1	Normal	<i>t</i> -Student	Plackett	
<i>GoF CvM</i>	0,814	0,812	0,790	
<i>GoF KendallCvM</i>	0,624	0,626	0,521	
<i>GoF KendallKS</i>	0,660	0,668	0,685	
<i>GoF Kernel</i>	0,148	0,154	0,095	
<i>GoF KS</i>	0,836	0,831	0,814	
<i>GoF PIOSRn</i>	0,385	0,456	0,782	
<i>GoF PIOSTn</i>	0,385	0,481	0,953	
<i>GoF RosenblattChisq</i>	0,707	0,706	0,713	
<i>GoF RosenblattGamma</i>	0,921	0,920	0,935	
<i>GoF RosenblattSnB</i>	0,626	0,624	0,658	
<i>GoF RosenblattSnC</i>	0,628	0,628	0,641	
<i>GoF_{hb}</i>	1,000	1,000	1,000	
<i>MME Kendall Tau</i>	-0,6351	-0,6357	-0,741	
<i>MME Spearman Rho</i>	-0,540	-	-0,680	
Testes Grupo 2	Frank	Clayton	Gumbel	Joe
<i>GoF ArchmChisq</i>	0,509	0,522	0,442	0,358
<i>GoF ArchmGamma</i>	0,601	0,570	0,656	0,632
<i>GoF ArchmSnB</i>	0,149	0,148	0,135	0,172
<i>GoF ArchmSnC</i>	0,153	0,168	0,187	0,219
<i>GoF CvM</i>	0,791	0,864	0,742	0,691
<i>GoF KendallCvM</i>	0,446	0,937	0,302	0,138
<i>GoF KendallKS</i>	0,449	0,981	0,361	0,213
<i>GoF Kernel</i>	0,100	0,106	0,111	0,041
<i>GoF KS</i>	0,819	0,848	0,769	0,734
<i>GoF PIOSRn</i>	0,817	-	0,571	0,807
<i>GoF PIOSTn</i>	0,930	-	0,917	0,911
<i>GoF RosenblattChisq</i>	0,726	0,723	0,668	0,652
<i>GoF RosenblattGamma</i>	0,919	0,871	0,919	0,894
<i>GoF RosenblattSnB</i>	0,672	-	0,793	0,841
<i>GoF RosenblattSnC</i>	0,633	-	0,638	0,613
<i>GoF White</i>	0,715	-	0,714	0,968
<i>GoF_{hb}</i>	1,000	-	1,000	0,656
<i>MME Kendall Tau</i>	-0,754	-1,167	-1,336	-1,940
<i>MME Spearman Rho</i>	-0,672	-0,911	-1,012	-
Testes Grupo 3	Galambos	Hüsler-Reiss	<i>t</i> -EV	
<i>GoF CvM</i>	0,769	0,784	0,771	
<i>GoF KendallCvM</i>	0,348	0,365	0,361	
<i>GoF KendallKS</i>	0,415	0,465	0,472	
<i>GoF Kernel</i>	0,052	0,126	0,161	
<i>GoF KS</i>	0,782	0,793	0,781	

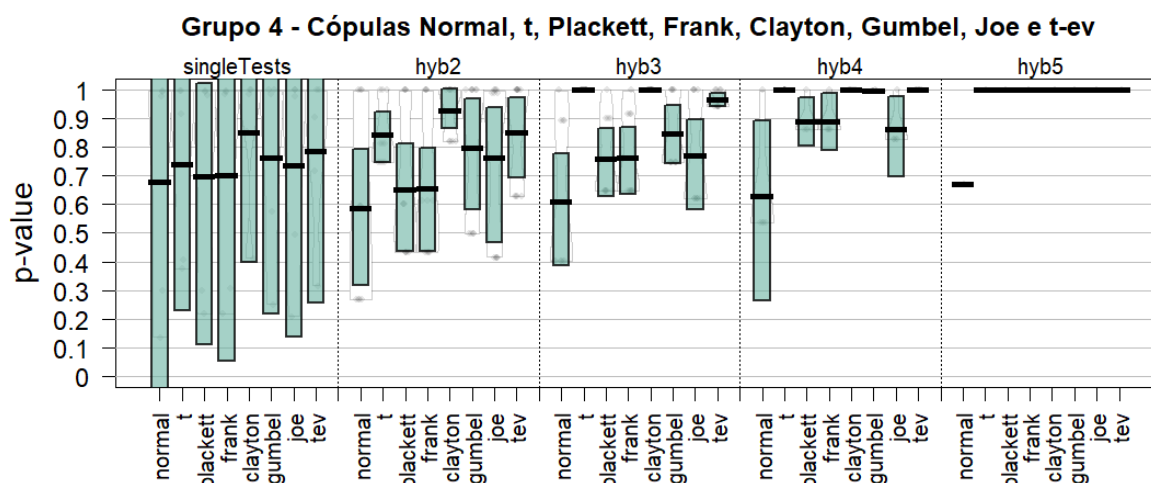
<i>GoF_{hb}</i>	0,260	0,630	0,805
<i>MME Kendall Tau</i>	-1,215	-1,144	-1,357
<i>MME Spearman Rho</i>	-0,885	-0,821	-1,021

Fonte: elaborado pelo autor.

VI - Testes *GoF* híbrido e *MME - PD(AxG)*

Figura 13. Teste *GoF* híbrido $PD(A \times G)$





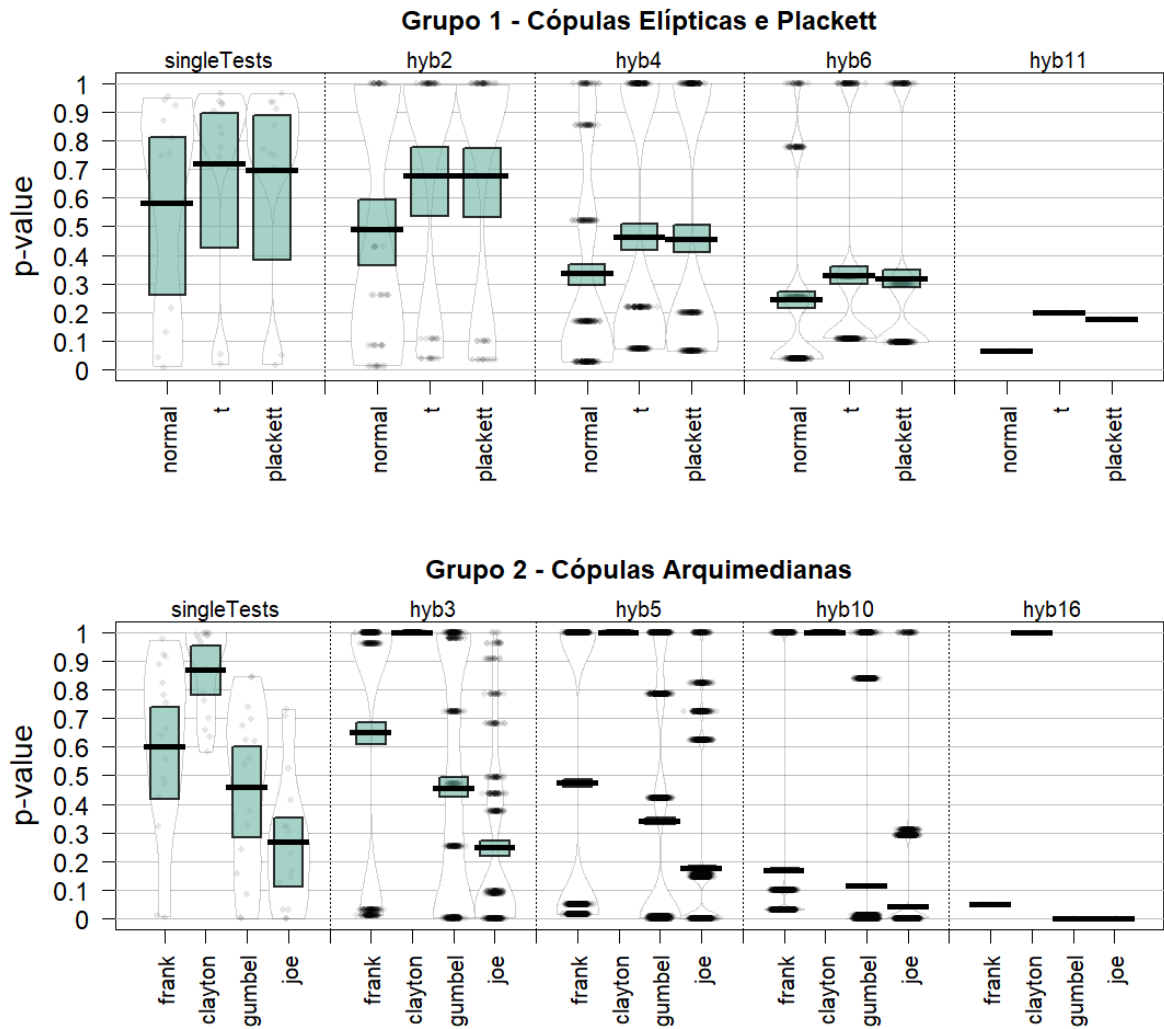
Fonte: elaborado pelo autor.

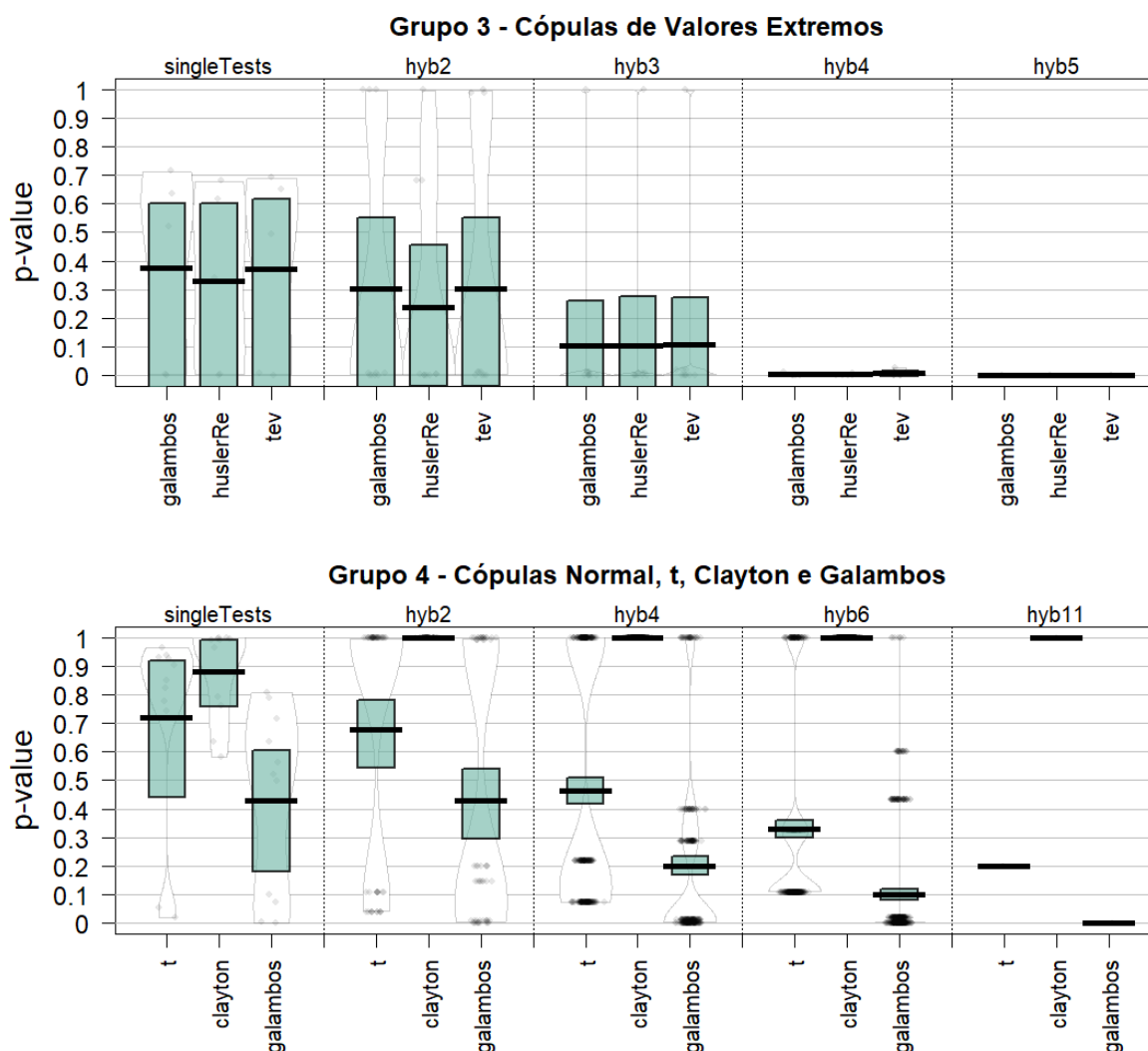
Tabela 33 p -valor dos testes GoF e estatística dos testes MME - $PD(A \times G)$

Testes Grupo 1	Normal	t -Student	Plackett	
GoF CvM	0,998	0,999	0,999	
GoF KendallCvM	0,298	0,407	0,300	
GoF KendallKS	0,134	0,374	0,216	
GoF Kernel	0,978	0,915	0,981	
GoF KS	0,993	0,996	0,994	
GoF PIOSRn	0,132	0,635	0,851	
GoF PIOSTn	0,233	0,815	0,716	
GoF RosenblattChisq	0,524	0,578	0,502	
GoF RosenblattGamma	0,776	0,773	0,78	
GoF RosenblattSnB	0,999	0,989	0,997	
GoF RosenblattSnC	0,999	1,000	1,000	
GoF_{hb}	1,000	1,000	1,000	
MME Kendal Tau	-1,104	-1,599	-0,972	
MME Spearman Rho	-1,052	-	-0,930	
Testes Grupo 2	Frank	Clayton	Gumbel	Joe
GoF ArchmChisq	0,821	0,904	0,869	0,856
GoF ArchmGamma	0,975	0,986	0,985	0,991
GoF ArchmSnB	0,996	0,996	0,995	0,997
GoF ArchmSnC	0,928	0,942	0,963	0,966
GoF CvM	0,999	1,000	1,000	1,000
GoF KendallCvM	0,306	0,854	0,575	0,494
GoF KendallKS	0,216	0,411	0,249	0,207
GoF Kernel	0,984	0,984	0,985	0,98
GoF KS	0,995	1	1,000	1,000
GoF PIOSRn	0,873	-	0,134	0,257
GoF PIOSTn	0,620	-	0,421	0,468
GoF RosenblattChisq	0,502	-	0,506	0,495
GoF RosenblattGamma	0,782	-	0,787	0,767
GoF RosenblattSnB	0,997	-	0,978	0,98
GoF RosenblattSnC	1,000	-	0,99	0,994
GoF White	0,818	-	0,148	0,200
GoF_{hb}	1,000	-	1,000	1,000
MME Kendal Tau	-0,986	-6,514	0,000	0,000
MME Spearman Rho	-0,940	-5,651	-0,012	-
Testes Grupo 3	Galambos	Hüsler-Reiss	t -EV	
GoF CvM	-	-	1,000	
GoF KendallCvM	-	-	0,718	

<i>GoF KendallKS</i>	-	-	0,315
<i>GoF Kernel</i>	-	-	0,904
<i>GoF KS</i>	-	-	1,000
<i>GoF_{hb}</i>	-	-	1,000
<i>MME Kendall Tau</i>	-	-	-0,820
<i>MME Spearman Rho</i>	-	-	-0,820

Fonte: elaborado pelo autor.

VII - Testes *GoF* híbrido e *MME - PD(BxE)*Figura 14. Teste *GoF* híbrido $PD(B \times E)$ 



Fonte: elaborado pelo autor.

Tabela 34. p -valor dos testes GoF e estatística dos testes $MME - PD(B \times E)$

Testes Grupo 1	Normal	t -Student	Plackett	
GoF_{CvM}	0,954	0,967	0,935	
$GoF_{KendallCvM}$	0,006	0,018	0,016	
$GoF_{KendallKS}$	0,042	0,054	0,050	
GoF_{Kernel}	0,869	0,826	0,751	
GoF_{KS}	0,924	0,941	0,935	
GoF_{PIOSRn}	0,130	0,742	0,854	
GoF_{PIOSTn}	0,214	0,779	0,912	
$GoF_{RosenblattChisq}$	0,746	0,852	0,706	
$GoF_{RosenblattGamma}$	0,943	0,932	0,968	
$GoF_{RosenblattSnB}$	0,754	0,904	0,756	
$GoF_{RosenblattSnC}$	0,809	0,929	0,772	
GoF_{hb}	0,066	0,198	0,176	
$MME_{Kendall\ Tau}$	14,846	15,487	15,480	
$MME_{Spearman\ Rho}$	15,201	-	15,576	
Testes Grupo 2	Frank	Clayton	Gumbel	Joe
$GoF_{ArchmChisq}$	0,426	0,661	0,622	0,321
$GoF_{ArchmGamma}$	0,755	0,874	0,673	0,733
$GoF_{ArchmSnB}$	0,554	0,991	0,327	0,125
$GoF_{ArchmSnC}$	0,664	0,983	0,374	0,165

<i>GoF CvM</i>	0,923	1,000	0,741	0,322
<i>GoF KendallCvM</i>	0,003	0,635	0,000	0,000
<i>GoF KendallKS</i>	0,010	0,581	0,001	0,000
<i>GoF Kernel</i>	0,889	0,997	0,625	0,029
<i>GoF KS</i>	0,917	1,000	0,698	0,303
<i>GoF PIOSRn</i>	0,471	0,792	0,241	0,227
<i>GoF PIOSTn</i>	0,321	0,965	0,157	0,145
<i>GoF RosenblattChisq</i>	0,823	0,762	0,844	0,708
<i>GoF RosenblattGamma</i>	0,978	0,993	0,849	0,412
<i>GoF RosenblattSnB</i>	0,783	0,995	0,559	0,524
<i>GoF RosenblattSnC</i>	0,639	0,999	0,541	0,262
<i>GoF White</i>	0,490	0,702	0,084	0,031
<i>GoF_{hb}</i>	0,048	1,000	0,000	0,000
<i>MME Kendall Tau</i>	14,364	20,344	10,461	-2,320
<i>MME Spearman Rho</i>	14,586	20,453	11,149	-
Testes Grupo 3	Galambos	Hüsler-Reiss	t-EV	
<i>GoF CvM</i>	0,716	0,681	0,694	
<i>GoF KendallCvM</i>	0,000	0,000	0,000	
<i>GoF KendallKS</i>	0,003	0,002	0,006	
<i>GoF Kernel</i>	0,521	0,341	0,495	
<i>GoF KS</i>	0,637	0,619	0,653	
<i>GoF_{hb}</i>	0,000	0,000	0,000	
<i>MME Kendall Tau</i>	9,835	7,945	10,588	
<i>MME Spearman Rho</i>	10,629	9,403	11,133	

Fonte: elaborado pelo autor.

Anexo B – Códigos do Software R

Anexo B – Códigos do Software R

B.1 - Carregar os dados e estimar a PD mensal das linhas de crédito

```

1 ##### Carregar as bibliotecas
2 library(DBI)
3 library(RODBC)
4 library(tidyverse)
5 library(formattable)
6 ##### Fazer conexao ODBC para os dados de 2014
7     con <- odbcConnect("agro_2014_64")
8
9 ##### Selecionar os dados de 10/2014
10    qry<- "SELECT * FROM
11         opr_agro_201410"
12
13    base_agro <- sqlQuery(con,qry)
14
15 ##### Filtrar as linhas de crédito rural excluindo operacoes com mais de 120
16     ↳ de atraso
17     base_agro <- filter(base_agro,
18         nm_prd %in% c("AGRONEGOCIOS-FIN. GARANTIA DE PREÇOS E ESTOCAGEM",
19         ↳ "AQUISICAO ATIVOS FINANCEIROS",
20         "CUSTEIO AGROPECUARIO","GIRO","INVESTIMENTO AGROPECUARIO",
21         "PRONAF AGRICULTURA FAMILIAR","PRONAMP"),
22         nr_dd_vcd_opr <= 120)
23
24 ##### Gerar dummy de default
25     base_agro <- mutate(base_agro, default = case_when(nr_dd_vcd_opr >= 90
26         ↳ ~ 1,
27         nr_dd_vcd_opr <90 ~ 0,TRUE ~ 0))
28
29 ##### Calculo a estatistica descritiva
30     descritiva <- summarise(base_agro,
31         mean_vl= mean(vl_sdo_ctb),
32         median_vl= median(vl_sdo_ctb),
33         max_vl= max(vl_sdo_ctb),
34         sd_vl = sd(vl_sdo_ctb))
35
36 ##### Contar numero total de operacoes e operacoes em default
37     contagem <- count(base_agro,nm_prd)
38     contagem <- mutate(contagem,data =
39         "10/2014")
40     contagem_default <- count(base_agro,nm_prd,default)
41     contagem_default <- filter(contagem_default, default ==1)
42     contagem_default <- rename(contagem_default, c("n_default" = "n"))
43     contagem <- left_join(contagem,contagem_default, by = "nm_prd")
44     contagem <- select(contagem,nm_prd,data,n_default,n)

```

```

42     base_agro <- group_by(base_agro, nm_prd)
43
44 ##### Gero a base de operacoes em nao default em 10/2014, isto servira para o
45 ↪ calculo da PD em 01/2015 (t+3)
46     n_90dias <- mutate(contagem, n = (n-n_default))
47     n_90dias <- select(n_90dias,nm_prd,n)
48     n_1410 <-rename(n_90dias, c("n_90dias" = "n"))
49
50 ##### Calculo a PD em 10/2014 tendo como referência a base de operações em
51 ↪ nao default em 07/2014 (t-3)
52
53 base_default <-mutate(base_default,pd_default = percent(n_default/n_90dias))
54     default_1410 <-select(base_default,nm_prd,data,n_default,n_90dias,
55     pd_default,mean_vl,median_vl,max_vl,sd_vl)
56
57 ##### Repito este processo para os 63 meses da base de dados (07/2014 a
58 ↪ 09/2019). Os tres primeiros meses da amostra (07, 08 e 09 de 2014) nao tem
59 ↪ a PD calculada, uma vez que o calculo da PD requer a base de operacoes em
60 ↪ nao default em t-3
61
62 ##### Consolido todas as PDs mensais em uma base
63 base_default <-rbind(default_1407,default_1408,default_1409,default_1410,
64 default_1411,default_1412,default_1501,default_1502,default_1503,default_1504,
65 default_1505,default_1506,default_1507,default_1508,default_1509,default_1510,
66 default_1511,default_1512,default_1601,default_1602,default_1603,default_1604,
67 default_1605,default_1606,default_1607,default_1608,default_1609,default_1610,
68 default_1611,default_1612,default_1701,default_1702,default_1703,default_1704,
69 default_1705,default_1706,default_1707,default_1708,default_1709,default_1710,
70 default_1711,default_1712,default_1801,default_1802,default_1803,default_1804,
71 default_1805,default_1806,default_1807,default_1808,default_1809,default_1810,
72 default_1811,default_1812,default_1901,default_1902,default_1903,default_1904,
73 default_1905,default_1906,default_1907,default_1908,default_1909)
74
75 base_default <-select(base_default,data,nm_prd,pd_default)
76
77 # Setting working directory to the same path where the R Project is
78 here::here()

```

B.2 - Figura 2. Taxa de default mensal - Aquisição de Ativos Financeiros

```

1 ##### Libraries
2     library(tidyverse)
3     library(hrbrthemes)
4     library(babynames)
5     library(viridis)
6     library(DT)
7     library(plotly)

```



```

8     library(tidyverse)
9
10    data_f <-filter(data, name == "F")
11
12    ggplot(data_f, aes(x=year, y=n)) +
13      geom_line( aes(color=name), color="#69b3a2", size=1.0 )+
14      scale_color_viridis(discrete = TRUE) +
15        scale_y_continuous(labels = scales::label_percent()) +
16        theme(
17          legend.position="none",
18          plot.title = element_text(size=14),
19          panel.grid = element_blank(),
20            axis.title = element_text(size=10),
21            panel.grid.major.x = element_blank(),
22            panel.grid.major.y = element_line( size=.1, color="black" ),
23            panel.grid.minor = element_blank(),
24            panel.background = element_blank()
25        ) +labs(y= "Taxa de Default", x = "")

```

B.3 - Matrizes de correlações – Coeficientes de Pearson, Kendall e Spearman

```

1  ##### Libraries e carregar base de dados
2  library(readxl)
3  library(tidyverse)
4  base_default_copula <-read_excel(here("base_default.xlsx"))
5
6  ##### Matriz de correlacao
7  cor(x, y, method = c("pearson", "kendall", "spearman"))
8
9  #Pearson
10 #A
11 cor(base_default_copula$A, base_default_copula$B, method = c("pearson"))
12 cor(base_default_copula$A, base_default_copula$C, method = c("pearson"))
13 cor(base_default_copula$A, base_default_copula$D, method = c("pearson"))
14 cor(base_default_copula$A, base_default_copula$E, method = c("pearson"))
15 cor(base_default_copula$A, base_default_copula$F, method = c("pearson"))
16 cor(base_default_copula$A, base_default_copula$G, method = c("pearson"))
17 cor(base_default_copula$A, base_default_copula$SFN, method = c("pearson"))
18
19 #B
20 cor(base_default_copula$B, base_default_copula$B, method = c("pearson"))
21 cor(base_default_copula$B, base_default_copula$C, method = c("pearson"))
22 cor(base_default_copula$B, base_default_copula$D, method = c("pearson"))
23 cor(base_default_copula$B, base_default_copula$E, method = c("pearson"))
24 cor(base_default_copula$B, base_default_copula$F, method = c("pearson"))
25 cor(base_default_copula$B, base_default_copula$G, method = c("pearson"))
26 cor(base_default_copula$B, base_default_copula$SFN, method = c("pearson"))

```

```

27
28 #C
29 cor(base_default_copula$C, base_default_copula$C, method = c("pearson"))
30 cor(base_default_copula$C, base_default_copula$D, method = c("pearson"))
31 cor(base_default_copula$C, base_default_copula$E, method = c("pearson"))
32 cor(base_default_copula$C, base_default_copula$F, method = c("pearson"))
33 cor(base_default_copula$C, base_default_copula$G, method = c("pearson"))
34 cor(base_default_copula$C, base_default_copula$SFN, method = c("pearson"))
35
36 #D
37 cor(base_default_copula$D, base_default_copula$D, method = c("pearson"))
38 cor(base_default_copula$D, base_default_copula$E, method = c("pearson"))
39 cor(base_default_copula$D, base_default_copula$F, method = c("pearson"))
40 cor(base_default_copula$D, base_default_copula$G, method = c("pearson"))
41 cor(base_default_copula$D, base_default_copula$SFN, method = c("pearson"))
42
43 #E
44 cor(base_default_copula$E, base_default_copula$E, method = c("pearson"))
45 cor(base_default_copula$E, base_default_copula$F, method = c("pearson"))
46 cor(base_default_copula$E, base_default_copula$G, method = c("pearson"))
47 cor(base_default_copula$E, base_default_copula$SFN, method = c("pearson"))
48
49 #F
50 cor(base_default_copula$F, base_default_copula$F, method = c("pearson"))
51 cor(base_default_copula$F, base_default_copula$G, method = c("pearson"))
52 cor(base_default_copula$F, base_default_copula$SFN, method = c("pearson"))
53
54 #G
55 cor(base_default_copula$G, base_default_copula$G, method = c("pearson"))
56 cor(base_default_copula$G, base_default_copula$SFN, method = c("pearson"))
57
58 #Kendall
59
60 #A
61 cor(base_default_copula$A, base_default_copula$B, method = c("kendall"))
62 cor(base_default_copula$A, base_default_copula$C, method = c("kendall"))
63 cor(base_default_copula$A, base_default_copula$D, method = c("kendall"))
64 cor(base_default_copula$A, base_default_copula$E, method = c("kendall"))
65 cor(base_default_copula$A, base_default_copula$F, method = c("kendall"))
66 cor(base_default_copula$A, base_default_copula$G, method = c("kendall"))
67 cor(base_default_copula$A, base_default_copula$SFN, method = c("kendall"))
68
69 #B
70 cor(base_default_copula$B, base_default_copula$B, method = c("kendall"))
71 cor(base_default_copula$B, base_default_copula$C, method = c("kendall"))
72 cor(base_default_copula$B, base_default_copula$D, method = c("kendall"))
73 cor(base_default_copula$B, base_default_copula$E, method = c("kendall"))
74 cor(base_default_copula$B, base_default_copula$F, method = c("kendall"))
75 cor(base_default_copula$B, base_default_copula$G, method = c("kendall"))
76 cor(base_default_copula$B, base_default_copula$SFN, method = c("kendall"))
77

```

```

78 #C
79 cor(base_default_copula$C, base_default_copula$C, method = c("kendall"))
80 cor(base_default_copula$C, base_default_copula$D, method = c("kendall"))
81 cor(base_default_copula$C, base_default_copula$E, method = c("kendall"))
82 cor(base_default_copula$C, base_default_copula$F, method = c("kendall"))
83 cor(base_default_copula$C, base_default_copula$G, method = c("kendall"))
84 cor(base_default_copula$C, base_default_copula$SFN, method = c("kendall"))
85
86 #D
87 cor(base_default_copula$D, base_default_copula$D, method = c("kendall"))
88 cor(base_default_copula$D, base_default_copula$E, method = c("kendall"))
89 cor(base_default_copula$D, base_default_copula$F, method = c("kendall"))
90 cor(base_default_copula$D, base_default_copula$G, method = c("kendall"))
91 cor(base_default_copula$D, base_default_copula$SFN, method = c("kendall"))
92
93 #E
94 cor(base_default_copula$E, base_default_copula$E, method = c("kendall"))
95 cor(base_default_copula$E, base_default_copula$F, method = c("kendall"))
96 cor(base_default_copula$E, base_default_copula$G, method = c("kendall"))
97 cor(base_default_copula$E, base_default_copula$SFN, method = c("kendall"))
98
99 #F
100 cor(base_default_copula$F, base_default_copula$F, method = c("kendall"))
101 cor(base_default_copula$F, base_default_copula$G, method = c("kendall"))
102 cor(base_default_copula$F, base_default_copula$SFN, method = c("kendall"))
103
104 #G
105 cor(base_default_copula$G, base_default_copula$G, method = c("kendall"))
106 cor(base_default_copula$G, base_default_copula$SFN, method = c("kendall"))
107
108 #Spearman
109
110 #A
111 cor(base_default_copula$A, base_default_copula$B, method = c("spearman"))
112 cor(base_default_copula$A, base_default_copula$C, method = c("spearman"))
113 cor(base_default_copula$A, base_default_copula$D, method = c("spearman"))
114 cor(base_default_copula$A, base_default_copula$E, method = c("spearman"))
115 cor(base_default_copula$A, base_default_copula$F, method = c("spearman"))
116 cor(base_default_copula$A, base_default_copula$G, method = c("spearman"))
117 cor(base_default_copula$A, base_default_copula$SFN, method = c("spearman"))
118
119 #B
120 cor(base_default_copula$B, base_default_copula$B, method = c("spearman"))
121 cor(base_default_copula$B, base_default_copula$C, method = c("spearman"))
122 cor(base_default_copula$B, base_default_copula$D, method = c("spearman"))
123 cor(base_default_copula$B, base_default_copula$E, method = c("spearman"))
124 cor(base_default_copula$B, base_default_copula$F, method = c("spearman"))
125 cor(base_default_copula$B, base_default_copula$G, method = c("spearman"))
126 cor(base_default_copula$B, base_default_copula$SFN, method = c("spearman"))
127
128 #C

```

```

129 cor(base_default_copula$C, base_default_copula$C, method = c("spearman"))
130 cor(base_default_copula$C, base_default_copula$D, method = c("spearman"))
131 cor(base_default_copula$C, base_default_copula$E, method = c("spearman"))
132 cor(base_default_copula$C, base_default_copula$F, method = c("spearman"))
133 cor(base_default_copula$C, base_default_copula$G, method = c("spearman"))
134 cor(base_default_copula$C, base_default_copula$SFN, method = c("spearman"))
135
136 #D
137 cor(base_default_copula$D, base_default_copula$D, method = c("spearman"))
138 cor(base_default_copula$D, base_default_copula$E, method = c("spearman"))
139 cor(base_default_copula$D, base_default_copula$F, method = c("spearman"))
140 cor(base_default_copula$D, base_default_copula$G, method = c("spearman"))
141 cor(base_default_copula$D, base_default_copula$SFN, method = c("spearman"))
142
143 #E
144 cor(base_default_copula$E, base_default_copula$E, method = c("spearman"))
145 cor(base_default_copula$E, base_default_copula$F, method = c("spearman"))
146 cor(base_default_copula$E, base_default_copula$G, method = c("spearman"))
147 cor(base_default_copula$E, base_default_copula$SFN, method = c("spearman"))
148
149 #F
150 cor(base_default_copula$F, base_default_copula$F, method = c("spearman"))
151 cor(base_default_copula$F, base_default_copula$G, method = c("spearman"))
152 cor(base_default_copula$F, base_default_copula$SFN, method = c("spearman"))
153
154 #G
155 cor(base_default_copula$G, base_default_copula$G, method = c("spearman"))
156 cor(base_default_copula$G, base_default_copula$SFN, method = c("spearman"))

```

B.4 - Figura 3. Taxa de default mensal - Linhas de Crédito Rural e SFN

```

1 ##### Libraries e carregar base de dados
2 library(readxl)
3 library(tidyverse)
4 library(hrbrthemes)
5 library(kableExtra)           options(knitr.table.format = "html")
6 library(babynames)
7 library(viridis)
8 library(DT)
9 library(plotly)
10
11 data <- read_excel(here("base_default_grafico.xlsx"))
12
13 ##### Gráfico das PD de Default com SFN
14 tmp <- data %>%
15   mutate(name2=name)
16 tmp %>%

```

```

17 ggplot( aes(x=year, y=n)) +
18   geom_line( data=tmp %>% dplyr::select(-name), aes(group=name2),
19   ↪ color="grey", size=0.6, alpha=0.6) +
20   geom_line( aes(color=name), color="#69b3a2", size=1.5) +
21   scale_color_viridis(discrete = TRUE) +
22   scale_y_continuous(labels = scales::label_percent(accuracy = 1)) +
23   theme(
24     legend.position="none",
25     plot.title = element_text(size=14),
26     panel.grid = element_blank(),
27     axis.title = element_text(size=14),
28     panel.grid.major.x = element_blank(),
29     panel.grid.major.y = element_line( size=.1, color="black" ),
30     panel.grid.minor = element_blank(),
31     panel.background = element_blank()
32   ) +
33   labs(y= "Taxa de Default", x = "")+
34   facet_wrap(~name, ncol = 2)

```

B.5 - Assimetria, Curtose e Teste de Normalidade

```

1 ##### Libraries e carregar base de dados
2 library(readxl)
3 library(tidyverse)
4 library(formattable)
5 library(moments)
6 library(tseries)
7
8 base_default_sfn <- read_excel(here("base_default_sfn.xlsx"))
9 base_default_sfn <- group_by(base_default_sfn, nm_prd)
10
11 ##### Assimetria e Curtose
12
13 descritiva <- summarise(base_default_sfn,
14   mean_pd= percent(mean(pd_default)),
15   sd_pd = percent(sd(pd_default)),
16   cv_pd = (sd(pd_default))/(mean(pd_default)),
17   skew_pd = skewness(pd_default),
18   kurt_pd = kurtosis(pd_default))
19
20 ##### Teste de Normalidade: Shapiro-Wilk e Jarque-Bera
21 A <-filter(base_default_sfn, nm_prd == "PRONAF AGRICULTURA FAMILIAR")
22 B <-filter(base_default_sfn, nm_prd == "PRONAMP")
23 C <-filter(base_default_sfn, nm_prd == "CUSTEIO AGROPECUARIO")
24 D <-filter(base_default_sfn, nm_prd == "GIRO")
25 E <-filter(base_default_sfn, nm_prd == "INVESTIMENTO AGROPECUARIO" )
26 F <-filter(base_default_sfn, nm_prd == "AQUISICAO ATIVOS FINANCEIROS" )
27 G <-filter(base_default_sfn, nm_prd ==
28   ↪ "AGRONEGOCIOS-FIN. GARANTIA DE PREÇOS E ESTOCAGEM")

```

```

28
29 SFN <-filter(base_default_sfn, nm_prd == "SFN")
30
31 shapiro.test(A$pd_default)
32 jarque.bera.test(A$pd_default)
33 shapiro.test(B$pd_default)
34 jarque.bera.test(B$pd_default)
35 shapiro.test(C$pd_default)
36 jarque.bera.test(C$pd_default)
37 shapiro.test(D$pd_default)
38 jarque.bera.test(D$pd_default)
39 shapiro.test(E$pd_default)
40 jarque.bera.test(E$pd_default)
41 shapiro.test(F$pd_default)
42 jarque.bera.test(F$pd_default)
43 shapiro.test(G$pd_default)
44 jarque.bera.test(G$pd_default)
45 shapiro.test(SFN$pd_default)
46 jarque.bera.test(SFN$pd_default)

```

B.6 - Figura 4. Representação conjunta das taxas de default das linhas de crédito rural avaliadas

```

1 ##### Libraries e carregar base de dados
2 library(readxl)
3 library(tidyverse)
4 library(gridExtra)
5 base_default_copula <- read_excel(here("base_default_copula.xlsx"))
6
7 ##### Selecionar os pares de dados
8 AxB <-select(base_default_copula, A,B)
9 AxC <-select(base_default_copula, A,C)
10 AxD <-select(base_default_copula, A,D)
11 AxE <-select(base_default_copula, A,E)
12 AxF <-select(base_default_copula, A,F)
13 AxG <-select(base_default_copula, A,G)
14 BxE <-select(base_default_copula, B,E)
15
16
17 ##### AxB #####
18 plot1 <- ggplot(AxB,aes(x=A, y=B))+
19 ggtitle("Taxa de Default - Linhas A e B") +
20 geom_point(color="#69b3a2")+
21 labs(y= "Taxa de Default - B", x = "Taxa de Default - A")+
22 scale_y_continuous(labels = scales::percent)+
23 scale_x_continuous(labels = scales::percent)+
24 theme(
25 plot.title = element_text(hjust = 0.5, size = 12),

```

```

26 axis.title = element_text(size=10),
27 axis.line=element_line(),
28   panel.grid.major.x = element_blank(),
29     panel.grid.major.y = element_blank(),
30     panel.grid.minor = element_blank(),
31     panel.background = element_blank()
32
33 ##### AxC #####
34 plot2 <- ggplot(AxC,aes(x=A, y=C))+
35 ggtitle("Taxa de Default - Linhas A e C") +
36 geom_point(color="#69b3a2")+
37 labs(y= "Taxa de Default - C", x = "Taxa de Default - A")+
38 scale_y_continuous(labels = scales::percent)+
39 scale_x_continuous(labels = scales::percent)+
40 theme(
41 plot.title = element_text(hjust = 0.5, size = 12),
42 axis.title = element_text(size=10),
43 axis.line=element_line(),
44   panel.grid.major.x = element_blank(),
45     panel.grid.major.y = element_blank(),
46     panel.grid.minor = element_blank(),
47     panel.background = element_blank()
48
49 ##### AxD #####
50 plot3 <-ggplot(AxD,aes(x=A, y=D))+
51 ggtitle("Taxa de Default - Linhas A e D") +
52 geom_point(color="#69b3a2")+
53 labs(y= "Taxa de Default - D", x = "Taxa de Default - A")+
54 scale_y_continuous(labels = scales::percent)+
55 scale_x_continuous(labels = scales::percent)+
56 theme(
57 plot.title = element_text(hjust = 0.5, size = 12),
58 axis.title = element_text(size=10),
59 axis.line=element_line(),
60   panel.grid.major.x = element_blank(),
61     panel.grid.major.y = element_blank(),
62     panel.grid.minor = element_blank(),
63     panel.background = element_blank()
64
65 ##### AxE #####
66 plot4 <-ggplot(AxE,aes(x=A, y=E))+
67 ggtitle("Taxa de Default - Linhas A e E") +
68 geom_point(color="#69b3a2")+
69 labs(y= "Taxa de Default - E", x = "Taxa de Default - A")+
70 scale_y_continuous(labels = scales::percent)+
71 scale_x_continuous(labels = scales::percent)+
72 theme(
73 plot.title = element_text(hjust = 0.5, size = 12),
74 axis.title = element_text(size=10),
75 axis.line=element_line(),
76   panel.grid.major.x = element_blank(),

```

```

77     panel.grid.major.y = element_blank(),
78     panel.grid.minor = element_blank(),
79     panel.background = element_blank()
80
81 ##### AxF #####
82 plot5 <-ggplot(AxF,aes(x=A, y=F))+
83 ggtitle("Taxa de Default - Linhas A e F*") +
84 geom_point(color="#69b3a2")+
85 labs(y= "Taxa de Default - F*", x = "Taxa de Default - A")+
86 scale_y_continuous(labels = scales::percent)+
87 scale_x_continuous(labels = scales::percent)+
88 theme(
89 plot.title = element_text(hjust = 0.5, size = 12),
90 axis.title = element_text(size=10),
91 axis.line=element_line(),
92     panel.grid.major.x = element_blank(),
93         panel.grid.major.y = element_blank(),
94         panel.grid.minor = element_blank(),
95         panel.background = element_blank())
96
97 ##### AxG #####
98 plot6 <-ggplot(AxG,aes(x=A, y=G))+
99 ggtitle("Taxa de Default - Linhas A e G") +
100 geom_point(color="#69b3a2")+
101 labs(y= "Taxa de Default - G", x = "Taxa de Default - A")+
102 scale_y_continuous(labels = scales::percent)+
103 scale_x_continuous(labels = scales::percent)+
104 theme(
105 plot.title = element_text(hjust = 0.5, size = 12),
106 axis.title = element_text(size=10),
107 axis.line=element_line(),
108     panel.grid.major.x = element_blank(),
109         panel.grid.major.y = element_blank(),
110         panel.grid.minor = element_blank(),
111         panel.background = element_blank())
112
113 ##### BxE #####
114 plot7 <-ggplot(BxE,aes(x=B, y=E))+
115 ggtitle("Taxa de Default - Linhas B e E") +
116 geom_point(color="#69b3a2")+
117 labs(y= "Taxa de Default - E", x = "Taxa de Default - B")+
118 scale_y_continuous(labels = scales::percent)+
119 scale_x_continuous(labels = scales::percent)+
120 theme(
121 plot.title = element_text(hjust = 0.5, size = 12),
122 axis.title = element_text(size=10),
123 axis.line=element_line(),
124 axis.line.x.top=element_line(),
125     panel.grid.major.x = element_blank(),
126         panel.grid.major.y = element_blank(),
127         panel.grid.minor = element_blank(),

```



```

128     panel.background = element_blank()
129
130 grid.arrange(plot1, plot2,ncol=2)
131 grid.arrange(plot3, plot4,ncol=2)
132 grid.arrange(plot5, plot6,ncol=2)
133 grid.arrange(plot7,ncol=2)

```

B.7 - Teste Goodness-of-Fit Híbrido e validacao por MME

```

1  ##### Libraries e carregar base de dados
2  library(copula)
3  library(gofCopula)
4  library(tidyverse)
5  library(readxl)
6  base_default_copula <- read_excel(here("base_default_copula.xlsx"))
7
8  ##### Selecionar os pares de dados
9  AxB <-select(base_default_copula, A,B)
10 AxC <-select(base_default_copula, A,C)
11 AxD <-select(base_default_copula, A,D)
12 AxE <-select(base_default_copula, A,E)
13 AxF <-select(base_default_copula, A,F)
14 AxG <-select(base_default_copula, A,G)
15 BxE <-select(base_default_copula, B,E)
16
17 ##### Realizar o teste GoF Híbrido para o par AxB (reprocessado para todos os
18 ↪ pares de dados)
19
20 #Grupo 1 - Copulas Elipticas
21 cop_1 = gof(x = AxB, M = 1000, MJ = 1000, processes = 7, seed.active = 10,
22 copula = c("normal", "t", "plackett"))
23
24 #Grupo 2 - Copulas Arquimedianas
25 cop_2 = gof(x = AxB, M = 1000, MJ = 1000, processes = 7, seed.active = 10,
26 copula = c("frank", "clayton", "gumbel", "joe"))
27
28 #Grupo 3 - Copulas de Valores Extremos
29 cop_3 = gof(x = AxB, M = 1000, MJ = 1000, processes = 7, seed.active = 10,
30 copula = c("galambos", "huslerReiss", "tev"))
31
32 #Grupo 4 - Metodos Cópulas Vencedoras
33 cop_4 = gof(x = AxB, M = 1000, MJ = 1000, processes = 7, seed.active = 10,
34 copula = c("t", "clayton", "tev"))
35
36 #Plotagem dos graficos
37 plot(cop_1, hybrid = c(1,2,4,6,11),
38 inf.f.col = ("#69b3a2"),
39 main = "Grupo 1 - Cópulas Elípticas e Plackett",
40 xlab = "Cópulas",

```

```

40     cex.axis = 1.2,
41     cex.lab = 1.4)
42
43 plot(cop_2, hybrid = c(1,3,5,10,16),
44      inf.f.col = ("#69b3a2"),
45      main = "Grupo 2 - Cópulas Arquimedianas",
46      xlab = "Cópulas",
47      cex.axis = 1.2,
48      cex.lab = 1.4)
49
50 plot(cop_3,
51      inf.f.col = ("#69b3a2"),
52      main = "Grupo 3 - Cópulas de Valores Extremos",
53      xlab = "Cópulas",
54      cex.axis = 1.2,
55      cex.lab = 1.4)
56
57 plot(cop_4,
58      inf.f.col = ("#69b3a2"),
59      main = "Grupo 4 - Cópulas t, Clayton e t-ev",
60      xlab = "Cópulas",
61      cex.axis = 1.2,
62      cex.lab = 1.4)
63
64 #####AxB - Validação do Best Fit por Copula Information Criterion (CIF)
65
66 ##### Estimação dos Parâmetros MPL itau e irho
67 ##### Estimação dos Parâmetros itau
68 txvMME_itau = array(data=NA,dim=10)
69 txvMME_itau[1] = xvcopula(normalCopula(),x=AxB,method="itau")
70 txvMME_itau[2] = xvcopula(tCopula(dispsr = "un"),x=AxB,method="itau.mpl")
71 txvMME_itau[3] = xvcopula(plackettCopula(),x=AxB,method="itau")
72 txvMME_itau[4] = xvcopula(francCopula(),x=AxB,method="itau")
73 txvMME_itau[5] = xvcopula(claytonCopula(),x=AxB,method="itau")
74 txvMME_itau[6] = xvcopula(gumbelCopula(),x=AxB,method="itau")
75 txvMME_itau[7] = xvcopula(joeCopula(),x=AxB,method="itau")
76 txvMME_itau[8] = xvcopula(galambosCopula(),x=AxB,method="itau")
77 txvMME_itau[9] = xvcopula(huslerReissCopula(),x=AxB,method="itau")
78 txvMME_itau[10] = xvcopula(tevCopula(),x=AxB,method="itau")
79
80 y <- as.matrix(txvMME_itau)
81
82 ##### Estimação dos Parâmetros irho
83 txvMME_irho = array(data=NA,dim=10)
84 txvMME_irho[1] = xvcopula(normalCopula(),x=AxB,method="irho")
85 txvMME_irho[2] = xvcopula(tCopula(dispsr = "un"),x=AxB,method="irho.mpl")
86 txvMME_irho[3] = xvcopula(plackettCopula(),x=AxB,method="irho")
87 txvMME_irho[4] = xvcopula(francCopula(),x=AxB,method="irho")
88 txvMME_irho[5] = xvcopula(claytonCopula(),x=AxB,method="irho")
89 txvMME_irho[6] = xvcopula(gumbelCopula(),x=AxB,method="irho")
90 txvMME_irho[7] = xvcopula(joeCopula(),x=AxB,method="irho")

```

```

91 txvMME_irho[8] = xvcopula(galambosCopula(),x=AxB,method="irho")
92 txvMME_irho[9] = xvcopula(huslerReissCopula(),x=AxB,method="irho")
93 txvMME_irho[10] = xvcopula(tevCopula(),x=AxB,method="irho")
94
95 z <- as.matrix(txvMME_irho)
96
97 #####Construção Tabela Estatísticas
98 gofOutputHybrid(cop_1,nsets = 11)
99 gofOutputHybrid(cop_2,nsets = 16)
100 gofOutputHybrid(cop_3,nsets = 5)

```

B.8 - Figura 6. Densidade de probabilidade conjunta das cópulas de maior aderência aos pares de PD

```

1 ##### Libraries e carregar base de dados
2 library(copula)
3 library(gofCopula)
4 library(readxl)
5 library(tidyverse)
6 library(cowplot)
7 base_default_copula <- read_excel(here("base_default_copula.xlsx"))
8
9 ##### AxB
10 AxB <-select(base_default_copula, A,B)
11 AxB <- as.matrix(AxB)
12 rho_copula <- 2.88735776524029
13
14 wire_AxB <-wireframe2(claytonCopula(dim=2,rho_copula),FUN=dCopula,
15 xlab="A",ylab="B",zlab="d",main="PD(AxB): Clayton (2.88)", col="#69b3a2")
16
17 ##### AxC
18 AxC <-select(base_default_copula, A,C)
19 AxC <- as.matrix(AxC)
20 rho_copula <- 0.577607454445608
21
22 wire_AxC <-wireframe2(claytonCopula(dim=2,rho_copula),FUN=dCopula,
23 xlab="A",ylab="C",zlab="d",main="PD(AxC): Clayton (0.57)", col="royalblue2")
24
25 ##### AxD Clayton#####
26 AxD <-select(base_default_copula, A,D)
27 AxD <- as.matrix(AxD)
28 rho_copula <- -0.073728693599996
29
30 wire_AxD_clayton <-wireframe2(claytonCopula(dim=2,rho_copula),FUN=dCopula,
31 xlab="A",ylab="D",zlab="d",main="PD(AxD): Clayton (-0.07)", col="purple")
32
33 ##### AxD t-ev
34 AxD <-select(base_default_copula, A,D)

```

```

35 AxD <- as.matrix(AxD)
36 rho_copula <- -0.639638960037685
37
38 wire_AxD_tev <-wireframe2(tevCopula(rho_copula,df =
  ↳ 16.7033424450924),FUN=dCopula,
39 xlab="A",ylab="D",zlab="d",main="PD(AxD): t-EV (-0,63)", col="plum")
40
41 #####AxE
42 AxE <-select(base_default_copula, A,E)
43 AxE <- as.matrix(AxE)
44 rho_copula <- 1.65649312053386
45
46 wire_AxE <-wireframe2(claytonCopula(dim=2,rho_copula),FUN=dCopula,
47 xlab="A",ylab="E",zlab="d",main="PD(AxE): Clayton (1.65)", col="cadetblue4")
48
49 ##### AxF
50 AxF <-select(base_default_copula, A,F)
51 AxF <- as.matrix(AxF)
52 rho_copula <- 0.835224159607038
53
54 wire_AxF <-wireframe2(tevCopula(rho_copula,df = 32.498169313415),FUN=dCopula,
55 xlab="A",ylab="F*",zlab="d",main="PD(AxF*): t-EV (0.83)", col="darkorange2")
56
57 #####AxG t-ev
58 AxG <-select(base_default_copula, A,G)
59 AxG <- as.matrix(AxG)
60 rho_copula <- -0
61
62 wire_AxG_tev <-wireframe2(tevCopula(rho_copula,df = 4),FUN=dCopula,
63 xlab="A",ylab="G",zlab="d",main="PD(AxG): t-EV (0.00)", col="coral")
64
65 ##### AxG Gulbel
66 AxG <-select(base_default_copula, A,G)
67 AxG <- as.matrix(AxG)
68 rho_copula <- 1
69
70 wire_AxG_gumbel <-wireframe2(gumbelCopula(rho_copula),FUN=dCopula,
71 xlab="A",ylab="G",zlab="d",main="PD(AxG): Gumbel (1.00)", col="coral3")
72
73 #####BxE
74 BxE <-select(base_default_copula, B,E)
75 BxE <- as.matrix(BxE)
76 rho_copula <- 1.97530847283902
77
78 wire_BxE <-wireframe2(claytonCopula(dim=2,rho_copula),FUN=dCopula,
79 xlab="B",ylab="E",zlab="d",main="PD(BxE): Clayton (1.97)", col="deeppink")
80
81 #####Plot dos graficos
82
83 plot_grid(wire_AxB,wire_AxC,wire_AxD_clayton, ncol=3)
84

```

```

85 plot_grid(wire_AxD_tev,wire_AxE,wire_AxF, ncol=3)
86
87 plot_grid(wire_AxG_tev,wire_AxG_gumbel,wire_BxE, ncol=3)

```

B.9 - Figura 8. Função de Distribuição Acumulada – Observações Empíricas e Dados Simulados sob a Distribuição Normal e Cópulas

```

1 #####AxB
2 # 1- Estimar a Acumulada Empírica F(AxB)
3 A<-select(base_default_copula, A)
4 A <- as.matrix(A)
5 Fn<- ecdf(A)
6 A_cdf<- cbind(A,Fn(A))
7
8 B<-select(base_default_copula, B)
9 B <- as.matrix(B)
10 Fn<- ecdf(B)
11 B_cdf<- cbind(B,Fn(B))
12
13 cdf_AxB <- cbind(A_cdf,B_cdf)
14 cdf_AxB <- cdf_AxB[,-(1)]
15 cdf_AxB <- cdf_AxB[,-(2)]
16
17 # 2- GERAR A SIMULAÇÃO DA CÓPULA NORMAL E DA BEST FIT
18 # 2.1. Normal
19 set.seed(10)
20 rho_normal = 0.797410104602365
21 u <- rCopula(60,normalCopula(dim=2,rho_normal))
22
23 # 2.2. Copula
24 set.seed(10)
25 rho_copula <- 2.88735776524029
26 v <- rCopula(60,claytonCopula(dim=2,rho_copula))
27
28 cdf_AxB <- as.data.frame(cdf_AxB)
29 u <- as.data.frame(u)
30 v <- as.data.frame(v)
31
32 # 3- Gráfico AxB - Dados x Normal
33 plot_normal<-ggplot() +
34   geom_point(data = cdf_AxB, aes(x = V1, y = V2, color ="Observada")) +
35   geom_point(data = u, aes(x = V1, y = V2,color ="Simulada Normal")) +
36   scale_color_manual(values = c("snow4", "firebrick"))+
37   ggtitle("PD(AxB) - FDA Observada e Simulada Normal") +
38   labs(y= "PD(B)", x = "PD(A)",colour="")+
39   theme(
40 plot.title = element_text(hjust = 0.5, size = 12),
41 axis.title = element_text(size=10),

```

```

42 axis.line=element_line(),
43   panel.grid.major.x = element_blank(),
44     panel.grid.major.y = element_blank(),
45     panel.grid.minor = element_blank(),
46     panel.background = element_blank(),
47     legend.position="bottom")
48
49 #4- Gráfico AxB - Dados x Cópula
50 plot_copula<-ggplot() +
51   geom_point(data = cdf_AxB, aes(x = V1, y = V2, color ="Observada")) +
52   geom_point(data = v, aes(x = V1, y = V2,color ="Simulada Clayton")) +
53   scale_color_manual(values = c("snow4", "#69b3a2"))+
54   ggtitle("PD(AxB) - FDA Observada e Simulada Clayton") +
55   labs(y= "", x = "PD(A)", colour="")+
56   theme(
57 plot.title = element_text(hjust = 0.5, size = 12),
58 axis.title = element_text(size=10),
59 axis.line=element_line(),
60   panel.grid.major.x = element_blank(),
61     panel.grid.major.y = element_blank(),
62     panel.grid.minor = element_blank(),
63     panel.background = element_blank(),
64     legend.position="bottom")
65
66 grid.arrange(plot_normal, plot_copula,ncol=2)
67
68 #5- Repetir o processo para os outros pares

```

B.10 - Erro Quadrático Médio – Simulações de normalidade e via Cópulas

```

1 #####AxB
2
3 # 1- Estimar a Acumulada Empírica F(AxB)
4 A<-select(base_default_copula, A)
5 A <- as.matrix(A)
6 Fn<- ecdf(A)
7 A_cdf<- cbind(A,Fn(A))
8
9 B<-select(base_default_copula, B)
10 B <- as.matrix(B)
11 Fn<- ecdf(B)
12 B_cdf<- cbind(B,Fn(B))
13
14 cdf_AxB <- cbind(A_cdf,B_cdf)
15 cdf_AxB <- cdf_AxB[,-(1)]
16 cdf_AxB <- cdf_AxB[,-(2)]
17

```

```

18 # 2- Contar a proporção no percentil até 90%
19 # 2.1- Base Empírica
20 cdf_AxB <- mutate(cdf_AxB,
21                   "98.33" = if_else(V1 <= 0.9833 & V2 <=
22                                     ↪ 0.9833,1,0))
23 cdf_AxB <- mutate(cdf_AxB,
24                   "96.67" = if_else(V1 <= 0.9667 & V2 <=
25                                     ↪ 0.9667,1,0))
26 cdf_AxB <- mutate(cdf_AxB,
27                   "95.00" = if_else(V1 <= 0.9500 & V2 <=
28                                     ↪ 0.9500,1,0))
29 cdf_AxB <- mutate(cdf_AxB,
30                   "93.33" = if_else(V1 <= 0.9333 & V2 <= 0.9333,1,0))
31 cdf_AxB <- mutate(cdf_AxB,
32                   "91.67" = if_else(V1 <= 0.9167 & V2 <=
33                                     ↪ 0.9167,1,0))
34 cdf_AxB <- mutate(cdf_AxB,
35                   "90.00" = if_else(V1 <= 0.9000 & V2 <=
36                                     ↪ 0.9000,1,0))
37 cdf_AxB <- mutate(cdf_AxB,
38                   "88.33" = if_else(V1 <= 0.8833 & V2 <=
39                                     ↪ 0.8833,1,0))
40 cdf_AxB <- mutate(cdf_AxB,
41                   "86.67" = if_else(V1 <= 0.8667 & V2 <=
42                                     ↪ 0.8667,1,0))
43 cdf_AxB <- mutate(cdf_AxB,
44                   "85.00" = if_else(V1 <= 0.8500 & V2 <=
45                                     ↪ 0.8500,1,0))
46 percentil_cdf_AxB = array(data=NA,dim=9)
47 percentil_cdf_AxB[1] = mean(cdf_AxB$"98.33")
48 percentil_cdf_AxB[2] = mean(cdf_AxB$"96.67")
49 percentil_cdf_AxB[3] = mean(cdf_AxB$"95.00")
50 percentil_cdf_AxB[4] = mean(cdf_AxB$"93.33")
51 percentil_cdf_AxB[5] = mean(cdf_AxB$"91.67")
52 percentil_cdf_AxB[6] = mean(cdf_AxB$"90.00")
53 percentil_cdf_AxB[7] = mean(cdf_AxB$"88.33")
54 percentil_cdf_AxB[8] = mean(cdf_AxB$"86.67")
55 percentil_cdf_AxB[9] = mean(cdf_AxB$"85.00")
56
57 percentil_cdf_AxB <- as.data.frame(percentil_cdf_AxB)
58
59 #2.2- Simulação Normal
60 #Normal
61 set.seed(10)
62 rho_normal = -0.797410104602365
63 u <- rCopula(100000,normalCopula(dim=2,rho_normal))
64 u <- as.data.frame(u)
65
66 u <- mutate(u,
67             "98.33" = if_else(V1 <= 0.9833 & V2 <=
68                               ↪ 0.9833,1,0))

```

```

60 u <- mutate(u,
61           "96.67" = if_else(V1 <= 0.9667 & V2 <=
        ↪ 0.9667,1,0))
62 u <- mutate(u,
63           "95.00" = if_else(V1 <= 0.9500 & V2 <=
        ↪ 0.9500,1,0))
64 u <- mutate(u,
65           "93.33" = if_else(V1 <= 0.9333 & V2 <= 0.9333,1,0))
66 u <- mutate(u,
67           "91.67" = if_else(V1 <= 0.9167 & V2 <=
        ↪ 0.9167,1,0))
68 u <- mutate(u,
69           "90.00" = if_else(V1 <= 0.9000 & V2 <=
        ↪ 0.9000,1,0))
70 u <- mutate(u,
71           "88.33" = if_else(V1 <= 0.8833 & V2 <=
        ↪ 0.8833,1,0))
72 u <- mutate(u,
73           "86.67" = if_else(V1 <= 0.8667 & V2 <=
        ↪ 0.8667,1,0))
74 u <- mutate(u,
75           "85.00" = if_else(V1 <= 0.8500 & V2 <= 0.8500,1,0))
76
77 percentil_u = array(data=NA,dim=9)
78 percentil_u[1] = mean(u$"98.33")
79 percentil_u[2] = mean(u$"96.67")
80 percentil_u[3] = mean(u$"95.00")
81 percentil_u[4] = mean(u$"93.33")
82 percentil_u[5] = mean(u$"91.67")
83 percentil_u[6] = mean(u$"90.00")
84 percentil_u[7] = mean(u$"88.33")
85 percentil_u[8] = mean(u$"86.67")
86 percentil_u[9] = mean(u$"85.00")
87
88 percentil_u <- as.data.frame(percentil_u)
89
90 #2.3- Simulação Copula Clayton #####
91 #Clayton
92 set.seed(10)
93 rho_copula <- 2.88735776524029
94 v <- rCopula(100000,claytonCopula(dim=2,rho_copula))
95 v <- as.data.frame(v)
96 v <- mutate(v,
97           "98.33" = if_else(V1 <= 0.9833 & V2 <=
        ↪ 0.9833,1,0))
98 v <- mutate(v,
99           "96.67" = if_else(V1 <= 0.9667 & V2 <=
        ↪ 0.9667,1,0))
100 v <- mutate(v,
101           "95.00" = if_else(V1 <= 0.9500 & V2 <=
        ↪ 0.9500,1,0))

```



```

102 v <- mutate(v,
103           "93.33" = if_else(V1 <= 0.9333 & V2 <= 0.9333,1,0))
104 v <- mutate(v,
105           "91.67" = if_else(V1 <= 0.9167 & V2 <=
106             ↪ 0.9167,1,0))
107 v <- mutate(v,
108           "90.00" = if_else(V1 <= 0.9000 & V2 <=
109             ↪ 0.9000,1,0))
110 v <- mutate(v,
111           "88.33" = if_else(V1 <= 0.8833 & V2 <=
112             ↪ 0.8833,1,0))
113 v <- mutate(v,
114           "86.67" = if_else(V1 <= 0.8667 & V2 <=
115             ↪ 0.8667,1,0))
116 v <- mutate(v,
117           "85.00" = if_else(V1 <= 0.8500 & V2 <= 0.8500,1,0))
118
119 percentil_v = array(data=NA,dim=9)
120 percentil_v[1] = mean(v$"98.33")
121 percentil_v[2] = mean(v$"96.67")
122 percentil_v[3] = mean(v$"95.00")
123 percentil_v[4] = mean(v$"93.33")
124 percentil_v[5] = mean(v$"91.67")
125 percentil_v[6] = mean(v$"90.00")
126 percentil_v[7] = mean(v$"88.33")
127 percentil_v[8] = mean(v$"86.67")
128 percentil_v[9] = mean(v$"85.00")
129
130 percentil_v <- as.data.frame(percentil_v)
131
132 #3.4- Calculo o EQM dos Percentis Extremos
133 percentil = c(98.33,96.67,95.00,93.33,91.67,90.00,88.33,86.67,85.00)
134
135 high_percentil_AxB = data.frame
136 (percentil,percentil_cdf_AxB,percentil_u,percentil_v)
137
138 high_percentil_AxB <- mutate(high_percentil_AxB,
139 erro_qdr_u = ((percentil_u-percentil_cdf_AxB)^2)^(1/2))
140
141 high_percentil_AxB <- mutate(high_percentil_AxB,
142 erro_qdr_v = ((percentil_v-percentil_cdf_AxB)^2)^(1/2))
143
144 mean(high_percentil_AxB$erro_qdr_u)
145 mean(high_percentil_AxB$erro_qdr_v)
146
147 # 3- Contar a proporção em todos os percentis
148
149 # 3.1- Base Empirica
150 #Percentis 99 a 01
151 cdf_AxB <- mutate(cdf_AxB,

```

```

149         "99" = if_else(V1 <= 0.99 & V2 <=
150           ↪ 0.99,1,0))
151 cdf_AxB <- mutate(cdf_AxB,
152         "98" = if_else(V1 <= 0.98 & V2 <=
153           ↪ 0.98,1,0))
154 cdf_AxB <- mutate(cdf_AxB,
155         "97" = if_else(V1 <= 0.97 & V2 <=
156           ↪ 0.97,1,0))
157 cdf_AxB <- mutate(cdf_AxB,
158         "96" = if_else(V1 <= 0.96 & V2 <= 0.96,1,0))
159 # [...]
160 cdf_AxB <- mutate(cdf_AxB,
161         "5" = if_else(V1 <= 0.5 & V2 <= 0.5,1,0))
162 cdf_AxB <- mutate(cdf_AxB,
163         "4" = if_else(V1 <= 0.4 & V2 <= 0.4,1,0))
164 cdf_AxB <- mutate(cdf_AxB,
165         "3" = if_else(V1 <= 0.3 & V2 <= 0.3,1,0))
166 cdf_AxB <- mutate(cdf_AxB,
167         "2" = if_else(V1 <= 0.2 & V2 <= 0.2,1,0))
168 cdf_AxB <- mutate(cdf_AxB,
169         "1" = if_else(V1 <= 0.1 & V2 <= 0.1,1,0))
170
171 percentil_cdf_AxB = array(data=NA,dim=99)
172 percentil_cdf_AxB[1] = mean(cdf_AxB$"99")
173 percentil_cdf_AxB[2] = mean(cdf_AxB$"98")
174 percentil_cdf_AxB[3] = mean(cdf_AxB$"97")
175 percentil_cdf_AxB[4] = mean(cdf_AxB$"96")
176 # [...]
177 percentil_cdf_AxB[95] = mean(cdf_AxB$"5")
178 percentil_cdf_AxB[96] = mean(cdf_AxB$"4")
179 percentil_cdf_AxB[97] = mean(cdf_AxB$"3")
180 percentil_cdf_AxB[98] = mean(cdf_AxB$"2")
181 percentil_cdf_AxB[99] = mean(cdf_AxB$"1")
182
183 percentil_cdf_AxB <- as.data.frame(percentil_cdf_AxB)
184
185 3.2- Simulação Normal
186 #Normal
187 set.seed(10)
188 rho_normal = -0.797410104602365
189 u <- rCopula(100000,normalCopula(dim=2,rho_normal))
190 u <- as.data.frame(u)
191
192 #Percentis 99 a 01
193 u <- mutate(u,
194         "99" = if_else(V1 <= 0.99 & V2 <=
195           ↪ 0.99,1,0))
196 u <- mutate(u,

```

```

196         "98" = if_else(V1 <= 0.98 & V2 <=
197         ↪ 0.98,1,0))
197 u <- mutate(u,
198         "97" = if_else(V1 <= 0.97 & V2 <=
199         ↪ 0.97,1,0))
199 u <- mutate(u,
200         "96" = if_else(V1 <= 0.96 & V2 <= 0.96,1,0))
201
202 # [...]
203
204 u <- mutate(u,
205         "5" = if_else(V1 <= 0.5 & V2 <= 0.5,1,0))
206 u <- mutate(u,
207         "4" = if_else(V1 <= 0.4 & V2 <= 0.4,1,0))
208 u <- mutate(u,
209         "3" = if_else(V1 <= 0.3 & V2 <= 0.3,1,0))
210 u <- mutate(u,
211         "2" = if_else(V1 <= 0.2 & V2 <= 0.2,1,0))
212 u <- mutate(u,
213         "1" = if_else(V1 <= 0.1 & V2 <= 0.1,1,0))
214
215 percentil_u = array(data=NA,dim=99)
216 percentil_u[1] = mean(u$"99")
217 percentil_u[2] = mean(u$"98")
218 percentil_u[3] = mean(u$"97")
219 percentil_u[4] = mean(u$"96")
220
221 # [...]
222
223 percentil_u[95] = mean(u$"5")
224 percentil_u[96] = mean(u$"4")
225 percentil_u[97] = mean(u$"3")
226 percentil_u[98] = mean(u$"2")
227 percentil_u[99] = mean(u$"1")
228
229 percentil_u <- as.data.frame(percentil_u)
230
231
232 #3.3 - Simulação Copula Clayton
233 #Clayton
234 set.seed(10)
235 rho_copula <- 2.88735776524029
236 v <- rCopula(100000,claytonCopula(dim=2,rho_copula))
237 v <- as.data.frame(v)
238
239 #Percentis 99 a 01
240 v <- mutate(v,
241         "99" = if_else(V1 <= 0.99 & V2 <=
242         ↪ 0.99,1,0))
v <- mutate(v,

```

```

243     "98" = if_else(V1 <= 0.98 & V2 <=
      ↪ 0.98,1,0))
244 v <- mutate(v,
245     "97" = if_else(V1 <= 0.97 & V2 <=
      ↪ 0.97,1,0))
246 v <- mutate(v,
247     "96" = if_else(V1 <= 0.96 & V2 <= 0.96,1,0))
248
249 # [...]
250
251 v <- mutate(v,
252     "5" = if_else(V1 <= 0.5 & V2 <= 0.5,1,0))
253 v <- mutate(v,
254     "4" = if_else(V1 <= 0.4 & V2 <= 0.4,1,0))
255 v <- mutate(v,
256     "3" = if_else(V1 <= 0.3 & V2 <= 0.3,1,0))
257 v <- mutate(v,
258     "2" = if_else(V1 <= 0.2 & V2 <= 0.2,1,0))
259 v <- mutate(v,
260     "1" = if_else(V1 <= 0.1 & V2 <= 0.1,1,0))
261
262 percentil_v = array(data=NA,dim=99)
263 percentil_v[1] = mean(v$"99")
264 percentil_v[2] = mean(v$"98")
265 percentil_v[3] = mean(v$"97")
266 percentil_v[4] = mean(v$"96")
267
268 # [...]
269
270 percentil_v[95] = mean(v$"5")
271 percentil_v[96] = mean(v$"4")
272 percentil_v[97] = mean(v$"3")
273 percentil_v[98] = mean(v$"2")
274 percentil_v[99] = mean(v$"1")
275
276 percentil_v <- as.data.frame(percentil_v)
277
278 #3.4- Calculo o EQM dos Percentis de toda a distribuição (99 a 01)
279
280 percentil = c(99:1)
281 base_completa_percentil_AxB =
      ↪ data.frame(percentil,percentil_cdf_AxB,percentil_u,percentil_v)
282
283 base_completa_percentil_AxB <- mutate(base_completa_percentil_AxB,
284     erro_qdr_u = ((percentil_u-percentil_cdf_AxB)^2)^(1/2))
285
286 base_completa_percentil_AxB <- mutate(base_completa_percentil_AxB,
287     erro_qdr_v = ((percentil_v-percentil_cdf_AxB)^2)^(1/2))
288
289
290 mean(base_completa_percentil_AxB$erro_qdr_u)

```

```
291 | mean(base_completa_percentil_AxB$erro_qdr_v)
292 |
293 | ##### O Mesmo processo é repetido para os outros pares de dados.
294 |
295 | ##### Este processo efetua o cálculo do EQM em Situação de Normalidade e de
    | ↪ Copula para a Tabela 24. Ele também foi aplicado ao cálculo das tabelas 26
    | ↪ e 27, com os respectivos ajustes na base de dados.
```