



Instituto Superior de Economia e Gestão

UNIVERSIDADE TÉCNICA DE LISBOA

DESDE 1911

MESTRADO FINANÇAS

TRABALHO FINAL DE MESTRADO DISSERTAÇÃO

**ANÁLISE DE RISCO DE CRÉDITO – MODELAÇÃO
DE DISTRIBUIÇÕES BASEADA EM CÓPULAS**

TIAGO ANDRÉ CARDOSO ANTUNES

ORIENTAÇÃO:

DR. CARLOS PESTANA BARROS

2ºSEMESTRE - 2011/2012

ANÁLISE DE RISCO DE CRÉDITO –
MODELAÇÃO DE DISTRIBUIÇÕES BASEADA
EM CÓPULAS

POR TIAGO ANTUNES

O risco de crédito é definido pela relação de incerteza em receber numa data futura o valor emprestado. Devido ao crescente volume de perdas de capital assim como ao maior rigor por parte das instituições reguladoras, a gestão do risco de crédito assume actualmente um papel fundamental na actividade bancária. A teoria de Cópulas surge como um modelo matemático que permite ajustar e compreender a relação de dependência entre diferentes distribuições marginais ou de variáveis. Os resultados obtidos confirmam a existência de relações de dependência entre perda de crédito e as diferentes variáveis de risco, demonstrando a possibilidade de modelação através da sua cópula.

Palavras-chave: risco de crédito, perda de capital, Cópula, distribuições marginais

The credit risk is defined as the ratio of uncertainty in receiving at a future date the amount borrowed. Due to the increasing volume of loss of capital, as well as the greater rigor on the part of regulatory entities, the management of credit risk plays currently a key role in banking activity. The Copula's theory appears such as a mathematical model that allows you to adjust and understand the relationship of dependence between different marginal distributions or risk variables. It is therefore expected that a multivariate distribution still that composed by different marginal distributions can have their dependence relations modeled by the Copula's families that best fit to each specific case.

Keywords: credit risk, capital loss, Copula, marginal distributions

ÍNDICE

1. Introdução	5
2. Revisão de Literatura	8
2.1. Estudos Científicos abordados	8
2.2. Risco de Crédito	10
2.3. Modelos de Risco	11
2.4. Perda de Crédito	12
2.5. Teoria de Cópulas	16
2.6. Definição da Função Cópula	17
2.7. Teorema de Sklar	18
2.8. Famílias de Cópulas	19
2.9. Testes de Independência Multivariada	21
2.10. Testes de Aleatoriedade (<i>Randomness</i>)	23
2.11. Testes de Ajustamento (<i>Goodness-of-Fit</i>)	23
2.12. Medidas de Associação	24
2.13. Métodos de Estimação dos Parâmetros da Cópula	25
2.13.1. <i>Tau</i> de <i>Kendall</i>	25
2.13.2. <i>Rho</i> de <i>Spearman</i>	26
2.13.3. Estimação Baseada em Medidas de Dependência	26
2.14. Aplicação Prática	27
3. Dados	27
4. Metodologia da Pesquisa Empírica	30
5. Análise de Resultados	34
6. Discussão e Conclusão	38
7. Referências Bibliográficas	41

ÍNDICE de TABELAS

Tabela I – Distribuição de grupos de risco	29
Tabela II – Distribuição de perda de crédito	34
Tabela III – Testes de Independência entre variáveis. $H_0 = \text{TRUE}$ (Independência entre variáveis)	35
Tabela IV - Testes de ajustamento de Cópulas. $H_0 = \text{FALSE}$ (Validade do modelo de cópula em análise para a distribuição bivariada)	36
Tabela V - Estimação do parâmetro da Cópula e correspondente erro padrão.	37
Tabela VI - Testes de ajustamento de Cópulas. $H_0 = \text{FALSE}$ (Distribuição multivariada)	38

ÍNDICE de FIGURAS

Figura 1 – Série temporal de perda de crédito da amostra.	28
Figura 2 – Histograma de observações mensais de perda de crédito percentual.	28
Figura 3 - Gráfico frequências acumuladas de incumprimento por grupo de risco	29

ÍNDICE de ANEXOS

Anexo 1 – Pseudo-observações por quebra aleatória de ligações intra-distribuições bivariadas.	
Anexo 2 – Taxas crédito consumo Banco Portugal	
Anexo 3 – Exemplos de comandos utilizados – Package Copula R	

1. Introdução

As recentes alterações nos mercados financeiros globais e o elevado volume de perdas de crédito, conduziram a que os modelos de gestão de risco de crédito assumam maior preponderância, quer na atividade de gestão bancária, quer na atividade académica, surgindo a cada dia novos desenvolvimentos das técnicas de avaliação e gestão de risco, que justificam a presente investigação (Bessis, 2002).

Historicamente a gestão de risco de crédito tem-se adaptado às diferentes exigências. O que em anos transatos apresentava-se como uma análise tradicional através de profissionais especialistas, passando depois por um processo de implementação de técnicas de análise estatística que lhe conferiam maior robustez, atualmente a análise do risco de crédito é um processo de elevada complexidade com a introdução de modelos de previsão e de modelos de dependência entre distribuições marginais de diferentes variáveis, como o modelo de Cópulas. Este progresso tem vindo a ser incentivado pelos órgãos de supervisão bancária, nomeadamente, pela implementação de um modelo regulatório através dos acordos de Basileia (Bluhm *et al.*, 2003).

A gestão do risco assume-se assim com uma proposta de estruturação e implementação de processos que determinam o risco, para que seja possível evitar, reduzir ou diversificar o mesmo fenómeno. Esta gestão pode assim ser vista como uma ferramenta que não só evita ou diminui a existência de cenários de perda de valor, mas também como uma ferramenta que cria valor às instituições financeiras (Froot *et al.*, 1993).

Já foi falado na gestão de risco de crédito, contudo, é importante efetuar uma breve abordagem ao conceito em si. Desta forma, conceptualizamos o risco de crédito “através da citação de Siqueira (2000), «uma consequência da decisão livre e consciente de expor-se a uma situação na qual há a expectativa de ganho sabendo-se que há a possibilidade de perda ou dano»”. Segundo Bessis (2002), risco pode ser definido como a variância do retorno esperado, ou como o risco de emprestar e não vir a receber futuramente o montante emprestado.

Atualmente a análise ao risco de crédito segue duas grandes correntes ou modelos: a primeira, e com base num modelo já estudado e amplamente aceite pela generalidade da comunidade de analistas de risco, modelo esse assente em conceitos probabilísticos de perda, ou seja, utilizam modelos de probabilidade para estimarem o número de eventos de incumprimento

possíveis. Outra corrente, que tem vindo a ganhar espaço essencialmente no meio empresarial, assenta num valor absoluto de perdas inerentes ao risco. Neste modelo estão inseridos modelos como CreditMetrics ou o modelo proposto no trabalho, com possibilidade de determinação do valor de perda de crédito associada à relação de dependência com os fatores de risco (Crouhy *et al.* 2002).

Ao longo dos últimos tempos, as instituições financeiras têm vindo a dar maior ênfase na utilização de modelos quantitativos como suporte às decisões de crédito, em substituição dos modelos probabilísticos até aqui utilizados. O principal motivo que conduz à adoção destes modelos está associado ao interesse das instituições financeiras em metodologias mais apuradas, que se focalizem na diminuição do valor das perdas de crédito, assim como, na determinação dos montantes de capital para constituição de provisões (Bessis, 2002).

Jorion (2007) afirma que uma abordagem tradicional no que respeita à análise de risco é baseada no pressuposto de existência de normalidade das distribuições conjuntas dos parâmetros de risco. Isto significa assumir que todos os fatores potenciadores de risco seguem obrigatoriamente uma distribuição normal, o que implica a existência de uma correlação linear entre fatores, sendo a sua distribuição conjunta também normal, e uma baixa probabilidade de ocorrência de eventos extremos. No entanto, é defendido pelo mesmo autor através de estudos empíricos, que estes pressupostos são cada vez mais questionáveis e que adoptá-los pode ser o sinónimo de uma subestimação no cálculo do risco real.

Desta forma, a utilização da teoria de cópulas para o cálculo e modelação de riscos financeiros tende a ter uma crescente aplicação nos anos futuros, devido essencialmente à forma flexível com que a distribuição conjunta dos fatores de risco pode ser elaborada, a partir de distribuições marginais conhecidas (Jorion, 2007).

A teoria de Cópulas tem a particularidade de proporcionar modelos bastante flexíveis para a modelagem de dependência entre variáveis. Cópulas são assim modelos matemáticos que podem ajudar a obter a distribuição conjunta entre variáveis (Yu & Chuang, 2012). No sector financeiro, mais concretamente no interior de uma carteira de crédito ao consumo, a ocorrência de acontecimentos extremos assume uma frequência regular, o que implica necessariamente o aumento do crescimento no interesse na utilização deste tipo de modelo matemático. A dependência entre valores extremos, mais concretamente a dependência entre

valores extremos de perda de crédito, é modelada de forma adequada pelas Cópulas (Genest *et al.* 2009a).

As funções cópula apresentam-se como uma ferramenta útil nos processos de construção e simulação de distribuições conjuntas multivariadas e bivariadas. Estas são o exemplo de uma ferramenta que permite de uma forma muito aceitável a modelação de uma estrutura de dependência, de forma independente a partir das distribuições marginais das variáveis aleatórias, como são o exemplo do *Credit Scoring*, idade, região da concessão de crédito e existência de eventos *default*. Desta forma, com base na função cópula mais ajustada é possível modelar uma distribuição conjunta, criando assim, uma correta estrutura de dependência entre variáveis aleatórias, quer as mesmas sejam independentes ou dependentes, podendo este modelo apresentar-se como uma ferramenta facilitadora para um gestor de risco formular a sua estratégia operacional (Clemente & Romano, 2004).

Goorah (2007) estabelece como vantagens da teoria de cópulas o facto de serem medidas de dependência mais robustas que os coeficientes de correlação linear, e o facto de poderem ser utilizadas a partir de diferentes distribuições marginais em variáveis que são dotadas de comportamentos distintos.

Como é do conhecimento geral e defendido por Nelsen (2006), as variáveis financeiras não são independentes, desta forma, o mesmo autor concluiu que é fundamental o estudo e compreensão da relação entre as referidas variáveis, de forma a obter-se uma correta avaliação do risco de determinada carteira de crédito, sendo que o principal intuito da função cópula é a representação da relação referida anteriormente. Assim, a teoria de cópulas fica diretamente relacionada ao estudo das medidas de dependência entre variáveis aleatórias, facto justificativo da mais-valia da utilização dos modelos no presente trabalho e da tendência crescente na sua utilização.

A hipótese de pesquisa em estudo é se o risco de crédito de uma carteira pode ser avaliado e modelado a partir da metodologia baseada na teoria de cópulas, aplicada à relação de dependência entre as variáveis explicativas (fatores de risco) e o conceito de perda de crédito.

São pouco frequentes os trabalhos publicados em Portugal que abordem a modelação de fatores de risco em carteiras de crédito ao consumo, com o objetivo de diminuir a exposição às perdas de crédito oriundas de eventos de incumprimento. Este trabalho é inovador no sentido

em que ao analisar o risco de determinada carteira de um banco Português através do modelo de Cópulas, vem completar a lacuna existente.

Assim, o principal objetivo do trabalho é a construção de um modelo eficiente que descreva a estrutura de dependência entre distribuições, esperando que o modelo de Cópulas seja capaz de modelar adequadamente o risco de crédito. Outro objetivo passa por no futuro permitir ao leitor replicar e utilizar o modelo proposto na análise de qualquer *portfolio* de crédito.

A estrutura do trabalho é definida no capítulo 2 sendo apresentada a fundamentação teórica que suporta toda a aplicação empírica do trabalho. São abordados os temas dos conceitos de risco e perda de crédito e teoria de Cópulas, que compreende a explanação sobre a teoria base do conceito de Cópulas, teorema de Sklar, a identificação das famílias de cópulas e todo um conjunto de testes para modelação da dependência e avaliação da validade estatística do modelo de Cópulas utilizado.

No capítulo 3 é apresentada a descrição da amostra assim como que inclui por exemplo o cálculo do valor de perda de crédito. O capítulo 4 descreve todos os procedimentos práticos ao longo do trabalho e aplicação dos dados empíricos. Os capítulos 5 e 6 têm como finalidade a exposição dos resultados obtidos e discussão relativa aos mesmos resultados, respetivamente, sendo que neste último capítulo são feitas as analogias dos resultados com a fundamentação teórica.

2. Revisão de Literatura

Neste capítulo serão abordados os conceitos teóricos que fundamentam a elaboração do presente trabalho. Esses conceitos são focados na definição de risco e perda de crédito, e fundamentação da teoria de Cópulas, desde o seu Teorema Central, aos testes de independência entre séries e observações, cálculo de parâmetros de Cópulas e testes de validade estatística.

2.1. Estudos Científicos abordados

Vários estudos foram revistos para a construção da presente investigação. Cada um com a sua questão de investigação, com diferentes metodologias no que respeita à utilização das diversas famílias de cópulas, apresentado desta forma um conjunto muito heterogêneo de resultados.

Segundo Tops (2010), que estuda o comportamento do modelo de Cópula Gaussiano na correlação entre distribuições marginais de séries financeiras de ativos de risco, indica que assumir que os valores dos ativos de risco e a sua correlação linear conduzem à correta aferição da estrutura de dependência, é um raciocínio insuficiente e desvia a estrutura de dependência do modelo e não considera o comportamento em conjunto dos fatores de risco. O autor defende que acontecimentos de perda de capital em simultâneo são eventos extremos pelo que deve existir um modelo capaz de processar a estrutura de dependência entre estes eventos. O estudo conclui que o modelo de cópula Gaussiana é efetivamente um bom modelo de um parâmetro, contudo, afirma que no caso de modelos com mais parâmetros a família de cópulas elípticas através da cópula *t*-Student é uma escolha mais vantajosa.

De maneira idêntica ao estudo anterior, Jouanin *et al.* (2001), aborda o problema da incorporação da correlação padrão nos modelos de risco baseados em eventos geradores de perda de valor. O estudo dá seguimento ao proposto por Li (2000), que utiliza os modelos de cópulas nas distribuições conjuntas de séries de incumprimento. Para o efeito, os autores utilizam a cópula Normal para modelar a relação de eventos de incumprimento na sua função de sobrevivência. Os resultados exprimem a ideia que a escolha de cópula tem um impacto muito forte na modelagem de eventos extremos em séries financeiras.

O estudo de Li (2000) propõe a utilização de cópulas para definir uma distribuição conjunta de sobrevivência, em que o coeficiente de correlação entre os eventos que compõem essa distribuição é definido como o coeficiente de correlação entre duas distribuições de variáveis de risco, como forma de resolver o problema da correlação linear entre eventos. O autor define como função de sobrevivência, a função que define se uma determinada observação atinge o limite temporal definido sem entrar em *default*. A metodologia do estudo passa por assumir duas distribuições marginais contínuas de duas séries financeiras distintas, especificando de seguida através das funções de cópula a distribuição conjunta da função de sobrevivência. Os resultados apresentados mostram que a utilização da família de cópulas Normal é equivalente ao modelo *CreditMetrics*, o que não é de estranhar dado que este modelo na sua essência utiliza a função de cópula Normal na sua fórmula de correlação padrão, mesmo considerando que não utiliza explicitamente o conceito de cópula.

O trabalho de Clemente & Romano (2004), de certa forma assemelha-se à presente investigação. Os autores apresentam uma metodologia para medir e otimizar o risco de uma carteira de crédito, tendo em consideração a não-normalidade da distribuição de perda de

crédito, e em particular, a probabilidade de eventos múltiplos de incumprimento. Neste estudo os autores abandonam a hipótese de modelagem da estrutura de dependência Gaussiana, assumindo uma estrutura de dependência com base na cópula *t*-Student.

O artigo de investigação de Kole *et al.* (2007) propõem-se a provar que a precisão no processo de seleção do melhor modelo de cópula para a gestão do risco é extremamente importante, dado que a escolha de um modelo inapropriado pode levar à construção de portfólios com avaliação imprecisa do risco de crédito. De acordo com outros autores citados anteriormente, os resultados sugerem que a família de cópulas elípticas, cópula Normal e *t*-Student são mais eficazes quando comparadas com as cópulas arquimedianas Gaussiana e de Gumbel.

2.2. Risco de Crédito

Do ponto de vista da instituição financeira, uma operação de crédito consiste em disponibilizar determinado valor a uma entidade devedora, mediante o compromisso de pagamento em data futura. Nesta perspetiva, o crédito está relacionado com a expectativa de receber um valor num determinado período de tempo, e o risco de crédito não é mais do que, as hipóteses dessa mesma expectativa não se cumprir. O risco de crédito pode ainda ser definido em função das perdas originadas por um evento *default* ou pela degradação da qualidade de crédito, um exemplo muito específico, é o facto de ao longo do tempo determinado devedor contrair mais créditos, diminuindo desta forma a sua capacidade de cumprir com as suas responsabilidades (Bessis, 2002).

O mesmo autor propõe uma divisão do risco em três classes:

- Risco de *default* – probabilidade de ocorrer um evento de incumprimento por parte do devedor;
- Risco de exposição – decorre da incerteza ao valor futuro da operação de crédito;
- Risco de recuperação – incerteza quanto ao valor possível de recuperar em caso de incumprimento.

Na mesma linha de raciocínio o risco de *default* é sem dúvida o mais importante a ser trabalhado, dado que representa uma medida de antecipação de risco, com uma filosofia de atuação proactiva e não reativa. Quer isto dizer que do ponto de vista de previsão de perdas, uma atuação antecipada ao acontecimento de perda com recurso a dados concretos obtidos pelas análises às distribuições conjuntas, apresentará tendencialmente resultados mais

satisfatórios quando comparados com análises de risco que atuem com foco nas perspetivas de recuperação das perdas registadas.

2.3. Modelos de Risco

Os modelos de risco de *portfolio* têm como objetivos básicos, a determinação da distribuição probabilística de perda por *default* ou o valor real de perda num determinado horizonte temporal, e ainda a quantificação dos riscos marginais. São estes objetivos com as suas próprias tradições de atuação, que distinguem duas das correntes em análise de risco. Uma primeira que opta pela aferição do acontecimento de *default* através de modelos probabilísticos, que contudo não será desenvolvido dada a opção tomada para o trabalho, e um segundo que opta pela aferição do valor de perda associado com medidas de dependência entre variáveis, e que será a base estrutural para o trabalho.

Ao longo dos tempos a medição do risco de crédito tem vindo a ser desenvolvida, pelo que diversos modelos foram publicados, como exemplo, CreditMetrics, CreditRisk+ ou KMV. Se as propostas mais antigas se baseavam em *ratings* proporcionados unicamente por instituições de avaliação externas ao próprio banco, as últimas propostas já permitem aos bancos utilizar os seus próprios modelos de risco de crédito, sujeitos claro está, à supervisão das entidades reguladoras (Chang & Chen, 2009).

Ainda segundo Hamilton & James (2001), a maioria das instituições financeiras atualmente desenvolve e mantém ativos os seus modelos internos de *rating*. Esta constatação é uma realidade necessária e natural, isto porque, as instituições financeiras necessitam de lidar com empresas não-classificadas pelas agências internacionais de *rating*. Por terem acesso a informações não disponíveis às agências, os sistemas internos são capazes de responder com mestria às alterações da qualidade do crédito. Assim, tanto os *ratings* públicos como os internos assumem um papel fulcral na avaliação da dimensão global do risco de crédito de uma instituição financeira.

A dispersão das perdas de crédito é um acontecimento que depende da correlação entre os diversos fatores de risco. Devido a esta constatação, surgiu a necessidade de desenvolver um modelo que descrevesse essa relação, pelo que o CreditMetrics desenvolvido pela JP Morgan foi um dos exemplos dos vários modelos desenvolvidos, que propõem uma distribuição normal binomial que descreve as estruturas de dependência entre fatores de risco (Chang & Chen, 2009).

Para modelar a estrutura referida e com base nas estruturas do CreditMetrics, no trabalho serão incorporados os modelos de Cópulas para distribuições normais bivariadas e distribuições multivariadas. Os modelos de Cópulas são um desenvolvimento do CreditMetrics.

2.4. Perda de Crédito

Este é um subcapítulo que assume um interesse fulcral no presente trabalho, dado que define o modelo para o cálculo do valor de perda de crédito em cada operação observada. Segundo Schuermann (2004), a perda de crédito por *default* de uma operação de crédito pode ser determinada de três formas distintas:

- Com base em dados explícitos do mercado – abordagem macroeconómica;
- Com base em dados implícitos do mercado;
- Com base nos fluxos de caixa durante a vida do empréstimo;

Desde a década de 90 que diversos autores têm vindo a propor a criação de modelos que possibilitem a criação de distribuições de perdas em carteiras de crédito. As propostas consistem numa abordagem econométrica para a obtenção das distribuições de perdas. Desta forma, é relacionada a perda num determinado sector ou tipologia da carteira, com variáveis macroeconómicas e variáveis específicas do setor (Chirinko & Guill, 1991).

Os mesmos autores, com base em diferentes premissas sobre possíveis estados de natureza ou histórico de estados de natureza propõem a criação de diferentes distribuições de perdas de crédito, estados estes que podem ser interpretados como diferentes cenários de condições económicas. Assim, são estabelecidas funções onde:

$$(1) \quad l_{i,s} = \alpha[\mathbf{Y}_{i,s}] + E \quad i = 1,2,3,\dots,I_i \quad s = 1,2,\dots,S$$

em que:

- $l_{i,s}$ é a perda de crédito na carteira i no estado da natureza s ;
- $\mathbf{Y}_{i,s}$ é um vetor de variáveis que influenciam a rentabilidade da carteira;
- $\alpha [.]$ é uma função estimada econometricamente;
- E_i é um fator específico da carteira;

Os parâmetros de perda de uma carteira ($\mathbf{Y}_{i,s}$) são vistos em função de fatores macroeconómicos, por exemplo, taxas de juros, comportamento do PIB, emprego, entre outras, em que:

$$(2) \quad Y_{i,s} = \Psi(\mathbf{X}_{J,s}, \mathbf{Z}_N) \quad i = 1, 2, 3, \dots, I_i \quad s = 1, 2, \dots, S$$

Onde:

- $\mathbf{X}_{J,s}$ é a matriz que contém os fatores macroeconómicos J , que variam em função do estado da natureza s ;
- \mathbf{Z}_N é um vetor de N fatores macroeconómicos que se mantêm constantes em todos os estados da natureza;
- Ψ é uma função estimada por meios macroeconómicos;

De acordo com os autores enunciados, por cada fator macroeconómico que compõe $\mathbf{X}_{J,s}$ é especificado um conjunto de possíveis resultados com probabilidades específicas. Por conseguinte, a probabilidade de ocorrência de cada estado de natureza, que é representado por um cenário de valores das variáveis J , ou seja por uma coluna da matriz $\mathbf{X}_{J,s}$, é obtida através das probabilidades de ocorrência de cada um dos resultados dos fatores macroeconómicos considerados.

A matriz $\mathbf{X}_{J,s}$ é utilizada para gerar a perda de crédito num determinado setor ou segmento da carteira de crédito condicionada ao estado de natureza s . Os valores de perda de segmento de carteira num determinado estado de natureza são ponderados pelo peso que representam na carteira global da instituição bancária, sendo associado com o vetor das probabilidades de ocorrência dos estados da natureza, para assim gerar a perda da carteira global num determinado estado s .

Uma outra abordagem possível de realizar, bem distinta dos conceitos macroeconómicos descritos nos parágrafos anteriores, é através dos fluxos de pagamentos de prestações devidas. No entanto, para a compreensão desta abordagem é necessário aceitar que apesar do conceito original de perda de crédito advir da expressão inglesa *Loss Given Default*, que incorpora os conceitos de taxa de recuperação, capital emprestado, juros remuneratórios e custos de execução e cobrança, por simplicidade e conveniência é possível definir o conceito de perda de crédito através da aplicação direta da seguinte função:

$$(3) \quad \text{Perda de Crédito} = 1 - \frac{\text{Valor presente dos pagamentos realizados}}{\text{Valor em dívida}}$$

Tal como referido no parágrafo anterior, por uma questão de simplicidade é possível assumir que os custos de cobrança são desprezíveis e que a taxa de recuperação comporta-se como uma variável independente dos eventos de *default* ou mesmo como uma variável exógena ao

modelo proposto para a perda de crédito. Apesar de esta premissa ser discutível, a maioria das bases de dados produzidas pelas instituições financeiras não contém este tipo de informação, desta forma, é perfeitamente aceitável adotar esta premissa. É de referir no entanto que existindo disponibilidade da informação excluída no presente modelo, a mesma deverá ser incorporada (Schuermann, 2004).

Dadas as características específicas de uma carteira de crédito ao consumo, uma avaliação baseada no fluxo de caixa será certamente uma medida viável quanto à estimação da perda de crédito, essencialmente devido a propriedades como:

- Processos de recuperação mais ágeis;
- Genericamente os processos de cobrança não chegam a via judicial;
- Genericamente não há complicações ligadas à venda de carteira de créditos;
- Não são utilizadas garantias reais, ou seja, o que se recupera é efetivamente apenas o que é pago;

Devido às propriedades indicadas, o processo de crédito ao consumo afirma-se como um processo simples, no qual é possível a recolha dos dados necessários para a construção dos fluxos de caixa da carteira, que por sua vez permitem calcular o valor da perda de crédito. A teoria de Cópulas considera como condição necessária o cálculo do valor de perda de crédito decorrente de cada operação (elemento individual). A análise ao valor de perda de crédito é decomposta em duas fases distintas: uma primeira em que é calculado o valor de perda de cada unidade em cada momento n , criando assim uma série temporal de distribuição de perdas; uma segunda e a mais importante para o trabalho, em que é analisado um determinado instante, quantificando o valor de perdas de acordo com as observações da amostra, parâmetro que é associado com as variáveis capazes de modelar o risco de uma operação.

Dada uma operação de crédito n , a perda de crédito na i -ésima prestação será calculada segundo a seguinte formulação:

(4)

$$P_{n,i} = \begin{cases} \bullet 1 & , \text{ se o pagamento não for realizado} \\ \bullet 1 - \frac{VF_{n,i}}{VF_{n,i}} & , \text{ se o pagamento for realizado} \end{cases}$$

onde,

- P_{nj} é o valor de perda de crédito na i -ésima prestação da operação n ;
- VD_{ni} é o valor devido da i -ésima prestação da operação n ;
- VP_{nj} é o valor efetivamente pago da i -ésima prestação da operação n ;
- $dtpag_{ni}$ é a data de pagamento a i -ésima prestação da operação n ;
- $dtvenc_{ni}$ é a data de vencimento da i -ésima prestação da operação n ;
- d é a taxa mensal de juro;

Pela equação acima, percebemos que o conceito de perda de crédito adotado para o presente trabalho, estabelece a relação entre o valor em dívida e o valor presente dos pagamentos efetuados (à data do vencimento), o que vai de encontro à proposta teórica segundo o trabalho de Schuermann (2004).

No caso de ser realizado o somatório de todos os momentos de pagamento da operação n , é possível calcular o valor da perda de crédito da operação referida, onde P_n é a perda de crédito. O mesmo raciocínio pode ser adotado para um portfólio de crédito, bastando aplicar os somatórios às prestações de todas as operações da carteira,

$$(5) \quad P_n = \sum \left\{ 1 - \frac{\sum \left[\frac{VP_{ni}}{(1+d)^{(dtpag_{ni}-dtvenc_{ni})/30}} \right]}{VD_{ni}} \right\}$$

Outro raciocínio possível de inferir trata-se de que a média da operação de crédito x ponderada pelo respetivo valor de cada prestação, conduz-nos à perda de crédito da mesma operação, assim:

$$(6) \quad P_j = \frac{P}{\sum VD_{ni}} * (\sum (P_{ni} * VD_{ni}))$$

em que:

- P_n é a perda de crédito da operação n ;
- P_{ni} é a perda de crédito da i -ésima prestação da operação n ;
- VD_{ni} é o valor devido da i -ésima prestação da operação n ;

Através da média ponderada da perda de crédito, relacionada com as prestações das operações de determinado *portfolio* com data de vencimento num determinado período, é possível obter uma distribuição contínua das perdas de crédito ao longo do período em estudo, assim:

$$(7) \quad P_j = \frac{1}{\sum_n VD_{n_j}} \left(\sum_n (P_{n_j} \times VD_{n_j}) \right)$$

em que,

- P_j é a perda de crédito da carteira no período j ;
- P_{n_j} é a perda de crédito da operação n no período j ;
- VD_{n_j} é o valor devido da prestação da operação n no período j ;

Importa ainda referir que a taxa d representa o custo de oportunidade de investimento para a instituição financeira, ou seja, na prática representa o custo da impossibilidade de investir o valor devido e não recebido.

2.5. Teoria de Cópulas

O estudo da teoria de cópulas assim como a sua aplicação para efeitos estatísticos, probabilísticos e processos estocásticos é um fenómeno muito recente com principal incidência nas últimas duas décadas, sendo uma área com largas possibilidades de desenvolvimento. A sua aplicação em setores específicos como é o caso do setor financeiro, mais concretamente na gestão de risco de crédito é ainda mais recente, porém, têm vindo a surgir alguns estudos sobre a temática (Nelsen, 2006; Li, 2000; Schonbucher & Schubert, 2001).

De acordo com Frey *et al.* (2001), existem diversas famílias de Cópulas, como por exemplo: a família Arquimediana composta pela cópula de Clayton, Frank e Gumbel; a família de cópulas elípticas, composta por cópula Normal ou cópula *T-Student*. Assim, dependendo das distribuições e do tipo de dados a serem analisados, é fundamental que sejam utilizadas as cópulas com o melhor ajuste a cada de distribuição multivariada.

Em relação à melhor função cópula a utilizar, existem diferentes opiniões de diferentes autores, e funções que se adaptam melhor a determinados tipos de distribuição do que outros. Assim, de acordo com Li (2000) a utilização da cópula Gaussiana ou de Clayton são dois exemplos bem-sucedidos para a modelação da estrutura de dependência de uma distribuição de incumprimento num determinado horizonte temporal. Contudo, no caso de distribuições com mais do que um parâmetro surgem as cópulas elípticas em clara vantagem.

De acordo com Yu & Chuang (2012), estes modelos operam sob a premissa que não são necessárias informações sobre os pressupostos das distribuições marginais e sem as limitações

dos modelos de distribuição normal. Nelsen (2006) propôs que a correta utilização de Cópulas pode ser uma ferramenta muito contributiva para a análise da dependência entre as distribuições dos valores extremos da amostra.

As cópulas apresentam uma grande vantagem em relação a outros modelos matemáticos, que é a forma como representam uma distribuição conjunta, identificando de forma rigorosa as relações entre os valores extremos das variáveis analisadas. Oferecem uma maior flexibilidade na agregação de riscos, pois as distribuições marginais são independentes da modelação da estrutura de dependência das variáveis estudadas (Cherubini *et al.*, 2004).

De acordo com o mesmo autor, essa flexibilidade permite, por exemplo, que ao construir-se uma função de distribuição conjunta composta por duas ou mais distribuições de variáveis aleatórias, cada uma delas pode ser modelada individualmente. Em simultâneo, a dependência entre essas variáveis pode assumir diferentes estruturas, mesmo que sejam de padrão não-linear, de acordo com o tipo de cópula utilizada.

De acordo com Yu & Chuang (2012), grande parte dos estudos empíricos realizados sobre a aplicação da teoria de Cópulas em investigações financeiras, refere que os resultados indicam um padrão de dependência assimétrica entre as variáveis estudadas. Ver-se-á se no presente trabalho se ocorre a mesma tendência.

Segundo Nelsen (2006) e Cherubini *et al.* (2004), o conceito da teoria de cópulas é válido para situações multidimensionais, o que permite a sua utilização para análises com mais de dois fatores de risco, propriedade de extrema importância, dado que, numa carteira de crédito são múltiplos os fatores ou parâmetros de risco que deverão ser considerados.

2.6. Definição da Função Cópula

A cópula de um vector aleatório d -dimensional é uma função d -dimensional multivariada que caracteriza a estrutura de dependência não sendo dependente das funções marginais. A função de densidade acumulada *c.d.f.* pode ser obtida por meio da cópula $C(x)$ e das suas funções marginais, por exemplo, $F(x) = C(F_{x_1}(x_1), \dots, F_{x_d}(x_d))$. É possível mostrar que a cópula de uma distribuição é invariante em processos de transformações estritamente crescentes dos componentes do vector aleatório. Assim, para todo o conjunto de d funções estritamente crescentes f_1, \dots, f_d , a cópula de (X_1, \dots, X_d) é a mesma que a cópula de $(f_1(x_1), \dots, f_d(x_d))$. O termo estrutura de dependência refere-se às características de uma distribuição que não dependem

das marginais e ao combinar esta estrutura com as marginais é possível reconstituir inteiramente a distribuição (Nelsen, 2006).

Segundo Hamilton & James (2001), de uma forma sucinta qualquer função de distribuição contínua $F(X_1, \dots, X_n)$ de variáveis aleatórias X_1, \dots, X_n pode ser decomposta em funções individuais de distribuição marginal $F_i(X_i)$ e uma função cópula $C(U_1, \dots, U_n)$, função (9). Ainda de acordo com o mesmo autor, Cópula é uma função de distribuição conjunta para variáveis uniformes U_i , $i = 1, \dots, N$, cada uma das quais tem uma distribuição uniforme *standard*.

Cópulas são funções que associam o ponto $(F(x_i), G(y_i))$ a um valor específico da sua função conjunta $F(x, y)$, definida no intervalo $[0, 1]$. O recente entusiasmo para o uso desta abordagem de modelação encontra a sua origem num antigo mas elegante Teorema de Sklar (Yan & Kojadinovic, 2010; Nelsen, 2006; Cherubini *et al.*, 2004).

2.7. Teorema de Sklar

Segundo Frey *et al.* (2001), Nelsen (2006) e Yan & Kojadinovic (2010) o teorema referido assenta na proposição que:

- Seja F uma função de distribuição conjunta d -dimensional com margens contínuas F_1, \dots, F_d e uniformemente distribuídas em $[0, 1]$ ($U(0, 1)$). Então, existe uma única função cópula d -dimensional $C : [0, 1]^d \rightarrow [0, 1]$, de tal forma que, para todo $X \in \mathbb{R}^d$, representada pela equação:

$$(8) \quad F(x_1, \dots, x_m) = C(F_1(x_1), \dots, F_m(x_m)),$$

Inversamente, da função anterior podemos afirmar que C é uma função cópula multivariada c.d.f. com margens F_1, \dots, F_m , caso F_1, \dots, F_m sejam funções de distribuição univariadas, isto porque $U_i = F_i(X_i)$, $i = 1, \dots, n$, é uma variável aleatória uniforme, e que a função F indicada em (a) é uma função de distribuição conjunta com margens F_1, \dots, F_m , cujas distribuições marginais são uniformes no intervalo $(0, 1)$.

Do Teorema de Sklar, considerado por diversos autores como o mais importante teorema sobre cópulas devido à sua aplicabilidade prática, poderá ser inferido o corolário que: se existir uma cópula que satisfaça a condição expressa anteriormente, é possível extrair assim uma única cópula C a partir da inversão da função de distribuição multivariada F com margens contínuas F_1, \dots, F_m para qualquer $u = (u_1, \dots, u_n)$ no intervalo $[0, 1]^n$, através da expressão:

$$(9) \quad C(u_1, \dots, u_m) = F(F_1^{-1}(u_1), \dots, F_m^{-1}(u_m)),$$

em que $F_1^{-1}, \dots, F_m^{-1}$ é regra geral inversa a F_1, \dots, F_m . Podemos assim denominar C como a cópula de F , ou de qualquer vector com função de distribuição F . Num ponto de vista da análise a circunstâncias particulares de crédito, se se assume que a variável oculta X tem uma função de distribuição multivariada Gaussiana com matriz de correlação R , então a cópula de X poderá ser representada pela fórmula:

$$(10) \quad C_R^{GA}(u_1, \dots, u_m) = \Phi_R(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_m)),$$

onde Φ_R representa a função de distribuição conjunta de um vector aleatório com normal d -dimensional e matriz de correlação R , sendo Φ a função de distribuição de uma normal padrão univariada. C_R^{GA} é conhecida como Cópula Gaussiana e é a estrutura de dependência da variável latente que está na base da maioria dos modelos padrão da indústria financeira. Pelo mesmo teorema é possível verificar que o método de inversão é um dos principais métodos para a criação de cópulas.

A densidade da função F pode ser descrita pela expressão abaixo,

$$(11) \quad f(x_1, \dots, x_n) = c(F_1(x_1), \dots, F_n(x_n)) \prod_{i=1}^n f_i(x_i)$$

em que $c(\dots)$ e $f_i(\dots)$ são as densidades da cópula e marginais, respetivamente.

Importa referir que o Teorema de Sklar oferece-nos uma afirmação muito valiosa, ou seja, a relação de dependência entre duas ou mais funções de distribuição marginal e a sua função de distribuição conjunta é completamente explicada pela sua cópula. Mais ainda, nos casos de funções de distribuição multivariadas contínuas, as funções marginais e a estrutura de dependência podem ser separadas (Nelsen, 2006; Cherubini *et al.*, 2004).

2.8. FAMILIAS DE COPULAS

Neste subcapítulo são explicadas quatro famílias de Cópulas. Duas que pertencem à classe de cópulas arquimedianas (Cópula de Gumbel e Clayton), e outras duas que pertencem à família de cópulas elípticas (Normal, ainda que com diferentes graus de liberdade, e t -Student) e são construídas a partir de distribuições multivariadas pelos métodos de inversão.

De acordo com Yan (2007) e Nelsen (2006), uma cópula arquimedianas é construída através de uma função geradora ϕ , sendo:

$$(12) \quad C(u_1, \dots, u_p) = \Phi^{-1} [\Phi(u_1) + \dots + \Phi(u_p)]$$

onde Φ^{-1} é a função inversa da variável geradora Φ . Cada variável Φ determina uma cópula arquimedianas específica.

Segundo Yu & Chuang (2012), no início da década de 60, Gumbel propôs o seu modelo de Cópulas como um modelo para valores extremos. Provavelmente a cópula de Gumbel é a cópula mais utilizada nestes tipos de distribuições. A função de dependência apresenta o parâmetro δ de Gumbel na seguinte expressão:

$$(13) \quad A(t) = (t^\delta + (1-t)^\delta)^{1/\delta}, \delta \geq 1$$

em que quando $\delta = 1$, não existe dependência, e quando $\delta = +\infty$, estamos perante dependência perfeita.

A Cópula de Clayton apresenta uma estrutura assimétrica, com dependência extrema na cauda inferior da distribuição. É obtida através da função geradora:

$$(14) \quad \phi(x) = \frac{1}{\alpha}(t^{-\alpha} - 1),$$

e é dada pela expressão:

$$(15) \quad C(u, v) = \max\left([u^{-\alpha} + v^{-\alpha} - 1]^{-\frac{1}{\alpha}}, 0\right)$$

A representação de uma cópula elíptica é uma função de uma distribuição elíptica proposta pelo Teorema de Sklar. De acordo com Yan (2007), é possível estabelecer uma cópula elíptica determinada a partir da função de densidade acumulada F , sendo a sua função já deduzida:

$$(16) \quad C(u_1, \dots, u_p) = F [F_1^{-1}(u_1), \dots, F_p^{-1}(u_p)]$$

Entenda-se que as cópulas elípticas tornaram-se muito populares em áreas como as finanças e análise de risco devido à facilidade da sua implementação. A possibilidade de obtenção de distribuições condicionais é outras das vantagens apresentadas por esta família de cópulas no que respeita à previsão de riscos (Frees & Wang, 2005).

Uma distribuição multivariada de t -Student apresenta um tipo de cópula denominada por analogia por Cópula de t -Student. Se \mathbf{X} for um vector com uma distribuição T-Student n -

variável com ν graus de liberdade e matriz de covariância $[V/(\nu-2)] \cdot R$ (para $\nu > 2$ caso não se confirme a condição, a matriz de covariância não será definida), representado por:

$$(17) \quad \mathbf{X} = \frac{\sqrt{\nu}}{\sqrt{S}} \mathbf{Y}$$

onde, $S \sim \chi_{\nu}^2$ e o vector aleatório $\mathbf{Y} \sim N_n(0, R)$ são independentes. Assim, a cópula do vector \mathbf{Y} é uma cópula T-Student com ν graus de liberdade. Pode ser representada analiticamente por:

$$(18) \quad C_{\nu, \mathbf{R}}^t(\mathbf{u}) = t_{\nu, \mathbf{R}}^n(t_{\nu}^{-1}(u_1), \dots, t_{\nu}^{-1}(u_n))$$

em que, $R_{ij} = \Sigma_{ij} / \sqrt{\Sigma_{ii}\Sigma_{jj}}$ para $i, j \in \{1, \dots, n\}$ e em que $t_{\nu, \mathbf{R}}^n$ representa a função de densidade multivariada c.d.f do vector $\sqrt{\nu}\mathbf{Y}/\sqrt{S}$, em que a função variável $S \sim \chi_{\nu}^2$ e o vector \mathbf{Y} são independentes. t_{ν} representa as margens de $t_{\nu, \mathbf{R}}^n$.

De acordo com Bouye *et al.* (2000), a partir da equação a seguir indicada podemos obter a densidade da Cópula T,

$$(19) \quad c(u_1, \dots, u_n; \mathbf{R}, \nu) = \frac{\Gamma((\nu+n)/2)[\Gamma(\nu/2)]^n (1 + \boldsymbol{\omega}^T \mathbf{R}^{-1} \boldsymbol{\omega})^{-(\nu+n)/2}}{|\mathbf{R}|^{1/2} \Gamma(\nu/2)[\Gamma((\nu+1)/2)]^n \prod_{i=1}^n (1 + \omega_i^2 / \nu)^{-(\nu+1)/2}}$$

Autores distintos defendem mais a utilização de determinadas famílias de cópulas em detrimento de outras. Dado um conjunto de dados, a escolha do melhor modelo de cópulas para o ajustamento dos dados é uma importante mas difícil tarefa. O verdadeiro mecanismo de geração de dados é desconhecido, para um determinado conjunto de dados, é assim possível que vários candidatos a cópulas ajustem os dados razoavelmente bem ou que nenhum dos candidatos consegue ajustar os mesmos dados de forma correta. Quando o método tau de *Kendall* é utilizado, a prática habitual é proceder ao ajustamento de dados com todas as cópulas e escolher a que apresenta maior probabilidade (Frees & Wang, 2005). Baseado no processo de *Kendall* é construída uma ferramenta gráfica para escolher a mais credível entre as famílias de cópulas (Fermanian, 2005).

2.9. Testes de Independência Multivariada

A modelação de distribuições multivariadas com margens contínuas baseada na teoria de Cópulas é descrita como uma sucessão de testes baseados em classificações: um teste multivariado de aleatoriedade seguido por um teste de independência mútua e uma série de testes de qualidade do ajuste. A generalidade dos testes descritos é baseada nas cópulas empíricas que contém estimadores não paramétricos (Kojadinovic & Yan, 2010).

A base do trabalho está assente da independência ou não dos valores que integram as variáveis da amostra. De acordo com o artigo de Yan & Kojadinovic (2010), a função matemática em baixo é sugerida como a base para o teste de independência mútua entre componentes de uma variável \mathbf{X} .

$$(20) \quad I_n = \int_{[0,1]^d} n \left\{ C_n(\mathbf{u}) - \prod_{i=1}^d u_i \right\}^2 d\mathbf{u}.$$

Um aspeto interessante do teste em consideração advém do facto de quando existe independência mútua das componentes X_1, X_2, \dots, X_n da variável \mathbf{X} , o processo empírico poder ser decomposto em dois subprocessos $2^d - d - 1$ que convergem em conjunto para os processos Gaussianos centrados na independência mútua. Uma propriedade fundamental desta decomposição é que a independência mútua entre X_1, X_2, \dots, X_n é equivalente a ter $M_A(\mathbf{u}) = 0$, para todos $\mathbf{u} \in [0,1]^d$ e para todos $A \subseteq \{1, \dots, d\}$ de tal forma que $|A| > 1$.

Em vez de um teste único I_n é proposto uma fórmula que englobe em conjunto os dois subprocessos, sendo:

$$(21) \quad M_{A,n} = \int_{[0,1]^d} n \{M_A(C_n)(\mathbf{u})\}^2 d\mathbf{u},$$

onde, $A \subseteq \{1, \dots, d\}$, $|A| > 1$, que são assintoticamente independentes entre si quando observada a hipótese nula da independência. A estatística $M_{A,n}$ é focada na dependência entre componentes de \mathbf{X} cujos índices podem ser observados em A . A combinação das propriedades dos dois subprocessos formam esta estatística específica. Para visualizar os resultados dos testes de independência quando baseados em estatística $M_{A,n}$ recorre-se à representação gráfica, denominada de dependograma. Para cada subconjunto $A \subseteq \{1, \dots, d\}$, $|A| > 1$, uma barra vertical é desenhada e cuja altura é proporcional ao valor estatístico de $M_{A,n}$ (Yan & Kojadinovic, 2010).

O teste descrito anteriormente pode ser obtido com as funções `indepTestSim` e `indepTest`. A função `indepTestSim` devolve um conjunto de dados que contem um vasto número de resultados independentes dos testes estatísticos sob independência mútua. A função `indepTest` baseia-se nos dados da função anterior e devolve os resultados estatísticos e os p -values aproximados.

2.10. Testes de Aleatoriedade (Randomness)

Segundo Kojadinovic & Yan (2010), que remete o seu raciocínio para Genest *et al.* (2002) e Genest & Rémillard (2003), afirma que os testes de independência multivariada podem ter a sua versão estendida para os testes de aleatoriedade. Ou seja, dada uma sequência univariada estacionária de variáveis aleatórias contínuas X_1, X_2, \dots, X_n , e um valor inteiro de $p > 1$, o primeiro vector aleatório de observações $\mathbf{Y}_i = (X_i, \dots, X_{i+p-1})$, $i \in \{1, \dots, n\}$, onde p é a dimensão incorporada. O p -value pode ser determinado com base na independência da séria e pode medido usando a estatística:

$$(22) \quad I_n^s = \int_{[0,1]^p} n \left\{ C_n^s(\mathbf{u}) - \prod_{k=1}^p u_k \right\}^2 d\mathbf{u},$$

onde C_n^s é o análogo de série da cópula empírica, calculado a partir de Y_1, Y_2, \dots, Y_n . Tal como secção anterior e seguindo o mesmo raciocínio, estas estatísticas ao serem combinadas numa estatística global com base nas regras de Fisher ou Tippett, apresentam testes mais poderosos do que os baseados unicamente em I_n^s . A aplicação computacional destes testes é feita através das funções *serialIndepTestSim* e *serialIndepTest*.

2.11. Testes de Ajustamento (Goodness-of-Fit)

Os testes de ajustamento baseiam-se todos na expressão empírica,

$$(23) \quad \mathbf{C}_n(\mathbf{u}) = \sqrt{n} \{ C_n(\mathbf{u}) - C_{\theta_n}(\mathbf{u}) \}, \quad \mathbf{u} \in [0, 1]^d,$$

em que C_n é a cópula empírica definida pela expressão,

$$(24) \quad C_n(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(\hat{\mathbf{U}}_i \leq \mathbf{u}), \quad \mathbf{u} \in [0, 1]^d.$$

sendo C_{θ_n} um estimador de C enquanto a hipótese $H_0 : C \in \{C_{\theta_n}\}$ se mantiver. O referido estimador adquire assim de forma idêntica a possibilidade de ser o estimador dos métodos de *Kendall's tau* e *Spearman's rho*, ou o estimador do método baseado em medidas de dependência (Yan & Kojadinovic, 2010; Genest *et al.* 2009b).

De acordo com o estudo de Kojadinovic *et al.* (2010), existe um procedimento para amostras compostas por um elevado número de observações, baseado no teorema multiplicativo do

limite central, que evita assim os altos custos computacionais necessários para o método paramétrico de ajustamento, além de que apresenta uma solução muito mais célere. Esta eficiência surge quando na presença de condições adequadas de regulamentação, o processo de ajustamento C_n pode ser descrito pela expressão,

$$(25) \quad \mathbb{C}_n(\mathbf{u}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{J}_i(\mathbf{u}) + R_n(\mathbf{u}),$$

onde J_1, \dots, J_n são processos de distribuição independentes e idênticos que dependem simultaneamente do estimador Θ_n e que a hipotética família de Copulas $\{C_\theta\}$ e $\sup |R_n(\mathbf{u})|$ tendam em probabilidade para 0, e também pelo facto dos termos de $J_{i,n}$ só por uma vez necessitarem de ser estimados (Genest *et al.* 2009b; Genest *et al.* 2006).

2.12. Medidas de Associação

Duas variáveis aleatórias X e Y consideram-se independentes no caso da sua distribuição conjunta ser equivalente à multiplicação das suas distribuições marginais, de acordo com a seguinte expressão:

$$(26) \quad F(x, y) = F(x)F(y)$$

Contudo, como de uma forma geral as variáveis financeiras não se figuram como independentes, pelo que se torna fundamental uma adequada compreensão da relação de dependência entre as mesmas para uma correta avaliação do risco de crédito. Como tem vindo a ser referido o principal intuito da função cópula é a representação dessas mesmas relações, desta forma, a teoria de cópulas surge diretamente relacionada com o estudo de medidas de dependência entre variáveis aleatórias (Nelsen, 2006).

De acordo com Pereira (2009), apesar das medidas de dependência não serem a ferramenta mais adequada para analisar a dependência entre variáveis aleatórias devido às suas limitações, poderão ser vistas como uma mais-valia para a escolha de qual a cópula a ser utilizada. Apesar da correlação linear ser a medida de dependência mais popular devido à simplicidade da sua estimação, de acordo a expressão matemática:

$$(27) \quad \rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X) \cdot \text{var}(Y)}}.$$

esta possui alguns problemas de afirmação, tais como: não capturam relações não-lineares entre variáveis, devido a esta limitação, o coeficiente de correlação não pode ser definido como uma função cópula; só pode ser utilizada na construção de distribuições conjuntas que sejam normalmente distribuídas; uma única observação da amostra pode ter um peso muito significativo na estimação da correlação, assim, a correlação não é uma medida de associação robusta, não devendo mesmo ser utilizado no caso de amostras que apresentem *outliers*.

As medidas de concordância ao terem a capacidade de capturar relações não-lineares podem ser aplicadas a outros tipos de distribuição conjunta, figurando-se como medidas robustas e invariantes nos casos de transformações estritamente crescentes das variáveis, evitando assim algumas das limitações do coeficiente de correlação. Ou seja, se duas variáveis são concordantes, as mesmas continuarão a sê-lo mesmo após sofrerem esse tipo de transformação (Pereira, 2009).

A ideia de concordância assenta na seguinte definição: se (x_i, y_i) e (x_j, y_j) forem duas observações do vector de variáveis aleatórias contínuas (X, Y) , podemos afirmar que as observações são concordantes se $x_i > x_j$ e $y_i < y_j$ ou se $x_i < x_j$ e $y_i > y_j$, e discordantes se $x_i < x_j$ e $y_i < y_j$ (Pereira, 2009).

Segundo Nelsen (2006) e Pereira (2009), o Tau de *Kendall* e Rho de *Spearman* são as medidas de concordância mais conhecidas e apresentam uma relação direta com a função de cópula.

2.13. Métodos de Estimação dos Parâmetros da Cópula

Segundo "Stevenson (1981), citado por Cordeiro (2009), «a estimação é o processo que consiste em utilizar dados amostrais para estimar os valores de parâmetros populacionais desconhecidos» ". Assim vai-se fazer uma breve abordagem teórica aos três principais métodos de estimação que serão os utilizados posteriormente.

2.13.1. Tau de Kendall

O Tau de Kendall é definido de acordo com Nelsen (2006), como a diferença entre a probabilidade de concordância e a probabilidade de discordância de duas variáveis aleatórias, através da expressão:

$$(28) \quad \tau = \tau_{X,Y} = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0]$$

em que (X_1, Y_1) e (X_2, Y_2) são vetores aleatórios independentes e identicamente distribuídos, formados a partir de observações (x_i, y_i) e (x_j, y_j) do vetor de variáveis aleatórias (X, Y) .

A estimação do Tau de *Kendall* pode ser utilizado para a estimação da função de cópula que relaciona X e Y (variáveis aleatórias contínuas com distribuição conjunta dada pela cópula C), dado que a relação,

$$(29) \quad \tau = 4 \iint_{I^2} C \, dC - 1 = 4 \iint_{I^2} C(u, v) c(u, v) \, du \, dv - 1$$

entre o Tau de *Kendall* e a cópula é válida. $C(u, v)$ é definido como a densidade da cópula (Nelsen, 2006).

2.13.2. Rho de Spearman

À semelhança do tau de *Kendall*, esta é uma medida de dependência baseada na concordância e discordância, sendo definida como proporção da probabilidade de concordância menos a probabilidade de discordância para dois vetores (X_1, Y_1) e (X_2, Y_2) , através da expressão,

$$(30) \quad \rho = \rho_{X,Y} = 3(P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0]).$$

O Rho de Spearman pode ser calculado através da cópula associada entre as variáveis X e Y por,

$$(31) \quad \rho_S = 12 \iint_{I^2} C(u, v) \, du \, dv - 3$$

Ambas as medidas tau de *Kendall* e rho de *Spearman* são medidas de probabilidade de concordância entre variáveis aleatórias dadas as cópulas, porém os seus respectivos valores ρ e τ são diferentes (Nelsen, 2006).

2.13.3. Estimação Baseada em Medidas de Dependência

Uma forma alternativa de estimar o parâmetro de uma cópula, é apresentado por Kojadinovic & Yan (2010), consiste na identificação da relação entre variáveis através da estimação de medidas das medidas de dependência já enunciadas. Este é um método bastante mais simples do que o anterior, contudo a sua aplicação pode revelar-se restrita, já que fica limitado a distribuições bivariadas, e para cópulas com apenas um parâmetro, como as cópulas arquimedianas. Para o presente estudo, quando se está a considerar a estimação da relação de

dependência entre duas variáveis, será este o modelo mais ajustado e o que proporciona os resultados com a melhor relação/exigência computacional.

2.14. Aplicação Prática

De acordo com Jorion (2007), existem diversas etapas a percorrer na utilização de Cópulas para o cálculo do risco de crédito de uma carteira composta por n fatores de risco, assim:

- Identificação das distribuições marginais dos fatores de risco;
- Testes de independência entre variáveis;
- Seleção do tipo de cópula mais ajustada para a integração das distribuições marginais em distribuições conjuntas;
- Estimação dos parâmetros das marginais e da função cópula;
- Modelação das variáveis de risco (distribuições marginais);

A abordagem do presente trabalho tem como objetivo os quatro primeiros itens, sendo que, a modelação das variáveis de risco é um subtema pertinente de continuidade futura

3. Dados

O procedimento empírico do trabalho teve como primeiro passo a seleção de variáveis de risco mais significativas, a recolha de dados referentes aos seus valores e o cálculo do valor de perda de crédito. Os testes estatísticos descritivos foram obtidos através do *software* Excel.

A amostra utilizada provém do *portfolio* de crédito ao consumo de uma instituição financeira Portuguesa, composta por 3.000 observações individuais realizadas entre 31-01-2009 e 31-12-2011, sendo que para a aplicação prática de Cópulas a criação das distribuições marginais para as variáveis de risco em estudo foi efetuada com análise a um único ou seja em 31-12-2011.

Com base na definição de taxa de perda descrita em capítulo 2, identificou-se que a amostra inclui algumas observações atípicas, nas quais a taxa de perda assume um valor negativo, e isto deve-se essencialmente a dois fenómenos:

- Existência de um valor em atraso recuperado no mês em curso em que o valor pago é superior ao valor da prestação mensal;
- Existência de um pagamento antecipado sem desconto.

Estas situações são algo frequentes em carteiras de crédito ao consumo, contudo, quando a amostra é suficientemente grande, é expectável que os casos sejam absorvidos pela mesma,

não existindo taxas agregadas de perda negativas em cada momento da avaliação. Para uma correta compreensão do comportamento da distribuição temporal das perdas de crédito, são apresentados os resultados na Figura 1 e Figura 2, que comportam os dados relativos às series temporais da variável “perda de crédito” durante a totalidade do período em estudo.

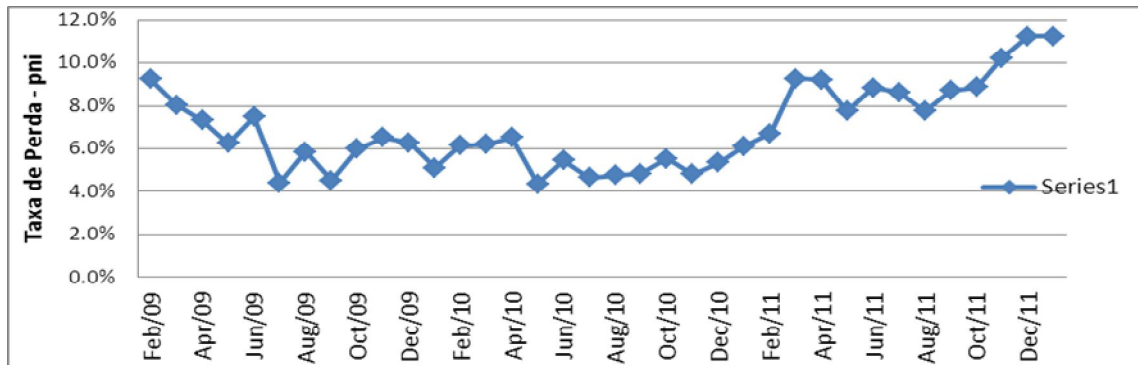


Fig 1 – Série temporal de perda de crédito da amostra. (Fonte própria).

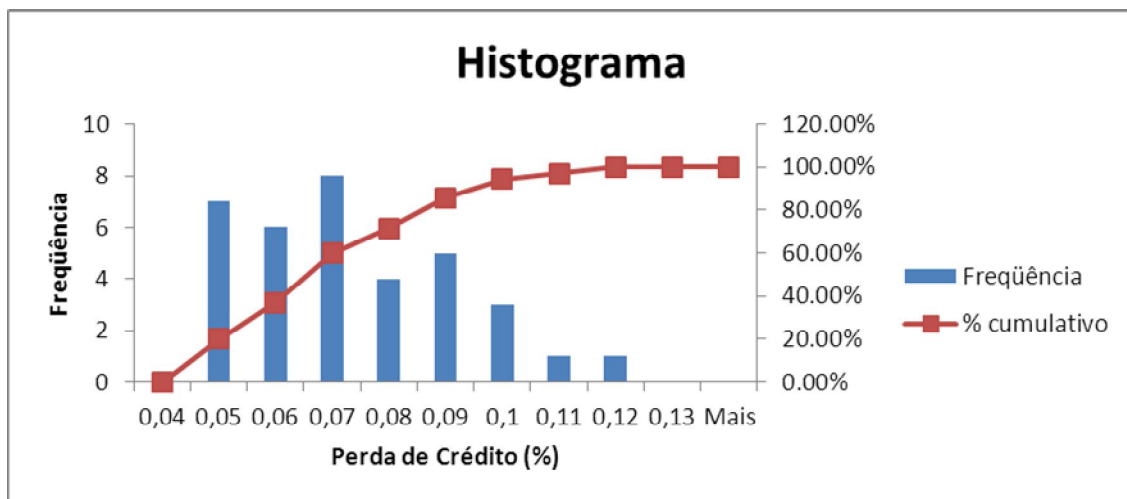


Fig 2 – Histograma de observações mensais de perda de crédito percentual. (Fonte própria).

Passando a uma análise detalhada à amostra utilizada, assumiu-se como uma das variáveis de risco: a variável “scoring”. Os valores observados que compõem a distribuição referente à variável em questão foram analisados de acordo com o sistema de classificação por grupos de risco, ou *Credit Scoring*. Modelo este adotado de acordo com a metodologia da instituição financeira em análise, desenvolvida internamente através de um modelo de *Credit Scoring*, assente em fatores de características pessoais do indivíduo tomador do crédito, características e comportamentos no pagamento da operação financeira em estudo e outros fatores comportamentais como são o caso do registo ou não de incidentes com cheques e histórico de registo de incidentes na Central de Responsabilidades do Banco de Portugal.

Assim, são definidos 4 grandes grupos de risco, estabelecida a sua ordem da seguinte forma:

- Grupo 1 – Risco mais baixo
- Grupo 2 – Risco Médio
- Grupo 3 – Risco Médio/Alto
- Grupo 4 – Risco mais alto

TABELA I

Distribuição de grupos de risco

Grupo De Risco	Frequência na Amostra	Frequência em Incumprimento
1	35,200%	2,210%
2	29,800%	1,934%
3	25,733%	40,331%
4	9,267%	55,525%

Fonte Própria

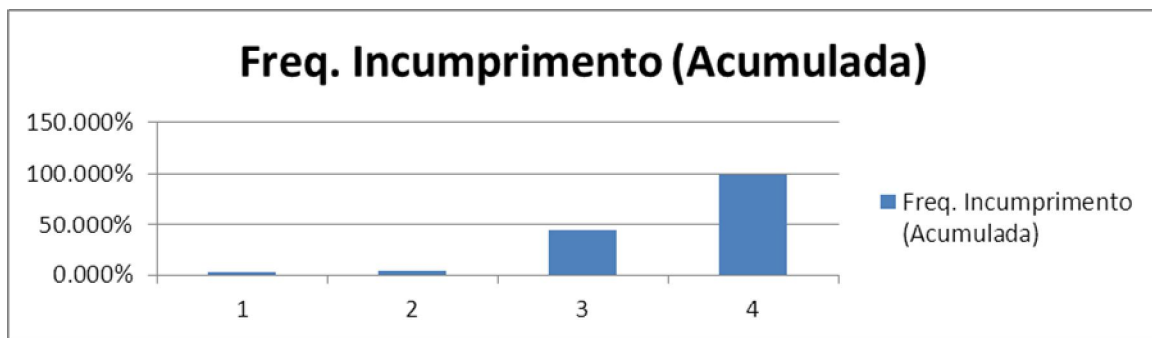


Fig 3 – Gráfico frequências acumuladas de incumprimento por grupo de risco. (Fonte própria)

Para a criação da amostra conforme já foi sendo referido, assumiu-se um conjunto de variáveis representativas dos fatores de risco. A primeira foi uma variável explicativa de perda de crédito - "*pnivdni*", as restantes variáveis foram assumidas de acordo com a literatura como os principais fatores impulsionadores de perdas, como são o caso concreto de: "*scoring*", "*idade*", "*regiao*", "*civil*" e "*evento*".

Para além da variável "*scoring*", a qual já foi devidamente explicada a sua construção e existência, segue-se a explicação para as restantes variáveis:

- "*pnivdni*" – representa o valor máximo de perda de crédito no momento da observação, conforme a fundamentação em capítulo 2.
- "*idade*" – representa a idade em valor absoluto do tomador de crédito à data da observação.

- “regiao” – representa a localização do tomador do crédito. Foram definidas 3 grandes regiões, assim: Grande Porto; Grande Lisboa; Restantes Regiões.
- “Evento” – ocorrência ou não de eventos de incumprimento superiores a 30 dias, no período iniciado na data de contratação até ao momento da observação.

Com base no estudo de Clemente & Romano (2004), que desenvolveu um modelo baseado na teoria de cópulas para a mensuração e otimização do risco de crédito de uma carteira de crédito ao consumo, também aqui todas as distribuições marginais assumem características de não-normalidade, com presença de assimetria e leptocurtose (caudas pesadas), inclusive as distribuições de “regiao”, “civil” e “evento”, caracterizam-se como discretas.

Como já referido em capítulo anterior a taxa d utilizada para o cálculo da variável marginal “perda de credito” é possível obter em tabelas que incorporem as taxas médias para operações de crédito ao consumo, fornecidas pelo Banco de Portugal e representadas em Taxas credito consumo Banco Portugal - Anexo 2. Por conveniência, optou-se por utilizar o valor da taxa média para o período em análise, em vez da utilização do valor da taxa em cada mês.

4. Metodologia da pesquisa empírica

A pesquisa empírica tem como primeira abordagem a elaboração das distribuições marginais de cada variável e definição das funções cópula mais ajustadas para cada distribuição bivariada e multivariada. A estimação de parâmetros e outros cálculos para modelagem da dependência entre variáveis via função cópula são realizados pela aplicação *Copula R Package* (Anexo 3).

O modelo de risco proposto no presente trabalho tem como objetivo final a modelação do comportamento de ocorrências de situações de perda de valor, através da criação do modelo de cópulas com o melhor desempenho no que respeita à relação de dependência entre as distribuições marginais assumidas. Como é expectável a ocorrência das situações enunciadas na plenitude de uma carteira não se figuram independentes entre si, dado que, existem diversos fatores macroeconómicos e outros intrínsecos a cada devedor, que originam tendências de alta ou baixa no número de situações de incumprimento. As relações de dependência entre valores de perda de crédito e as variáveis de risco representam o principal elemento na formulação do nosso modelo de risco de crédito.

Identificadas as variáveis, passamos à construção das cinco distribuições bivariadas para análise prática, que são: *pnivdni – scoring*; *pnivdni-idade*; *pnivdni-regiao*; *pnivdni-civil*; *pnivdni-*

evento. Serão apresentados os resultados relativos aos testes de independência entre distribuições, teste de aleatoriedade, testes de ajustamento de cada família de cópulas e testes erro padrão do parâmetro de ajustamento.

Facto importante a referir, é que é analisando o número de observações com valores diferentes em cada distribuição marginal verifica-se que este é muito inferior ao número total de observações - 3000, por exemplo, para as variáveis "pnivdni" e "idade" existem 363 e 53 valores diferentes, respetivamente, o que significa que há um conjunto de ligações ou laços entre dados. Como veremos mais tarde, ignorar estas ligações pode afetar as conclusões da análise qualitativa. Assente na pouca consistência do pressuposto que todos os testes associados à teoria de Cópulas necessitam que as distribuições marginais sejam contínuas, facto que implica necessariamente que a dependência entre valores da mesma variável tenha probabilidade zero, para ultrapassar esta situação e lidar com as dependências entre valores de uma forma mais satisfatória, foi proposto a construção de pseudo-observações, que originam a quebra das ligações entre observações da mesma distribuição e transformam as distribuições marginais discretas em contínuas, através do método *random rank* proposto por Kojadinovic & Yan (2010).

Pelo facto das distribuições indicarem a não existência do elemento temporal, dado que todas as variáveis estão relacionadas com a variável "pnivdni" e inseridas num único momento temporal n , de acordo com Yan & Kojadinovic (2010), não há razão para testar a dependência temporal das distribuições, pelo que se passará ao trabalho de modelação de cópulas.

O próximo passo na construção de um modelo de cópulas passa pelo teste de independência entre variáveis. Mais uma vez serão testados os binómios já identificados. Os testes estatísticos de independência apresentam um quadro muito completo de resultados, dado que a amostra é composta por 3000 observações, com a criação de $N=1000$ simulações independentes. A conjugação de um elevado número de simulações com o número de observações implica a utilização de uma máquina com elevado poder computacional, o que não sucedeu, facto que originou que o tempo de execução fosse demasiado moroso e pouco prático. Foi definida como hipótese nula para os testes de independência a expressão: $H_0 =$ independência entre variáveis.

Devido à rejeição da hipótese nula de independência na maioria das distribuições bivariadas, o próximo passo é por conseguinte, a realização de um vasto conjunto de testes de ajustamento.

Temos assim de considerar as diferentes famílias de Cópulas, como exemplo, Gumbel, Clayton, normal e t com n graus de liberdade (4, 10 e 20). Assumindo as técnicas propostas pelo estudo de Kojadinovic & Yan (2010), o número de graus de liberdade da cópula t deve ser fixado e por conseguinte não deve ser considerado como um parâmetro a ser estimado. Por *default* o *software* utilizado apresenta 4 graus de liberdade.

Para qualquer uma das famílias de cópulas, pode ser utilizado qualquer um dos métodos de estimação: itau (inversão de *Kendall's tau*) e irho (inversão de *Spearman's rho*), a escolha final recairá no método que apresenta os melhores resultados.

A estimação dos parâmetros das cópulas, conforme descrito em capítulo 2, apresenta dois métodos de simulação possíveis: o método multiplicativo (não-paramétrico) e o método paramétrico. Ambas as propostas apresentaram resultados muito semelhantes com p -values muito similares, contudo o método multiplicativo revelou-se muito mais rápido, fator decisivo na escolha do mesmo para a análise e discussão de resultados. Este fator de decisão torna-se tanto mais importante à medida que o número de observações e de simulações aumenta.

Finalmente são calculados os valores de erro padrão da estimativa do parâmetro das famílias de cópulas que obtiveram p -values com valores estatisticamente significativos, que são:

- *pnivdni*-idade – Cópula de Clayton e Cópula Normal
- *pnivdni*-civil - Cópula de Clayton e Cópula Normal

Este cálculo é utilizado para avaliar a capacidade qualitativa do ajuste de cada modelo de cópula, sendo realizado através do comando *fitCopula*.

Para uma análise mais profunda ao comportamento das relações de dependência entre a globalidade das variáveis em estudo, e partindo do pressuposto que as observações de cada variável da amostra são independentes entre si, através da aplicação da função *ties.method*, *random*, prosseguimos para o teste de independência mútua entre as variáveis definidas.

Em consideração aos resultados da aplicação de cópulas nas diferentes distribuições bivariadas, os quais assumem a relação de dependência entre as distribuições marginais "*pnivdni*", "idade" e "civil", assim como, ao bom ajuste que as cópulas proporcionam em cada distribuição conjunta "*pnivdni*-idade" e "*pnivdni*-civil", tomou-se como facto importante para análise, o comportamento da distribuição conjunta das três variáveis descritas. As famílias de

cópulas utilizadas são: duas famílias com um parâmetro, *Clayton* e *Gumbel*, e outras quatro famílias com três parâmetros, a Normal e a *t* com 4, 10 e 20 graus de liberdade.

De forma similar, serão realizados todos os procedimentos relativamente aos testes já identificados anteriormente, com a inclusão do dependograma, dado a mais-valia desta ferramenta, pois permite visualizar graficamente o resultado dos testes de independência de todas as relações possíveis da distribuição conjunta. Esta função é conseguida através do comando *dependogram*.

Mais uma vez, pelo facto da proposta de simulação multiplicativa apresentar *p-values* semelhantes à proposta paramétrica, com tempos de resposta cerca de 300 vezes mais rápido, será dada maior ênfase aos resultados de ajustamento obtidos pela simulação multiplicativa. Assim a análise ao erro padrão de cada parâmetro estimado, será realizada de forma perfeitamente arbitrária, a partir da família de cópulas com o maior *p-value*.

5. Análise de Resultados

O capítulo tem como principal objetivo a apresentação dos resultados obtidos pelos testes realizados na aplicação da teoria de cópulas às distribuições conjuntas, quer sejam bivariadas quer sejam multivariadas. A escolha do modelo de cópula a utilizar em cada distribuição, terá sempre em consideração o modelo com maior p -value nos testes de ajustamento.

O cálculo do valor de perda de crédito para cada observação, realizado de acordo com capítulo 2, conduz-nos à construção de uma distribuição marginal, a qual é alvo da aplicação da estatística descritiva, o permite em primeira instância avaliar algumas características da amostra, como a média, desvio padrão e coeficiente curtose.

TABELA II

Distribuição de perda de crédito.

<i>Perda de Credito (em valor)</i>	
Média	13,93€
Erro padrão	1,06€
Mediana	0
Modo	0
Desvio padrão	58,46€
Variância da amostra	3418,47419
Curtose	51,16164337
Assimetria	4,972685287
Intervalo	1370,60€
Mínimo	-484,57€
Máximo	886,03€
Soma	41797,89€
Contagem	3000
Nível de confiança (95,0%)	2,093046686
Perda Credito/Valor	
Carteira (%)	2,3221 %

Fonte: Própria

Numa amostra em que o valor médio de cada operação de crédito é aproximadamente 6.000€, é apresentado um valor global de perda no montante de 41.797,89€ e um rácio de perda de 2,3221%. Uma das primeiras ilações a retirar da análise da tabela II é o valor extremamente elevado do coeficiente de curtose, o que significará que estamos na presença de uma distribuição afunilada, assimétrica, com presença de caudas pesadas. Significado prático deste fenómeno é a existência de observações que se afastam da média em vários múltiplos do

desvio padrão. Ainda a registar a ideia que a distribuição de perdas não tem características de normalidade, ideia demonstrada pela característica intrínseca da própria distribuição, a assimetria positiva, com desvio para a esquerda do gráfico de distribuição. Este facto deve-se essencialmente à maior percentagem de observações com perda igual a zero, contudo, também à presença de valores elevados de perda, que se apresentam claramente como *outliers* à direita da distribuição.

A primeira análise às figuras 1 a 10 – Anexo 1, permite-nos verificar que para o procedimento de quebra de ligações entre observações, após ser aplicado o método *random rank: ties.method*, em cada distribuição bivariada as ligações são partidas passando a distribuição a ter um padrão uniforme, apresentando-se como uma distribuição de margens contínuas. Para facilitação foram construídos gráficos em três dimensões para cada distribuição bivariada, que permite compreender de forma quase intuitiva os resultados do procedimento realizado.

TABELA III

Testes de Independência entre variáveis. $H_0 = \text{TRUE}$ (Independência entre variáveis)

	Teste Independência - <i>In</i>	<i>p</i> -value	Decomposição Mobius <i>MA, n</i> - (<i>p</i> -value)		Análise
<i>pnivdni</i> -idade	0,04582787	0.1003996	0.1003996	0.1003996	Pouca ou nenhuma evidência contra H_0
<i>pnivdni</i> -scoring	2,70941100	0.004950495	0.004950495	0.004950495	Rejeito H_0
<i>pnivdni</i> -regiao	0,01641222	0.7673267	0.7673267	0.7673267	Aceito H_0
<i>pnivdni</i> -civil	0,07268631	0.004950495	0.004950495	0.004950495	Rejeito H_0
<i>pnivdni</i> -evento	0,08626648	0.004950495	0.004950495	0.004950495	Rejeito H_0

Fonte Própria

A análise dos resultados da tabela III demonstra a evidência estatística dos mesmos para um nível de confiança de 5%. Com exceção para a distribuição *pnivdni* – região, em que a hipótese nula de independência é claramente aceite, o que por inerência leva a aceitar que não existe dependência entre as observações da variável “*região*” e os acontecimentos de perda de crédito, nas restantes distribuições bivariadas a hipótese nula ou é claramente rejeitada ou não existe forte evidência para que a mesma seja aceite. Assim, para os casos de *pnivdni*-idade, *pnivdni*-scoring, *pnivdni*-civil e *pnivdni*-evento, faz sentido considerar a possibilidade de utilização das diferentes famílias de cópulas como forma modelar a dependência entre variáveis.

Os resultados dos testes de ajustamento aplicados para as diferentes famílias de cópula descritos na tabela IV, mostram-nos uma diversidade muito grande de possíveis utilizações em cada distribuição bivariada. Assim, para a distribuição *pnivdni*-idade verificamos que as cópulas de Clayton e Normal, com valores de 0,06843 e 0,08741, respetivamente, não são rejeitadas para um nível de significância de 5%. Logo é possível deduzir que ambos os modelos são eficientes no ajustamento do parâmetro de cópula.

Nos casos das distribuições *pnivdni-scoring* e *pnivdni-evento*, as quais se situam num patamar diferente das restantes, dado que, apresentam um *p*-value idêntico em todos os casos na ordem de valores de 0,000495. Esta constatação é a base da ideia que para qualquer uma das famílias de cópulas utilizadas, o ajustamento do parâmetro é rejeitado, inclusive, para um nível de confiança de 1%, afirmando claramente que as estruturas de dependência assumidas não apresentam valor estatístico significativo.

TABELA IV

Testes de ajustamento de Cópulas. $H_0 = \text{FALSE}$ (Validade do modelo de cópula em análise para a distribuição bivariada)

	Gumbel Copula		Clayton Copula		Normal		t (4 gl)	
	estatística	<i>p</i> -value	estatística	<i>p</i> -value	estatística	<i>p</i> -value	estatística	<i>p</i> -value
<i>pnivdni</i> -idade	0.0434	0.005495	0.0301	0.06843	0.0272	0.08741	0.0445	0.005495
<i>pnivdni-scoring</i>	0.4304	0.000495	1,53410	0.000495	0.7835	0.000495	0.7393	0.000495
<i>pnivdni-regiao</i>	0.0166	0.6089	0.0167	0.5495	0.0162	0.5	0.0485	0.00495
<i>pnivdni-civil</i>	0.0212	0.2592	0.0162	0.6349	0.0166	0.527	0.0522	0.001499
<i>pnivdni-evento</i>	0.057	0.000495	0.0873	0.000495	0.0702	0.000495	0.0735	0.000495

Fonte Própria

Para a distribuição *pnivdni-regiao* apesar dos resultados apresentarem valor estatístico significativo, em que a cópula de Gumbel apresenta o valor de 0,6089, devido à presença de independência entre distribuições marginais de acordo com os resultados expostos na tabela III, estes últimos resultados não são tidos em consideração. Para a distribuição *pnivdni-civil* com exceção da cópula de *t*-Student, que apresenta um *p*-value de 0,001499, as restantes são aceites como estruturas de modelação da dependência para um nível de significância de 5%.

Os resultados obtidos ainda de acordo com a tabela anterior mostram que para a distribuição *pnivdni*-idade o resultado com maior valor de *p*-value (0,08741) é obtido através da família de

Cópula Normal, e para o caso de *pnivdni-civil* é a Cópula de Clayton que obtém o maior valor (0,6349). Arbitrariamente optamos pela definição da Cópula que melhor ajusta cada distribuição, pela que apresenta o maior valor de *p-value*.

Os resultados apresentados anteriormente foram obtidos através do método de estimação de *Kendall's tau*, em conjunto com o método de simulação multiplicativo devido à sua maior velocidade e equivalência de resultados obtidos com os restantes métodos, quer de estimação quer de simulação. No entanto, para os casos nos quais não foi identificada relevância estatística nos modelos de cópulas utilizados, deverão ser realizados testes com $N = 10.000$, cruzados com todos os métodos de estimação e modelos de simulação apresentados, de forma a se obterem *p-values* com valores mais precisos. Para este nível de testes, é absolutamente necessária uma elevada capacidade computacional, pelo que se torna de certa forma pouco atrativo no meio empresarial.

TABELA V

Estimação do parâmetro da Cópula e correspondente erro padrão.

	Gumbel Copula			Clayton Copula		
	Parâmetro	Std. Error	p-value	Parâmetro	Std. Error	p-value
<i>pnivdni-idade</i>	1,00	0.01244131	0	-0.03076619	0.02412297	0.2021713
<i>pnivdni-scoring</i>	1,2288340	0.01945635	0	0.4576689	0.0389127	0
<i>pnivdni-regiao</i>	1,00	0.01232091	0	-0,0082389	0.02443923	0.736028
<i>pnivdni-civil</i>	1,0301960	0.01299589	0	0.06039237	0.02599177	0.02015129
<i>pnivdni-evento</i>	1,0202840	0.01331397	0	0.0405673	0.02662794	0.127637
	Normal			Student t		
	Parâmetro	Std. Error	p-value	Parâmetro	Std. Error	p-value
<i>pnivdni-idade</i>	-0,02453876	0.01953688	0.2091078	-0.02453876	0.01953688	0.2091078
<i>pnivdni-scoring</i>	0.2883612	0.01937953	0	0.2883612	0.01937953	0
<i>pnivdni-regiao</i>	0,006497537	0.01935324	0.7370716	-0,00649754	0.01935324	0.7370716
<i>pnivdni-civil</i>	0.04602551	0.01921434	0.01660345	0.04602551	0.01921434	0.01660345
<i>pnivdni-evento</i>	0.03122299	0.02008046	0.1199713	0.03122299	0.02008046	0.1199713

Fonte Própria

Os resultados obtidos na tabela V, revelam-se conclusivos para ambas as distribuições analisadas, *pnivdni-idade* e *pnivdni-civil*, para os casos concretos das cópulas assumidas como válidas, Normal, Clayton e Normal, com os valores de erro padrão a revelarem-se significativamente pequenos, e *p-values* que aceitam o erro padrão calculado, com exceção para a família de cópulas Gumbel na distribuição *pnivdni-civil*, que não aceita o erro padrão.

Tal como seria de esperar, o teste de independência mútua para a distribuição multivariada das variáveis “*pnivdni*”, “idade” e “civil”, comprova a forte evidência contra a hipótese nula de independência mútua, com o valor estatístico de 1,507693 e *p*-value 0,0004950495. O dependograma relativo ao teste indicado (Figura 11 – Anexo 1) mostra a rejeição de independência mútua para todos os subconjuntos de variáveis, incluindo para a distribuição conjunta que comporta simultaneamente as três distribuições marginais.

TABELA VI

Testes de ajustamento de Cópulas. $H_0 = \text{FALSE}$ (Distribuição multivariada)

	<i>t</i> c/ 4 GL	<i>t</i> c/ 10 GL	<i>t</i> c/ 20 GL	Gumbel	Normal	Clayton
	<i>p</i> -value	<i>p</i> -value	<i>p</i> -value	<i>p</i> -value	<i>p</i> -value	<i>p</i> -value
<i>pnivdni-idade-civil</i>	0.001499	0.0004995	0.001499	0.001499	0.0004995	0.002498

Fonte Própria

Relativamente aos resultados obtidos pelos dois métodos distintos: paramétrico e multiplicativo, de referir que ambos apresentam resultados muito idênticos com valores *p*-value praticamente iguais, contudo, a exemplo dos testes anteriores a proposta pelo método multiplicativo demonstrou ser manifestamente mais rápido.

Os resultados apresentados na tabela VI mostram que todas as famílias de cópulas são claramente rejeitadas para um nível de significância de 5%. Assim, não é possível assumir que a modelação de dependência da distribuição conjunta em análise é corretamente aferida através de qualquer que seja a família de cópulas.

6. Discussão e Conclusão

A agregação entre os diferentes fatores de risco deverá ser sempre feita com consideração à correlação existente entre os mesmos, através de modelos que avaliem a dependência de forma adequada, de maneira a que o valor não seja sobrestimado ou subestimado. Desta forma, as cópulas podem ser uma ferramenta muito útil, dado que permitem a construção de modelos multivariados com diferentes tipos de dependência entre as distribuições marginais.

Pelo facto dos testes associados aos modelos de cópulas abordados no trabalho serem baseados no pressuposto das distribuições marginais serem contínuas, é fundamental não ignorar a existência de ligações entre observações, essencialmente em amostras grandes. Por conseguinte, foi utilizado um processo computacional que gera as pseudo-observações através do processo de randomização.

O presente trabalho sendo baseado na randomização ou aleatoriedade de observações de determinada distribuição, apesar de não ser completamente satisfatório dado que os valores obtidos pelas randomizações não são mais do que pseudo-observações, que implica a impossibilidade de visualização de acontecimentos no plano prático, não deixa de constituir o primeiro passo a dar na utilização de cópulas em séries financeiras. Esta randomização é absolutamente necessária, dado que conforme visto em capítulo anterior, em cada distribuição marginal existe uma grande quantidade de observações com o mesmo valor, o que poderia indicar a presença de ligações ou laços entre observações.

Verificou-se que os resultados da construção dos diversos modelos de cópulas apresentam valores muito distintos para os testes de aleatoriedade, independência e ajustamento. Isto revela claramente e indo de encontro à bibliografia explorada, que para cada distribuição existem famílias de cópulas que são melhores modelos de ajustamento do que outras, especialmente em distribuições com valores extremos, com assimetria à esquerda da distribuição, conclui-se que a cópula de Clayton é substancialmente mais eficaz. Isto sucede concretamente nas distribuições "*pnivdni-idade*" e "*pnivdni-civil*".

Este estudo revelou a existência de fortes correlações entre as perdas de capital e os fatores de risco "idade" e "civil", quando analisadas em distribuição bivariada, conclusão confirmada pelos valores dos parâmetros de independência. Ambas as distribuições bivariadas são caracterizadas por uma estrutura assimétrica, indo de encontro à revisão de literatura sobre a Cópula de Clayton e ainda de acordo com o estudo de (Yu & Chuang, 2012) sobre a distribuição de séries financeiras. Como conclusões práticas, podemos inferir sobre a necessidade de ajustar a distribuição das idades e do estado civil, como forma de revisão em baixa para os valores de perda de crédito.

No que respeita às distribuições bivariadas *pnivdni-scoring* e *pnivdni-evento*, verifica-se que as famílias de Cópulas utilizadas não conseguem realizar uma boa estimação dos parâmetros, isto poderá ser devido quer aos tipos de distribuição assimétrica, com maior dependência da cauda inferior, como também impossibilidade de realizar os testes de ajustamento de definição de parâmetros com maior número de simulações, por exemplo $N = 10.000$, o que poderia representar um valor de *p-value* mais preciso.

No que respeita à distribuição multivariada em estudo, é possível concluir que apesar da hipótese de independência entre variáveis ser fortemente rejeitada, a modelação pelas famílias de cópulas não apresenta um resultado estatisticamente significativo.

Estes modelos de cópulas devem ser encarados não unicamente como uma proposta académica, mas também como uma plataforma de análise de dados reais conforme utilizado no presente trabalho, e um princípio para novos estudos na área.

7. Referências Bibliográficas

- Bessis J. (2002). *Risk Management in Banking*. 2ª Ed. John Wiley & Sons, Ltd.
- Bluhm C., Overbeck L. & Wagner C. (2003). *An introduction to credit risk modeling*. Boca Raton: CRC Press LLC
- Bouyé E., Durrleman, A. Nikeghbali G., Riboulet G. & Roncalli T. (2000). *Copulas for finance – a reading guide and some applications*. Working Paper. Financial Econometrics Research Centre.
- Chang J. & Chen A. (2009). *Copula, Correlated Defaults and Credit VaR*. Department of Quantitative Finance. National Tsing Hua University
- Cherubini U., Luciano E. & Vecchiato W. (2004). *Copula Methods in Finance*. 1ª Ed. Chichester: John Wiley & Sons Ltd.
- Chirinko, R. & Guill G. (1991). *A framework for assessing credit risk in depository institutions: toward regulatory reform*. Páginas 785-804. Journal of Banking and Finance.
- Clemente A. & Romano C. (2004). *Measuring and optimizing portfolio credit risk: a copula-based approach*. Working Paper nº 1. Centro Interdipartimentale sul Diritto e l'Economia dei Mercati
- Cordeiro F. (2009). *Aplicação da Teoria de Cópulas para o cálculo do VAR*. Escola de Economia de São Paulo.
- Crouhy M., Galai D. & Mark R. (2000). *A comparative analysis of current Credit Risk Models*. Páginas 59-117. Journal of Banking and Finance.
- Fermanian J. (2005). *Goodness-of-fit test for Copulas*. Páginas 119-152. Journal of Multivariate Analysis.
- Frees E. & Wang P. (2005). *Credibility using Copulas*. Páginas 31-48. North American Actuarial Journal.
- Frey R., Macneil A. & Nyfeler M. (2001). *Copulas and Credit Models Risk*. Working Paper. Swiss Banking Institute
- Froot K., Scharfstein D. & Stein J. (1993). *Risk Management: Coordinating corporate investment and financing policies*. Vol1. 48, Páginas 1629-1658. The Journal of Finance.

- Genest C., Quessy J. & Rémillard B. (2002). *Test of serial independence based on Kendall's process*. Vol. 30, N 3, Páginas 441-461. The Canadian Journal of Statistics.
- Genest C. & Rémillard B. (2003). *Tests of Independence and Randomness Based on the Empirical Copula process*. SSC Annual Meeting.
- Genest C., Quessy J. & Rémillard B. (2006). *Goodness-of-fit procedures for copula models based on the probability integral transformation*. Scandinavian Journal of Statistics.
- Genest C., Gendron M. & Bourdeau-Brien M. (2009a). *The advent of Copulas in Finance*. 15:7:8, 609-618. The European Journal of Finance.
- Genest C., Rémillard B. & Beaudoin D. (2009b). *Goodness-of-fit test for copulas: A review and a power study*. Páginas 199-213. Insurance: Mathematics and Economics.
- Goorah A. (2007). *Real Estate Risk Management with Copulas*. Vol. 24. Páginas 289-311. Journal of Property Research.
- Hamilton D. & James J. (2001). *Copula Methods and the Analysis of Credit Risk*. University of Warwick: Warwick Business School.
- Jorion P. (2007). *Value at Risk: The new benchmark for managing financial risk*. 3ª Ed. McGraw-Hill.
- Jouanin J.F., Rapuch G., Riboulet G. & Roncalli T. (2001). *Modeling dependence for credit derivatives with copulas*. Crédit Lyonnais: Groupe de Recherche Opérationnelle.
- Kojadinovic I., Yan J. & Holmes M. (2010). *Fast Large-sample Goodness-of-Fit for Copulas*. Statistical Sinica. Forthcoming.
- Kojadinovic I. & Yan J. (2010). *Modeling Multivariate Distribution with Continuous Margins Using the copula R Package*. Vol. 34. Issue 9. Journal of Statistical Software.
- Kole E., Koedijk K & Verbeek (2007). *Selecting copulas for risk management*. Páginas 2405-2423. Journal of Banking & Finance 31.
- Li D. (2000). *On Default Correlation: A Copula Function Approach*. Working Paper. Páginas 99-07. The RiskMetrics Group.

- Nelsen R. (2006). *An Introduction to Copulas*. 2ª Ed. Springer.
- Pereira P. (2009). *Cópulas – Uma alternativa para a estimação de modelos de risco multivariados*. Working paper. Escola de Economia de São Paulo.
- Schonbucher P. & Schubert D. (2001). *Copula-Dependent default risk in intensity models*. Working Paper. Bonn University, Economics Faculty, Department of Statistics
- Schuermann T. (2004). *What do we know about loss-given-default*. Working paper. Federal Reserve Bank of New York.
- Tops R. (2010). *Copulas and correlation in credit risk*. University of Amsterdam, KdV Institute for mathematics
- Siqueira J. (2000). *Risco: da filosofia à administração*. Conventit Selecta 2. Editora Mandruvá
- Yan J. (2007). *Enjoy the joy of Copulas: with a package copula*. Volume 21, Issue 4. Journal of Statistical Software.
- Yan J. & Kojadinovic I. (2010). *Copula: Multivariate dependence with Copulas. R package version 0.9-5*. «Disponível: <http://cran.r-project.org/web/packages/copula/>»
- Yu T. & Chuang C.C. (2012). *The relationship between efficiency and leading market power: a longitudinal view of the life insurance industry in Taiwan*. Applied Economics

Anexo 1

Pseudo-observações por quebra aleatória de ligações intra-distribuições bivariadas

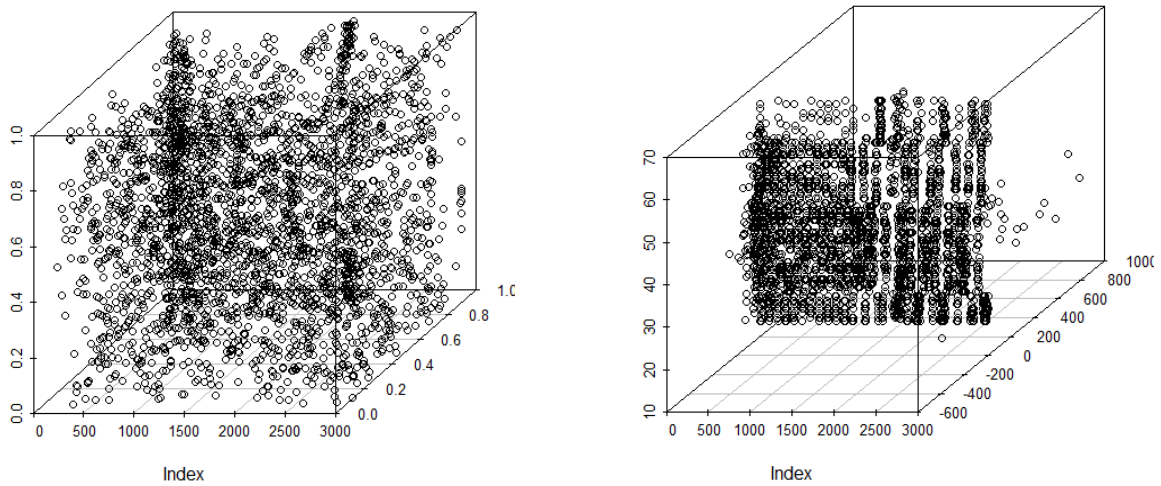


Fig. 1 – Distribuição randomizada de *pnvldni-idade* após quebra de ligações (Esquerda). Distribuição *pnvldni-idade* (Direita)

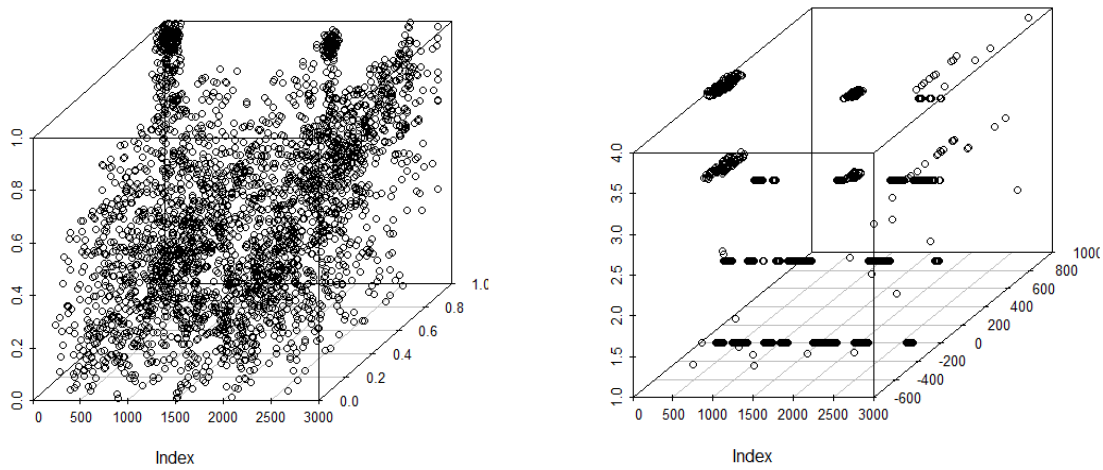


Fig. 2 – Distribuição randomizada de *pnvldni-scoring* após quebra de ligações (Esquerda). Distribuição *pnvldni-scoring* (Direita)

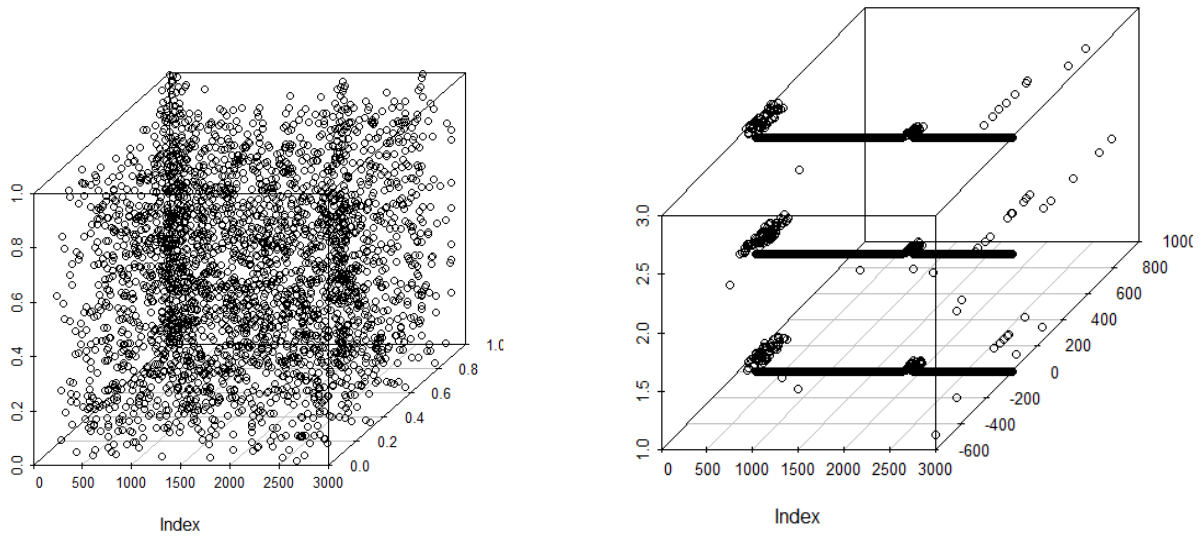


Fig. 3 – Distribuição randomizada de *pnivdni-regiao* após quebra de ligações (Esquerda). Distribuição *pnivdni-regiao* (Direita)

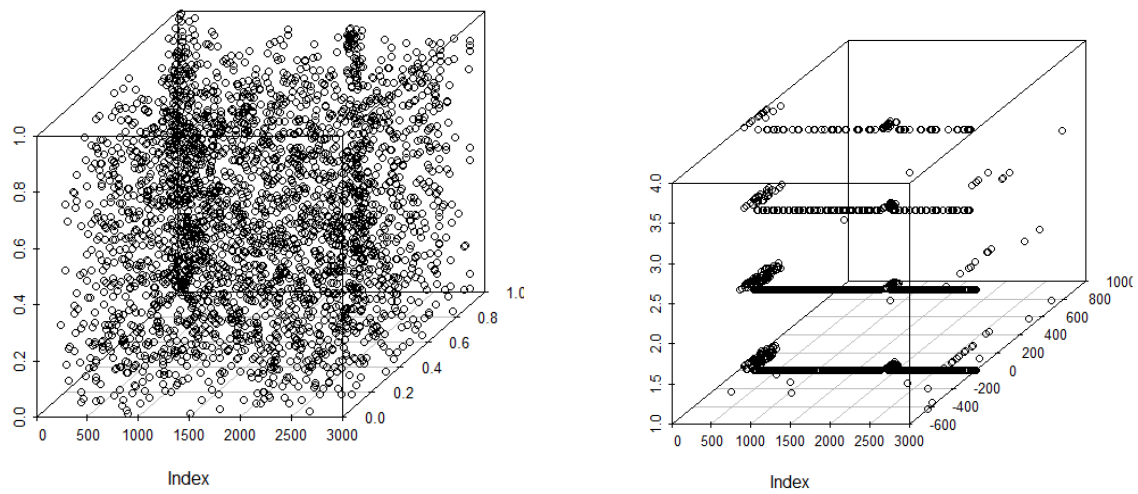


Fig. 4 – Distribuição randomizada de *pnivdni-civil* após quebra de ligações (Esquerda). Distribuição *pnivdni-civil* (Direita)

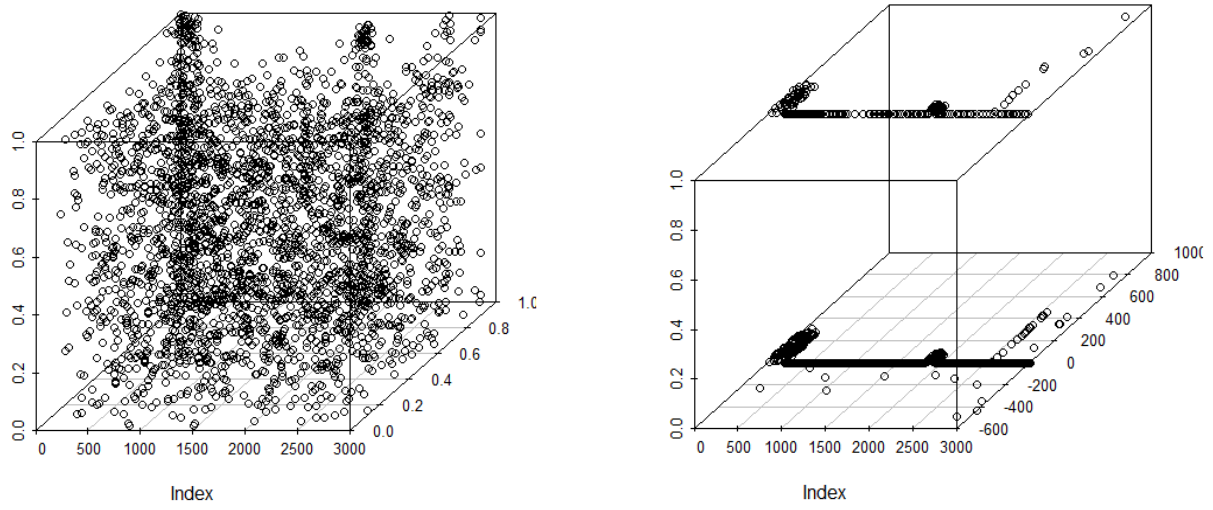


Fig. 5 – Distribuição randomizada de *pnv dni*-evento após quebra de ligações (Esquerda). Distribuição *pnv dni*-evento (Direita)

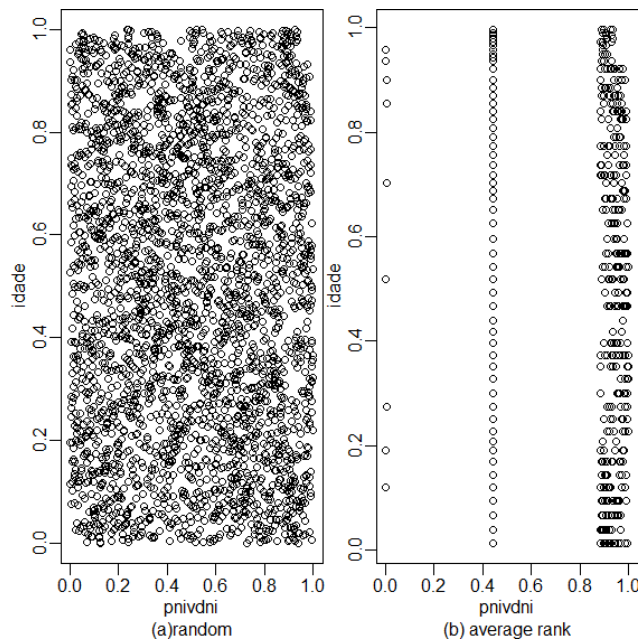


Fig 6 – Pseudo-Observações do binómio *pnv dni*-idade.

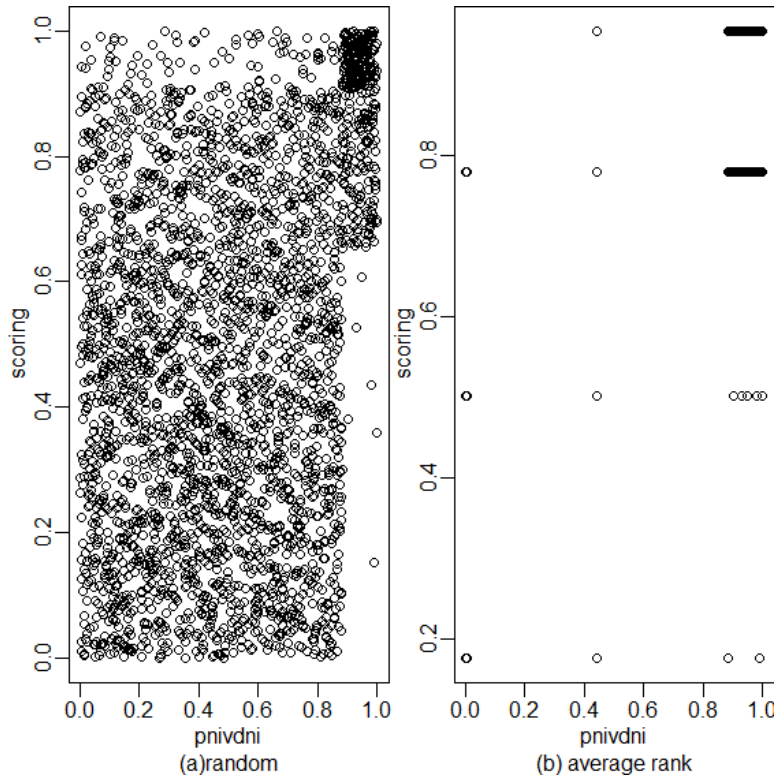


Fig 7 – Pseudo-Observações do binómio *pnivdni-scoring*.

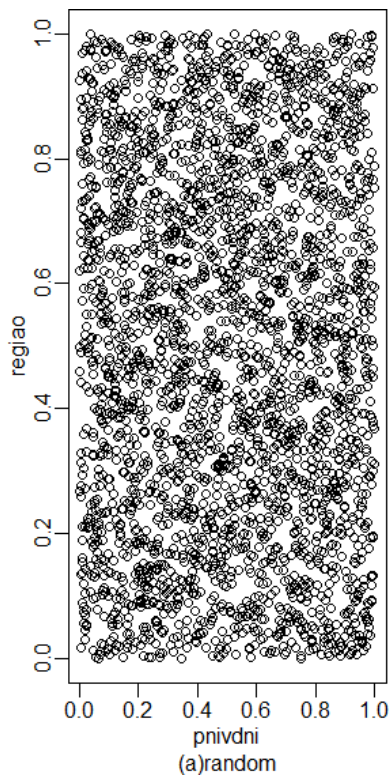


Fig 8 – Pseudo-Observações do binómio *pnivdni-região*.

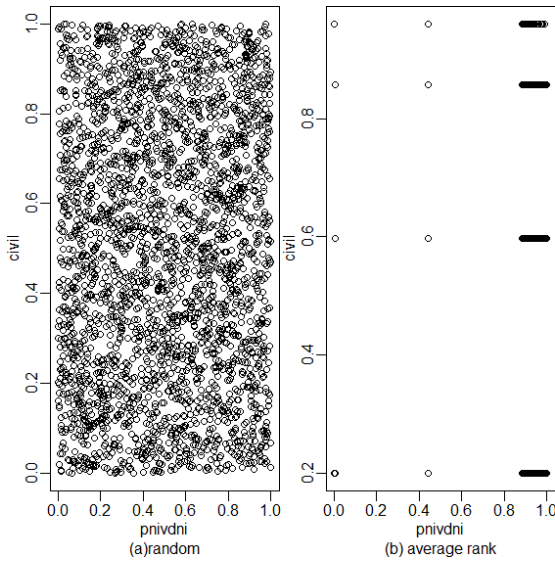


Fig 9 – Pseudo-Observações do binómio *pnivdni-civil*.

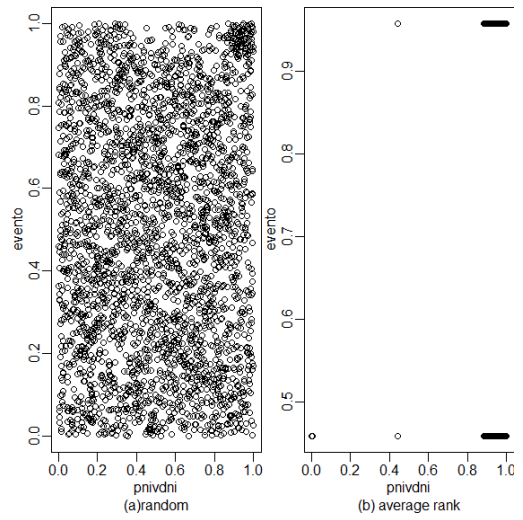


Fig 10 – Pseudo-Observações do binómio *pnivdni-evento*.

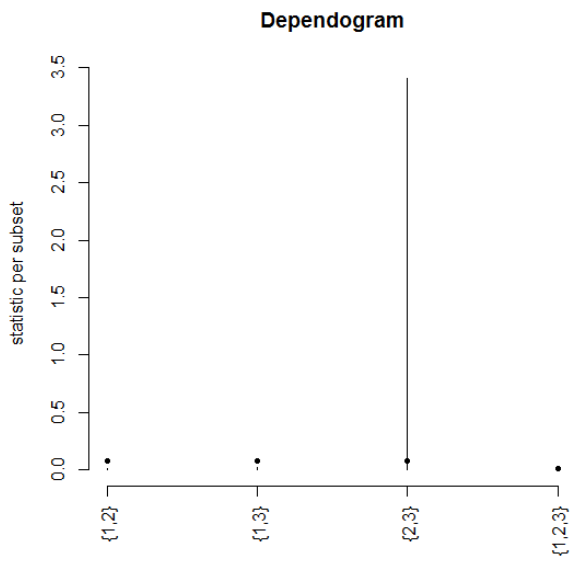


Fig 11 – Dependograma com os resultados dos testes de independência mútua para a distribuição multivariada *pnivdni-idade-civil*

Anexo 2

Taxas crédito consumo Banco Portugal

Por 1 milhão de euros	Particular	Particular	Particular	Particulares - Para consumo - Total	
Private individuals over EUR 1 million	Private individuals	Private individuals	Private individuals	Private individuals - consumer credit - Total	
	4	5	6	7	
				8	
2009 Jan	5.32	4.13	4.12	4.93	10.05
2009 Fev	4.48	3.45	3.40	4.16	9.91
2009 Mar	4.19	3.07	3.02	3.88	9.34
2009 Abr	3.96	2.81	2.77	3.52	8.92
2009 Mai	4.09	2.65	2.62	3.39	8.84
2009 Jun	3.71	2.57	2.54	3.23	8.57
2009 Jul	3.70	2.54	2.51	3.13	8.85
2009 Ago	3.14	2.39	2.35	2.94	8.57
2009 Set	3.42	2.32	2.29	2.88	8.82
2009 Out	3.54	2.28	2.24	2.84	8.69
2009 Nov	3.18	2.26	2.23	2.83	8.65
2009 Dez	3.27	2.25	2.22	2.81	8.06
2010 Jan	3.75	2.17	2.13	2.72	7.80
2010 Fev	3.15	2.20	2.16	2.80	8.20
2010 Mar	3.20	2.23	2.20	2.89	8.11
2010 Abr	3.63	2.26	2.23	3.01	7.99
2010 Mai	3.85	2.31	2.27	3.13	8.16
2010 Jun	3.82	2.28	2.25	3.37	8.72
2010 Jul	3.35	2.36	2.33	3.44	8.87
2010 Ago	3.62	2.58	2.54	3.64	9.11
2010 Set	4.28	2.68	2.65	3.76	9.10
2010 Out	3.75	2.73	2.69	3.79	9.08
2010 Nov	3.94	2.84	2.80	3.91	8.96
2010 Dez	4.57	2.98	2.96	4.08	8.39
2011 Jan	4.23	2.94	2.91	4.04	9.18
2011 Fev	4.35	3.05	3.03	4.13	9.51
2011 Mar	4.85	3.21	3.18	4.20	9.84
2011 Abr	4.88	3.44	3.41	4.46	9.85
2011 Mai	5.02	3.60	3.57	4.65	9.79
2011 Jun	5.16	3.76	3.74	4.82	9.80
2011 Jul	5.30	3.92	3.89	4.85	10.03
2011 Ago	5.81	4.15	4.11	5.10	10.35
2011 Set	6.07	4.20	4.16	5.16	10.45
2011 Out	6.12	4.27	4.24	5.25	10.38
2011 Nov	6.17	4.37	4.33	5.34	10.36
2011 Dez	5.88	4.28	4.25	5.21	10.04

Anexo 3

-

Exemplos de comandos utilizados – Package Copula R

Instalar package "xlsreadwrite"

```
library(xlsReadWrite)
```

```
xls.getshlib()
```

Instalar package "copula"

```
library(copula)
```

```
library(pspline)
```

Instalar package "scatterplot3d"

```
library(scatterplot3d)
```

```
banif<- read.xls("dados final.xls")
```

```
head(banif)
```

```
banif$civil <- factor(banif$civil, label = c("casado", "solteiro", "divorciado", "viuvo"),  
+ levels = 1:4)
```

```
banif$regiao <- factor(banif$regiao, label = c("porto", "lisboa",  
+ "outro"), lev = c(1, 2, 3))
```

```
nrow(banif)
```

```
apply (banif[, 1:6], 2, function(x) length(unique(x)))
```

```
mybanif <- subset(banif, select =c("pnivdni","idade"))
```

```
set.seed(123)
```

```
pseudobanif<- sapply(mybanif, rank, ties.method = "random") /  
+ (nrow(mybanif) + 1)
```

```
pseudobanif.ave<- sapply(mybanif, rank)/(nrow(mybanif) + 1)
```

```
par(mfrow = c(1,2), mar = c(2, 2, 1, 1), oma = c(1, 1, 0, 0),  
+ mgp = c(2, 1, 0))
```

```
plot(pseudobanif, sub = "(a)random")
```

```

plot(pseudobanif.ave, sub = "(b) average rank")

mybanif<- subset(banif, select =c("pnivdni", "scoring"))

pseudobanifsc<- sapply(mybanif, rank, ties.method = "random") /
  + (nrow(mybanif) + 1)

pseudobanifsc.ave<- sapply(mybanif, rank)/(nrow(mybanif) + 1)

plot(pseudobanif, sub = "(a)random")

plot(pseudobanif.ave, sub = "(b) average rank")

system.time (empsamp<- indepTestSim(nrow(pseudobanifida), p = 2,
  + N = 1000, print.every = 0))

indepTest(pseudobanifida, empsamp)

system.time(banifgof.gumbel.mult<- gofCopula(gumbelCopula(1), pseudobanif,
  + method = "itau", simulation = "mult", N = 1000))

banifgof.gumbel.mult

system.time(banifgof.clayton.mult<- gofCopula(claytonCopula(1),
  + pseudobanifsc, method = "itau", simulation = "mult", N = 1000))

banifgof.clayton.mult

system.time(banifgof.normal.mult<- gofCopula(normalCopula(0),
  + pseudobanifsc, method = "itau", simulation = "mult", N = 1000))

banifgof.normal.mult

system.time(banifgof.t.mult<- gofCopula(tCopula(0, df = 4, df.fixed = TRUE),
  + pseudobanifida, method = "itau", simulation = "mult", N = 1000))

banifgof.t.mult

Raciocinio:

fitCopula(gumbelCopula(1), pseudobanifida, method = "itau")

myAnalysis<- function(mybanif) {
+ pseudobanif<- sapply(mybanif, rank, ties.method = "random") /
+ (nrow(mybanif) + 1)
+ indTest<- indepTest(pseudobanif, empsamp)$global.statistic.pvalue

```

```

+ gof.g<- gofCopula(gumbelCopula(1), pseudobanif, method = "itau",
+ simulation = "mult")$pvalue
+ gof.c<- gofCopula(claytonCopula(1), pseudobanif, method = "itau",
+ simulation = "mult")$pvalue
+ gof.n<- gofCopula(normalCopula(0), pseudobanif, method = "itau",
+ simulation = "mult")$pvalue
+ gof.t <- gofCopula(tCopula(0, df = 4, df.fixed = TRUE), pseudobanif,
+ method = "itau", simulation = "mult")$pvalue
+ fit.g<- fitCopula(gumbelCopula(1), pseudobanif, method = "itau")
+ c(indep = indTest, gof.g = gof.g, gof.c = gof.c,
+ gof.n = gof.n, gof.t = gof.t, est = fit.g@estimate,
+ se = sqrt(fit.g@var.est)) + }
myReps<- t(replicate(100, myAnalysis(mybanif)))
round(apply(myReps, 2, summary), 3)
set.seed(123)
pseudobanif<- apply(banif[,1:6],2,rank)/(nrow(banif)+1)
system.time(empsamp<- indepTestSim(nrow(pseudobanif), p = 6, N = 1000,
+ print.every = 0))
srMultIndepTest<- indepTest(pseudobanif, empsamp)
srMultIndepTest
dependogram(srMultIndepTest)
system.time(srGof.t.mult<- gofCopula(tCopula(c(0,0,0), dim=3,
+ dispstr = "un", df = 4, df.fixed = TRUE), pseudobanif, method = "mpl",
+ simulation = "mult"))
srGof.t.mult
fitCopula(tCopula(c(0,0,0), dim = 3, dispstr = "un", df = 4,
+ df.fixed = TRUE), pseudobanif, method = "mpl")

```