

# Estimativas de *loss given default* em portfólios de crédito simulados

Herbert Kimura<sup>1</sup>  
Gustavo de Magalhães Rezende<sup>2</sup>

## RESUMO

O acordo de Basileia II permite que os bancos utilizem modelos internos que sirvam de base para o cálculo dos requisitos mínimos de capital em virtude do nível de exposição ao risco de crédito. Dentre os principais componentes estimados, estão a probabilidade de inadimplência (PD – *probability of default*), a perda dada a inadimplência (LGD – *loss given default*) e a exposição à inadimplência (EAD – *exposure at default*). Este estudo tem como objetivo investigar mecanismos de estimação de LGD utilizando modelos de regressão e transformações logit e Beta. Considerando que o mercado brasileiro ainda se encontra em um estágio incipiente na análise de LGD e a disponibilidade de dados de taxas de recuperação é restrita, serão simulados portfólios de crédito por meio de técnicas de Monte Carlo. A dependência entre LGD e as variáveis explicativas da fração de perdas em caso de inadimplência é modelada por meio de cópulas gaussianas e de matrizes *doubly stochastic*. Os resultados sugerem que a análise de regressão multivariada, usando diversas funções de transformação, possibilita identificar adequadamente as variáveis que explicam a LGD. Mais particularmente, os resultados indicam que o modelo de Hamerle, Knapp e Wildenauer (2009) para estimação de LGD apresenta desempenho superior ao do desempenho do modelo discutido por Gupton e Stein (2005).

**Palavras-chave:** Risco de crédito. Perda dada inadimplência. Acordo da Basileia. Matriz bi-estocástica. Cópulas gaussianas.

## 1 INTRODUÇÃO

A concessão de crédito é requisito essencial para o desenvolvimento de um país. Nos últimos anos, a oferta de crédito no Brasil tem aumentado de maneira significativa, influenciada por diversos fatores como, por exemplo, a queda nas taxas de juros, o aumento do prazo médio das operações e a estabilidade econômica.

Segundo dados do Banco Central do Brasil (BACEN, 2008, 2011, 2012), o estoque total de empréstimos bancários no sistema financeiro em relação ao PIB passou de 31,6% em 2007 para 47,2% em 2010, chegando a 53,2% em 2012. Essa ampliação de recursos destinados à concessão de crédito causa também o aumento dos riscos e a necessidade de maior controle por parte das instituições financeiras e do regulador do mercado.

<sup>1</sup> Doutor em Administração de Empresas. Universidade de Brasília - herbert.kimura@gmail.com

<sup>2</sup> Mestre em Administração de Empresas. Banco Panamericano - gustavo\_m.rezende@yahoo.com.br

Em se mantendo a tendência atual no aumento do volume de empréstimos, será cada vez mais importante para os bancos precificarem e gerenciarem o risco de crédito de maneira mais precisa, utilizando sistemas mais sofisticados de gestão. Adicionalmente, considerando a maior complexidade do mercado financeiro e o aumento do risco sistêmico, organismos internacionais e órgãos reguladores nacionais vêm desenvolvendo diretrizes e regras para mensuração e gestão do risco de crédito, agindo de modo preventivo para evitar crises econômicas de grandes proporções.

Em particular, a implantação do acordo de Basileia II tem sido para muitas instituições financeiras em todo o mundo um dos maiores e mais importantes desafios dos últimos anos no que concerne à gestão de riscos. Além da introdução dos sistemas internos de *ratings*, o acordo é também um estímulo ao aprimoramento das práticas de gestão de riscos adotadas pelos bancos, gerando, como consequência, o aumento da competitividade entre eles (MIU; OZDEMIR, 2009).

Neste contexto, o presente trabalho pretende estudar o comportamento das perdas em carteiras de crédito. Mais particularmente, utiliza técnicas de simulação de Monte Carlo e, além disso, realiza estimativas do parâmetro LGD (*Loss Given Default*) com dois diferentes modelos descritos na literatura e compara os resultados obtidos por cada um deles.

Dada a escassez de dados de perdas em carteiras de empréstimos disponíveis ao público, conforme apontam Dermine e Carvalho (2005), este estudo utilizou bases de dados simuladas com técnicas de Monte Carlo. Para a criação de dependência entre algumas variáveis explicativas aleatórias, foram empregadas/usadas cópulas gaussianas e, no processo de combinação das variáveis com as observações de LGD, foram utilizadas matrizes estocásticas duplas, que serão detalhadas na descrição da metodologia.

Seguindo o procedimento utilizado por Hlawatsch e Ostrowski (2011), primeiramente, os valores de LGD e das variáveis explicativas que compõem as bases de dados do estudo serão simulados. Em seguida, essas bases serão utilizadas para realização de estimativas do parâmetro LGD por meio: (i) do modelo desenvolvido por Hamerle, Knapp e Wildenauer (2009) e (ii) do LossCalc e da Moody's KMV, descritos por Gupton e Stein (2005). Finalmente, os resultados obtidos pelos modelos de estimação de LGD dentro dos portfólios de crédito resultantes serão comparados entre si.

## 2 REFERENCIAL TEÓRICO

O risco de crédito tem origem no fato de que a contraparte tomadora de um empréstimo pode não efetuar o pagamento de suas obrigações contratuais tanto por vontade própria quanto por sua incapacidade financeira (JORION, 2001). Segundo Hayden e Porath (2009), os modelos estatísticos de classificação de risco utilizam variáveis explicativas, representativas de características dos tomadores, que fornecem informação sobre a deterioração de um determinado crédito.

A despeito de instituições financeiras utilizarem, já há bastante tempo, modelos para gerenciamento do risco de crédito, a partir dos anos 1980, as maiores economias do mundo iniciaram esforços na tentativa de adaptar as diferentes regulamentações bancárias para um padrão internacional mais abrangente e comum entre os países.

Em 1988, o primeiro acordo de Basileia (Basileia I) estabeleceu recomendações ao sistema bancário mundial com os objetivos de reduzir os riscos do sistema bancário internacional e minimizar as desigualdades competitivas provenientes da arbitragem regulatória, isto é, das diferenças na alocação de capital exigido de bancos situados em países distintos (WAGSTER, 1996).

Apesar de representar um primeiro esforço para padronizar requisitos de capital e conferir maior segurança ao sistema financeiro, a Basileia I não evitou diversas crises financeiras da década de 90. Nesse sentido, o segundo acordo de Basileia (Basileia II) surge em 2004, com uma proposta final em 2006, com o objetivo principal de fortalecer a robustez e a estabilidade do sistema bancário internacional, por meio de novas recomendações que visavam (i) assegurar uma alocação de capital mais sensível ao risco; (ii) separar o risco de crédito do risco operacional; (iii) reduzir a arbitragem regulatória e (iv) encorajar os bancos a desenvolverem sistemas de gestão de riscos internos cada vez melhores e mais modernos (DIERICK et al, 2005).

Como resposta à crise financeira global iniciada em 2008 no mercado americano de empréstimos sub-prime, um terceiro acordo, Basileia III, foi emitido em 2010 com revisões em 2011 com a finalidade de incorporar aprimoramentos referentes a requisitos de capital e, em 2013, para contemplar aspectos de gestão de liquidez. Com relação especificamente a risco de crédito, uma das preocupações da Basileia III foi impor requisito de capital em razão de potenciais flutuações do risco de crédito da contraparte (BCBS, 2011)

No contexto de aperfeiçoamento dos modelos internos de gestão de riscos de crédito, um importante avanço ocorreu com a Basileia II, que começou a exigir requisitos de capital baseados em estimativas dos parâmetros de risco de crédito: (i) PD (*Probability of Default* ou probabilidade de inadimplência), representativa da chance de uma inadimplência ocorrer em um dado horizonte de tempo; (ii) LGD (*Loss Given Default* ou perda no caso de ocorrência de inadimplência), associada à fração da exposição, já descontadas as recuperações ocorridas, que serão perdidas caso o crédito entre em inadimplência; (iii) EAD (*Exposure At Default* ou exposição no momento da inadimplência) referente à quantidade devida pela contraparte no momento da inadimplência e (iv) M (*Maturity* ou maturidade), ou seja, o prazo efetivo de vencimento (SCHUERMANN, 2004; STEPHANOU; MENDOZA, 2005).

Enquanto a PD é bastante descrita e estudada na literatura de risco de crédito, o parâmetro LGD ainda tem sido pouco explorado (ALTMAN, 2008) e será, portanto, foco do presente trabalho. A seguir, são discutidos diferentes conceitos de LGD e apresentado um levantamento sobre pesquisas que estudam esse parâmetro de risco de crédito.

## 2.1 Definições de LGD

O parâmetro LGD representa o percentual do saldo do crédito que não é recuperado quando o tomador se torna inadimplente. Enquanto as entradas de caixa são consideradas recuperações e reduzem o saldo inadimplente, as saídas de caixa são os custos incorridos no processo de recuperação e aumentam o saldo. Por estarem distribuídos em momentos diferentes do tempo, as recuperações e os custos devem ser descontados, por uma taxa apropriada, até o momento da inadimplência (PETER, 2009).

Segundo Altman (2008), a LGD é expressa como sendo igual à unidade menos a taxa de recuperação, isto é, o percentual do que não foi recuperado após o momento de inadimplência. Peter (2009), por sua vez, expressa a LGD como um percentual da EAD em uma exposição de crédito e, para um dado produto  $j$  que esteja em inadimplência, seu valor é igual a:

$$LGD_j(t_{DF}) = \frac{EAD_j(t_{DF}) - NPV(Rec_j(t), t \geq t_{DF}) + NPV(Costs_j(t), t \geq t_{DF})}{EAD_j(t_{DF})}$$

onde NPV se refere ao valor presente líquido, Rec e Costs representam, respectivamente, todas as recuperações e os custos observados em dado momento  $t$ , e  $t_{DF}$  representa o instante da inadimplência.

De acordo com Peter (2009), a taxa de desconto utilizada pode afetar significativamente o resultado da perda econômica e de LGD. Schuermann (2004) sugere que a taxa de desconto deve representar o risco de um ativo semelhante. Neste contexto, a taxa livre de risco, por exemplo, não deveria ser utilizada. Nessa conceituação, as recuperações ocorridas após a inadimplência resultam da venda de colaterais, garantias, massa falida entre outros. Os custos ligados às exposições em inadimplência podem ser diretos ou indiretos, notadamente os custos legais, administrativos, as taxas de seguros, custos de estocagem, manutenção, reparos de ativos, entre outros (PETER, 2009).

O parâmetro LGD é, portanto, o valor de perda *ex-post* expresso como um percentual da EAD para uma exposição de crédito, caso o tomador do empréstimo esteja em inadimplência. No caso de um tomador que ainda não esteja em inadimplência, a LGD é a estimativa *ex-ante* da perda, também expressa como um percentual da EAD, representada por uma variável aleatória que deve ser estimada (BENNETT; CATARINEU; MORAL, 2005).

Para o cálculo de LGD das exposições que não estejam em inadimplência, existem métodos subjetivos, baseados na experiência e julgamento por parte dos especialistas diretamente envolvidos no processo, e métodos objetivos, que utilizam dados de perdas realizadas para o desenvolvimento de modelos (BENNETT; CATARINEU; MORAL, 2005).

Os métodos objetivos podem ser explícitos ou implícitos. Segundo Peter (2009), os métodos explícitos são aqueles em que as informações analisadas em bases de dados permitem o cálculo direto da LGD. São dois os métodos utilizados (BENNETT; CATARINEU; MORAL, 2005): (i) *market* LGD, abordagem que se baseia na observação dos preços de mercado de títulos ou empréstimos negociáveis, logo após a inadimplência; (ii) *workout* LGD, abordagem baseada no desconto dos fluxos de caixa resultantes do processo de recuperação desde a data da inadimplência até o fim do período de recuperação de crédito.

Já nos métodos implícitos, de acordo com Bennett, Catarineu e Moral (2005), os valores de LGD são derivados de perdas e estimativas de PD, e não podem ser calculados diretamente pelas informações existentes nas bases de dados. Existem dois métodos implícitos distintos utilizados: (i) *implied market* LGD – abordagem na qual as estimativas de LGD são derivadas dos preços de mercado de títulos com risco que

não estejam em inadimplência por meio de um modelo de precificação de ativos; (ii) *implied historical LGD* – que consiste na inferência da LGD pelas perdas realizadas e estimativas de PD (BENNETT; CATARINEU; MORAL, 2005). Segundo Peter (2009), a abordagem implícita de mercado é útil em portfólios com poucas observações de inadimplência, enquanto a abordagem histórica implícita é utilizada em carteiras do varejo que, notadamente, possuem uma quantidade maior de eventos de inadimplência.

Segundo Miu e Ozdemir (2006), a Basileia II requer a utilização do valor de LGD em momento de recessão econômica (*downturn*) ou no pior momento de um ciclo econômico, quer dizer, o máximo entre a LGD média ponderada de longo prazo (*long-run default-weighted average*) e a LGD estressada (*stressed LGD*). Além disso, a Basileia II estabelece que os bancos (i) identifiquem as condições apropriadas de recessão e as dependências adversas, se existirem, entre as taxas de inadimplência e as taxas de recuperação e (ii) incorpore-as para produzir parâmetros de LGD para as exposições dos bancos, que sejam consistentes com as condições de recessão identificadas (MIU; OZDEMIR, 2006).

Além disso, existem duas filosofias de LGD que os bancos podem adotar. Na filosofia conhecida como PIT (*point-in-time*), comumente utilizada pelos bancos, a LGD é uma medida cíclica que reflete a LGD esperada, normalmente, nos doze meses seguintes. Em contrapartida, na filosofia denominada TTC (*through-the-cycle*), a LGD estimada é uma medida acíclica definida pela média no ciclo econômico e relativamente constante ao longo do ciclo (MIU; OZDEMIR, 2006).

## 2.2 Estudos sobre LGD

Em abordagens avançadas de risco de crédito, o órgão regulador estabelece que instituições financeiras possam utilizar *ratings* internos para cálculo da exigência de capital. No caso específico da LGD, os bancos devem utilizar modelos próprios para gerar estimativas. De acordo com Jacobson, Linde e Roszbach (2006), os *ratings* de crédito internos resumem as propriedades de risco da carteira de empréstimos de um banco e são utilizados para gerenciamento de seu risco.

Em estudo realizado por Altman e Kishore (1996), foram investigados mais de setecentos títulos em inadimplência entre os anos de 1978 e 1995 e concluiu-se que a senioridade tem um papel importante na recuperação das perdas. Acharya, Bharath e Srinivasan (2003) reportaram, em pesquisa realizada no período de 1982 a 1999, uma taxa de recuperação média de 48% para títulos de dívida sênior com garantia e 51% para títulos de dívida sênior sem garantia. Além disso, verificaram que a recuperação dos empréstimos é afetada não somente pela senioridade do título ou por seu nível de segurança, mas também pelas condições da indústria no momento da inadimplência.

Araten, Jacobs Jr. e Varshney (2004) encontraram em seu estudo com 3.761 títulos em inadimplência entre os anos de 1982 e 1999, que o valor médio da LGD foi de 39,8%, com uma taxa anual de desconto de 15%. Também, o estudo verificou que a LGD é sensível à taxa de desconto e sua distribuição é bimodal.

Dermine e Carvalho (2005) estimaram a LGD em uma amostra com 374 empréstimos corporativos de um banco europeu entre os anos de 1995 e 2000, por meio/por intermédio do valor descontado dos fluxos de caixa após a inadimplência (*workout LGD*), e encontraram uma recuperação média de 71% nos empréstimos ruins ou duvidosos. Além disso, reportaram que a frequência de distribuição da LGD parece ser bimodal.

Caselli, Gatti e Querci (2008) utilizaram uma amostra de 11.649 financiamentos imobiliários e empréstimos a pequenas e médias empresas, no mercado italiano, e apontaram diversas variáveis macroeconômicas explicativas para os referidos tipos de empréstimos. Estimaram ainda a LGD tomando por base as recuperações ocorridas pelos fluxos de caixa descontados.

Em estudo realizado com dados de cartões de crédito no Reino Unido, Bellotti e Crook (2012) usaram alguns modelos de estimação de LGD, focando na identificação de variáveis importantes que explicam as perdas nas carteiras de crédito estudadas como, por exemplo, o tempo de conta, o *rating*, a renda e a idade dos clientes. Além disso, os autores verificaram que a inclusão de variáveis explicativas macroeconômicas e estatisticamente significantes nos modelos como, por exemplo, as taxas de juros praticadas pelos bancos e o nível de desemprego do país, melhorou o desempenho dos modelos de previsão de inadimplência. Dentre os modelos estatísticos empregados, estão os de árvore de decisão, Tobit e regressões mediante/por meio do método dos mínimos quadrados ordinários (OLS – *Ordinary Least Squares*).

Diversos autores apontam a independência entre PD e LGD como um fator questionável. Estudos realizados por Carey (1998), Frye (2000) e Hu e Perraudin (2002), entre outros, mostraram que os parâmetros PD e LGD são positivamente correlacionados e que, se tal dependência não for levada em conta, os cálculos de perdas das carteiras de empréstimos e o capital que deveria ser alocado estarão incorretos. Em particular, quando a PD é alta, a LGD também é alta, o que sugere a existência de um risco sistêmico

afetando ambos os parâmetros (HUANG; OOSTERLEE, 2011).

Alguns modelos que levam em consideração a dependência entre PD e LGD foram desenvolvidos nos últimos anos. De acordo com Frye (2000), as inadimplências são guiadas por um único fator sistêmico – o estado da economia – em vez de diversos parâmetros correlacionados. Análises empíricas sugerem ainda que, em momentos de recessão exacerbada, a recuperação dos títulos de dívida pode ser reduzida de 20 a 25% do valor de sua média anual. Pykhtin (2003), Dullmann e Trapp (2004), Rosch e Scheule (2005) também propõem modelos estruturais que levam em conta a relação da LGD com o risco sistêmico, conseguindo identificar o ciclo econômico e capturar os valores mais altos de LGD, oriundos do pior momento do ciclo.

No Brasil, Silva, Marins e Neves (2009) realizaram trabalho sobre perdas em operações prefixadas no mercado brasileiro, com base em informações existentes no Sistema de Informações de Crédito (SCR) do BACEN, em uma amostra de 9.557 operações de não varejo prefixadas. O estudo identificou que a LGD de sua amostra de dados apresentava uma distribuição bimodal e encontrou valores médios de LGD entre 47% e 92%, reportando variáveis que influenciaram as perdas estimadas no estudo como, por exemplo, o nível de atividade econômica, a presença de garantia e o valor da operação, a existência de renegociação, entre outras.

### 3 METODOLOGIA

A metodologia do presente trabalho assemelha-se à desenvolvida por Hlawatsch e Ostrowski (2011) para a simulação de portfólios de crédito, incorporando modelos de estimação de LGD baseados em Hamerle, Knapp e Wildenauer (2009), e em Gupton e Stein (2005). Em resumo, a metodologia segue três etapas principais: (i) desenvolvimento de portfólios de crédito utilizando simulação de Monte Carlo, cada um contendo 10.000 observações de LGD, seguindo o procedimento de Hlawatsch e Ostrowski (2011) e de quatro variáveis explicativas de clientes em inadimplência; (ii) realização de estimativas de LGD das carteiras simuladas, utilizando dois modelos descritos na literatura (HAMERLE; KNAPP; WILDENAUER, 2009; GUPTON; STEIN, 2005) e (iii) comparação dos resultados obtidos com a aplicação dos dois modelos de estimação de LGD.

Mais especificamente, o processo de simulação de dados envolve (i) a simulação de carteiras de crédito com o uso de cópulas gaussianas para a modelagem de dependência entre as variáveis associadas à inadimplência e (ii) a combinação das variáveis de crédito da pesquisa, por meio do confronto entre percentis da variável LGD e das variáveis independentes, levando-se em consideração uma matriz bi-estocástica como referência. A estimação do modelo de LGD por meio/por intermédio/com o uso dos dados gerados por simulação de Monte Carlo é realizada usando-se regressão linear múltipla, com transformações logística e Beta.

A simulação de dados é fundamental no desenvolvimento deste artigo, dadas as dificuldades na obtenção de dados reais de LGD em carteiras de crédito no Brasil. Adicionalmente, o uso de simulação possibilita que o fenômeno em estudo seja investigado sem a influência de variáveis exógenas que, comumente, são de difícil controle em estudos empíricos com dados reais de mercado. Assim sendo, o presente estudo deve resultar em uma ferramenta teórica importante de análise de carteiras para as instituições financeiras, que poderão controlar melhor suas perdas esperadas e inesperadas, seguindo ao mesmo tempo as diretrizes do acordo de Basileia II e as regulamentações do BACEN.

#### 3.1 Simulação dos portfólios de crédito

Esta primeira etapa da metodologia pode ser subdividida em três partes: (i) simulação das observações de LGD, (ii) simulação das variáveis explicativas utilizando técnicas de Monte Carlo e (iii) junção ou combinação das observações de LGD e das variáveis explicativas para formar diferentes carteiras de crédito em inadimplência.

##### 3.1.1 Simulação das observações de LGD

Alguns autores como, por exemplo, Altman e Kalotay (2010) e Gupton e Stein (2005), descrevem a distribuição Beta como uma boa aproximação da distribuição de LGD em carteiras de crédito, tanto por sua característica bimodal quanto por pertencer ao intervalo entre zero e a unidade. No presente estudo, opta-se por utilizar uma mistura de duas distribuições Beta para simular observações de LGD, visando estabelecer uma distribuição bimodal mais evidente, seguindo evidências empíricas de Araten, Jacobs Jr. e

Varshney (2004), e Dermine e Carvalho (2005). Além disso, são definidos intervalos para o valor esperado, a variância, os parâmetros  $\alpha$  e  $\beta$ , e a ponderação de cada uma das distribuições.

Dessa maneira, é possível simular diversas distribuições de LGD diferentes formadas pela mistura de duas distribuições originais. Neste estudo, as simulações de LGD foram realizadas por meio de rotinas programadas na linguagem VBA do Excel. Embora na prática, em casos especiais, a LGD possa assumir valores menores que zero e maiores que a unidade, o mecanismo de simulação não permitirá a ocorrência desses eventos. Por exemplo, nos casos em que não existe recuperação alguma após a inadimplência e somente custos relacionados à cobrança, a LGD pode ser maior que 1. Por outro lado, se a recuperação é maior que a EAD, a LGD pode ser negativa.

Para a criação das observações de LGD, são definidas, primeiramente, duas variáveis aleatórias com distribuição Beta. A primeira delas,  $X_1$ , possui assimetria à direita e valor esperado  $E(X_1)$  pertencente ao intervalo entre 0,059 e 0,3; enquanto a segunda, denotada por  $X_2$ , possui assimetria à esquerda e valor esperado  $E(X_2)$  pertencente ao intervalo entre 0,7 e 0,941. Além disso, seguindo Hlawatsch e Ostrowski (2011), ambas as variáveis possuem variância de no mínimo 0,003, para evitar que as observações de LGD se concentrem próximas de 0 ou de 1. O fator  $w$  pondera a distribuição  $X_1$  e é um valor escolhido aleatoriamente entre 0,5 e 1. Os parâmetros  $\alpha$  e  $\beta$  das variáveis aleatórias  $X_1$  e  $X_2$  são obtidos como função do valor esperado e da variância das distribuições Beta.

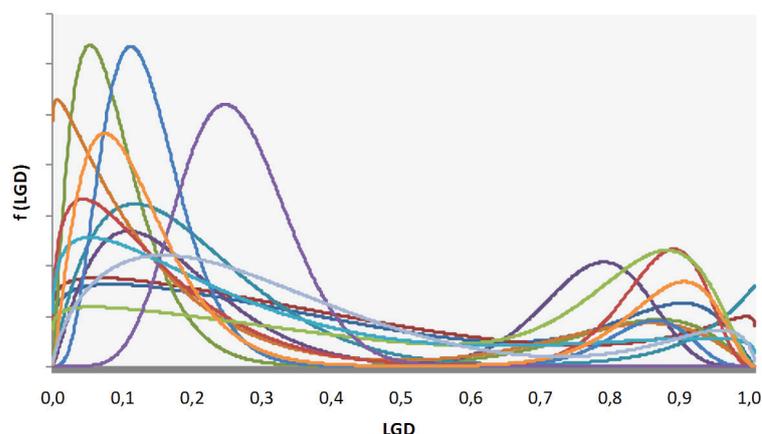
Definidos os intervalos necessários, as duas distribuições Beta devem ser combinadas para a geração das 10.000 observações de LGD mediante/com simulações de Monte Carlo. Para tanto, foi utilizado um método numérico de geração de números aleatórios, semelhante ao de Hlawatsch e Ostrowski (2011). Inicialmente, divide-se o intervalo entre zero e a unidade em 10.001 pontos equidistantes para a obtenção de 10.000 observações diferentes com valor igual a 1/10.000 cada um.

Em seguida, para cada observação, calcula-se o valor acumulado das observações até aquele ponto, ou seja, a décima e a milésima observação, por exemplo, possuem valor acumulado de 10/10.000 e 1.000/10.000, respectivamente. Esse valor acumulado será utilizado no cálculo da distribuição Beta acumulada de cada uma das variáveis aleatórias,  $X_1$  e  $X_2$ , que serão ponderadas por  $w$  e  $1 - w$ , respectivamente, e resultarão na mistura de distribuições Beta. A mistura de distribuições conduz a uma nova função de probabilidade acumulada dada por  $F(v_i) = w \times F_1(v_i) + (1 - w) \times F_2(v_i)$

Finalmente, são gerados 10.000 números aleatórios uniformes e comparados, um a um, aos valores da mistura da distribuição Beta acumulada. Assim, a LGD será igual ao valor da observação, não acumulada, na qual a variável aleatória se iguala ao valor acumulado da mistura da distribuição Beta. Matematicamente, o valor de LGD será igual ao inverso da função da mistura da distribuição Beta acumulada de cada valor gerado aleatoriamente.

A Figura 1 mostra possíveis distribuições de probabilidade de LGD em carteiras de crédito simuladas por Monte Carlo, variando entre zero e um.

Figura 1 – Possíveis distribuições de LGD simuladas via Monte Carlo



Fonte: Os autores (2013)

A criação das observações de LGD neste estudo foi realizada pela combinação de duas distribuições obtidas aleatoriamente pelo procedimento descrito. Na etapa da simulação a seguir, serão criadas quatro variáveis explicativas pertencentes às bases de dados aqui utilizadas.

### 3.1.2 Simulação de variáveis explicativas com dependência

As carteiras de crédito que serão simuladas devem conter, cada uma delas, quatro variáveis explicativas fictícias,  $A$ ,  $B$ ,  $C$  e  $D$ , com diferentes distribuições umas das outras. A variável  $A$  possui distribuição Beta com valores de  $\alpha = \beta = 5$ . A variável  $B$ , por sua vez, possui distribuição normal com média 0,05 e desvio-padrão igual a 0,2. A variável  $C$  tem distribuição de Bernoulli, podendo ser igual a zero ou à unidade, com probabilidade  $p = 0,7$  de pertencer a uma das classes. Finalmente, a variável  $D$  possui distribuição Beta com  $\alpha = 2$  e  $\beta = 10$ .

Essas variáveis representam potenciais características de tomadores conforme sugeridas em estudos de Altman e Kishore (1996), Dermine e Carvalho (2005), Hamerle, Knapp e Wildenaer (2009), Peter (2009) como, por exemplo, existência de garantias ou colaterais, senioridade do crédito, idade, saldo devedor, tempo de relacionamento entre banco e cliente, alavancagem do tomador, etc. As diferentes distribuições para variáveis  $A$ ,  $B$ ,  $C$  e  $D$  possibilitam modelar diferentes características dos tomadores.

Sob o aspecto de simulação, as variáveis  $B$ ,  $C$  e  $D$  serão geradas para serem correlacionadas negativamente à LGD, ao contrário de  $A$ , que terá correlação positiva. Entre si, as variáveis  $B$ ,  $C$  e  $D$  deverão manter alguma dependência. Já a variável  $A$  será simulada de forma independente das outras variáveis. Os parâmetros de todas as variáveis criadas seguem os valores estipulados por Hlawatsch e Ostrowski (2011).

A dependência entre as variáveis  $B$ ,  $C$  e  $D$  será formada por meio de uma cópula gaussiana tridimensional. De acordo com Nelsen (1999), cópulas são funções que agrupam funções de distribuição multivariadas em funções de distribuição marginais uniformes pertencentes a uma única dimensão. A relação entre cópulas e funções de distribuição de variáveis aleatórias é garantida pelo teorema de Sklar (1973), que estabelece que, dada uma função de distribuição  $G$  com  $n$  dimensões e marginais contínuas  $F_1, \dots, F_n$ , existe uma cópula  $C: [0, 1]^n \rightarrow [0, 1]$ , com  $n$  dimensões, de tal forma que  $G(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n))$ .

Destaca-se que a teoria de cópulas vem sendo bastante empregada como ferramenta para a modelagem multivariada em Finanças, Ciências Atuariais, entre outras, visto que muitas relações de dependência entre as variáveis não são totalmente lineares. Na gestão de risco de crédito, como exemplo, as cópulas são utilizadas na modelagem de *credit scoring* e da probabilidade de inadimplência (YAN, 2007).

Para a geração de números aleatórios com dependência, o procedimento usado consiste na criação de uma matriz de correlação  $M$ , positiva semidefinida, entre as três variáveis. Em seguida, são geradas três variáveis aleatórias com distribuição normal padrão  $Z_1, Z_2$  e  $Z_3$  que devem ser transformadas em três variáveis dependentes  $X_1, X_2$  e  $X_3$ , pela decomposição de Cholesky da matriz de correlação  $M$ . Neste estudo, serão criados alguns cenários de dependência entre as variáveis, nos quais serão alterados os valores da correlação na matriz, para verificar se há diferenças na estimação de LGD dependendo da correlação existente.

As variáveis dependentes criadas devem, então, ser acumuladas na função de distribuição normal padrão, para resultar em três variáveis aleatórias dependentes uniformemente distribuídas,  $U_1, U_2$  e  $U_3$ , que são utilizadas como *inputs* para a geração das observações marginais da cópula pelo método de inversão. Vale notar que as observações marginais criadas correspondem às três variáveis explicativas fictícias e dependentes nas bases de dados simuladas, ou seja, a variável  $U_1$  dará origem a  $B$ ,  $U_2$  a  $C$  e  $U_3$  a  $D$ . A geração da variável  $A$  se dará diretamente pelo método de inversão, uma vez que ela independe das outras variáveis. Por fim, o procedimento descrito deve ser repetido até que sejam criadas 10.000 observações de cada um dos parâmetros explicativos.

A parte final da simulação dos portfólios de crédito, descrita a seguir, consiste na combinação entre as observações simuladas de LGD e das variáveis explicativas.

### 3.1.3 Correspondência de LGD com as variáveis explicativas

A etapa final do trabalho utiliza dois procedimentos pseudoaleatórios distintos, nos quais as observações de cada uma das variáveis  $A$ ,  $B$ ,  $C$  e  $D$  serão combinadas às observações de LGD de forma diferente. No desenvolvimento dessa etapa, foram criadas rotinas implementadas em linguagem de programação no VBA do Excel.

As observações das quatro variáveis explicativas e de LGD devem ser ordenadas para a criação de quintis. Na sequência, utilizando-se uma matriz conhecida como biestocástica (*bistochastic*) ou estocástica dupla (*doubly-stochastic matrix*), em que a soma de suas linhas e colunas é igual à unidade, as observações de LGD são combinadas às variáveis  $A$ ,  $B$ ,  $C$  ou  $D$ , separadamente. O procedimento envolve a combinação de observações pertencentes aos quintis de cada variável e de LGD. Para melhor compreensão, será descrita a metodologia utilizada na combinação entre a variável  $A$  e LGD, que deve ser repetida para as outras variáveis,

de acordo com suas matrizes bi-estocásticas.

Primeiramente, define-se uma matriz estocástica dupla que determinará a probabilidade com que as observações pertencentes a cada quintil da variável  $A$  serão combinadas observações dos quintis do parâmetro LGD. Trata-se de uma matriz quadrada  $E$  de ordem 5, em que cada elemento  $m_{ij}, i, j = 1, \dots, 5$  descreve a probabilidade de uma observação do quintil  $j$  da variável  $A$  se combinar a uma observação do quintil  $i$  de LGD. As matrizes utilizadas no estudo para as variáveis  $B, C, e D$ , tanto na base de dados “boa” quanto na base de dados “ruim”, foram retiradas de Hlawatsch e Ostrowski (2011); enquanto as matrizes usadas para a variável  $A$  resultaram de elaboração própria. As Figuras 2 a 5 mostram as matrizes bi-estocásticas utilizadas. A base de dados “boa” é formada por uma matriz com altas probabilidades de combinação em suas diagonais, ao contrário da base de dados “ruim”, com probabilidades menores nas diagonais.

Uma vez definidas as matrizes, a primeira observação do primeiro quintil de LGD deve ser selecionada. Em seguida, gera-se um número aleatório uniforme  $N_R$  que é comparado à primeira linha da matriz de probabilidade acumulada de  $E$ . Por exemplo, supondo que o número seja 0,75, a primeira observação de LGD será combinada a uma variável escolhida de forma aleatória e pertencente ao (i) primeiro quintil da base de dados “boa” ou (ii) terceiro quintil da base de dados “ruim”. Depois de combinada, essa observação da variável  $A$  deve ser excluída do conjunto de observações iniciais, para que não venha a ser escolhida novamente. Tal procedimento foi repetido até que todas as observações da variável  $A$  estivessem combinadas às observações de LGD. Um dos problemas ocorridos durante o processo foi a escolha de números aleatórios que apontaram quintis sem observações da variável  $A$ . Hlawatsch e Ostrowski (2011) sugerem a seleção do próximo quintil plausível. No entanto, a solução utilizada no estudo foi diferente, sendo gerados tantos números aleatórios quantos foram necessários até que um quintil com observações de  $A$  fosse selecionado.

Figura 2 – Matriz estocástica dupla da base “boa” para a variável  $A$

85,0%	5,0%	5,0%	3,0%	2,0%
5,0%	85,0%	6,0%	3,0%	1,0%
5,0%	6,0%	86,0%	2,0%	1,0%
3,0%	3,0%	2,0%	85,0%	7,0%
2,0%	1,0%	1,0%	7,0%	89,0%

Fonte: Os autores (2013)

Figura 3 – Matriz estocástica dupla da base “boa” para as variáveis  $B, C e D$

2,1%	2,4%	2,6%	2,9%	90,0%
1,3%	1,5%	3,6%	90,0%	3,6%
3,0%	3,3%	87,5%	3,3%	3,0%
3,6%	90,0%	3,6%	1,5%	1,3%
90,0%	2,9%	2,6%	2,4%	2,1%

Fonte: Os autores (2013)

Figura 4 – Matriz estocástica dupla da base “ruim” para a variável A

55,0%	15,0%	10,0%	12,0%	8,0%
15,0%	45,0%	25,0%	10,0%	5,0%
10,0%	25,0%	40,0%	15,0%	10,0%
12,0%	10,0%	15,0%	43,0%	20,0%
8,0%	5,0%	10,0%	20,0%	57,0%

Fonte: Os autores (2013)

Figura 5 – Matriz estocástica dupla da base “ruim” para as variáveis B, C e D

2,5%	5,0%	10,0%	20,0%	62,5%
5,0%	10,0%	20,0%	45,0%	20,0%
10,0%	20,0%	40,0%	20,0%	10,0%
20,0%	45,0%	20,0%	10,0%	5,0%
62,5%	20,0%	10,0%	5,0%	2,5%

Fonte: Os autores (2013)

### 3.2 Modelos de estimação de LGD

A segunda etapa da metodologia consiste na estimação de LGD das carteiras simuladas previamente, utilizando dois modelos encontrados na literatura. Uma vez que os valores de LGD variam entre 0 e 1, não é possível realizar estimativas por/com o uso de regressão linear, pois estimativas geradas pela técnica podem não pertencer a esse intervalo. Assim, Hamerle, Knapp e Wildenauer (2009) descrevem um procedimento de transformação do tipo Logit nas observações de LGD da base de dados de desenvolvimento, a fim de que sejam realizadas as estimativas dos valores de LGD, com a seguinte relação:

$$y_{t(i)} = \ln \frac{LGD_{t(i)}}{1 - LGD_{t(i)}}$$

Em seguida, os coeficientes de uma regressão linear são obtidos para cada uma das variáveis e, assim, os valores transformados de LGD,  $y_{t(i)}$ , podem ser estimados. Finalmente, é necessário transformar os valores de  $y_{t(i)}$  em observações de LGD, de maneira que:

$$LGD_{t(i)} = \frac{\exp(y_{t(i)})}{1 + \exp(y_{t(i)})}$$

Para os dados deste artigo, serão realizadas regressões lineares dos parâmetros A, B, C e D contra os valores transformados das observações de LGD, utilizando-se o método dos mínimos quadrados ordinários (OLS).

O segundo modelo de estimação de LGD, baseado no LossCalc, da Moody's KMV, é descrito por Gupton e Stein (2005). O LossCalc considera a taxa de recuperação no momento da inadimplência como uma variável estocástica independente da PD e igual a  $1 - LGD$ . Essa estimativa é utilizada por investidores e instituições financeiras nas estimativas de LGD para inadimplências ocorridas imediatamente ou que venham a ocorrer dentro de um ano.

Gupton e Stein (2005) observaram que as recuperações das perdas existentes na base de dados utilizada no desenvolvimento do modelo não possuíam uma distribuição normal. Verificaram também que uma distribuição Beta poderia ser utilizada como uma boa alternativa de aproximação das recuperações

e que, transformando tais dados em valores equivalentes de uma distribuição normal, eles se tornariam “melhor comportados” para serem modelados. A transformação e a modelagem dos dados foram feitas para cada tipo de dívida: empréstimos, títulos e ações preferenciais.

Embora existam outros modelos que utilizam a distribuição Beta em seu desenvolvimento, como Gordy e Jones (2002), Ivanova (2004), Onorato e Altman (2005) e Pesaran et al (2006), há alguns críticos desse tipo de aproximação nas estimativas de LGD. Por exemplo, Dermine e Carvalho (2005) afirmam que os modelos de empréstimos baseados em estimativas de recuperação fixas ou em distribuições Beta não capturam a característica bimodal das perdas em portfólios de crédito. Baixauli e Alvarez (2009) avaliam em seu estudo, por sua vez, que o desempenho dos modelos que utilizam como aproximação de suas perdas a distribuição Beta não é satisfatório para a construção de intervalos analíticos de previsão.

Partindo das observações de LGD nas bases de dados simuladas, que possuem distribuição Beta, característica bimodal e variam entre 0 e 1, é necessário criar uma variável dependente normalmente distribuída por uma transformação, adotando a expressão a seguir:

$$Y_i = N^{-1}(B(LGD, \alpha, \beta))$$

em que  $N^{-1}$  é a inversa da distribuição normal acumulada,  $B$  é a distribuição Beta acumulada,  $LGD$  é o valor da variável na base de dados que será transformada e  $\alpha$  e  $\beta$  são os parâmetros da distribuição Beta combinada de LGD.

Em seguida, é realizada uma regressão linear para estimar os coeficientes apropriados dos fatores preditivos  $A$ ,  $B$ ,  $C$  e  $D$ , tendo como variável dependente  $Y$ . Como os valores de  $Y$  pertencem ao espaço normal, a etapa final consiste em aplicar a transformada inversa da distribuição Beta para levar os valores de  $Y$  do “espaço normal” para o “espaço de LGD”, obtendo, dessa maneira, os resultados estimados de LGD para cada exposição. Assim:

$$LGD = B^{-1}(N(Y_i, \mu, \sigma))$$

em que  $B^{-1}$  é a inversa da distribuição Beta acumulada,  $N$  é a distribuição normal acumulada,  $Y$  é o valor da variável transformada e  $\mu$  e  $\sigma$  são a média e o desvio-padrão, respectivamente, das observações de LGD da base de dados.

Uma vez que os valores de  $\alpha$  e  $\beta$  da distribuição de LGD não são conhecidos, é necessário estimá-los. Gupton e Stein (2005) utilizam a média ( $\mu$ ) e o desvio-padrão ( $\sigma$ ) de LGD para os cálculos dos parâmetros  $\alpha$  e  $\beta$ , seguindo as expressões:

$$\alpha = \frac{\mu^2(1-\mu)}{\sigma^2} - \mu \quad \text{e} \quad \beta = \alpha \left( \frac{1}{\mu} - 1 \right)$$

#### 4 ANÁLISE DE RESULTADOS

As análises nas carteiras de crédito simuladas foram realizadas utilizando-se programação no *software* R. Os resultados foram obtidos em bases de dados geradas aleatoriamente por meio de regressões lineares dos parâmetros  $A$ ,  $B$ ,  $C$  e  $D$  contra os valores transformados das observações de LGD, seguindo a metodologia de simulação descrita anteriormente. São alterados também os valores das correlações existentes entre as variáveis  $B$ ,  $C$  e  $D$  com a finalidade de sensibilizar a dependência entre elas.

Nas estimativas de LGD, a comparação entre os modelos de Gupton e Stein (2005), e de Hamerle, Knapp e Wildenauer (2009) foi realizada em bases com correlações distintas entre as variáveis simuladas  $B$ ,  $C$  e  $D$ , totalizando quatro conjuntos de dados – tipo 1, tipo 2, tipo 3 e tipo 4. A Tabela 1 resume as correlações definidas na criação das variáveis pertencentes à cópula.

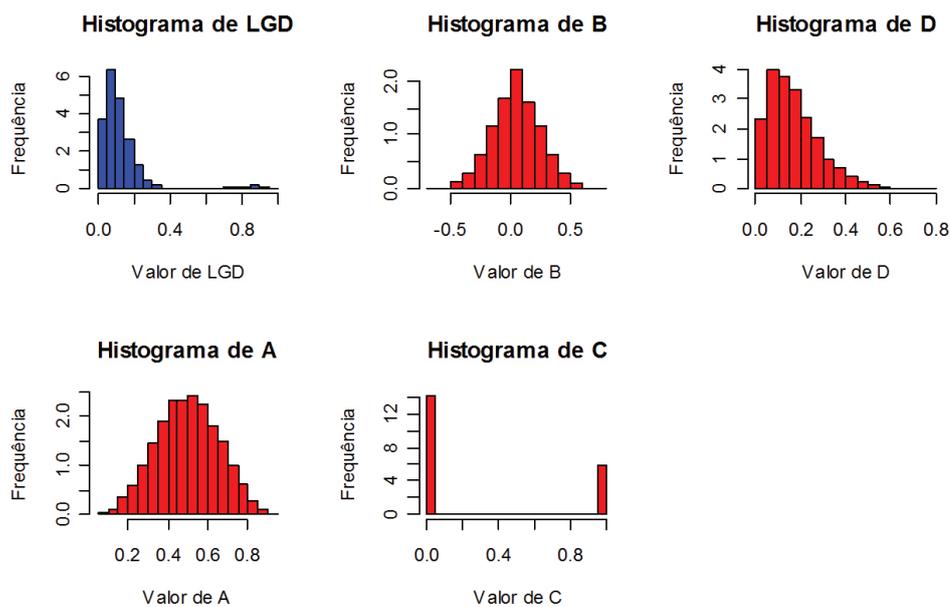
Tabela 1 - Correlação pré-definida para a criação das variáveis *B*, *C* e *D*

Correlação	$\rho_{BC}$	$\rho_{BD}$	$\rho_{CD}$
Tipo 1	0,75	0,50	0,75
Tipo 2	0,25	0,00	0,25
Tipo 3	-0,25	0,00	-0,25
Tipo 4	-0,75	-0,25	-0,75

Fonte: Os autores (2013)

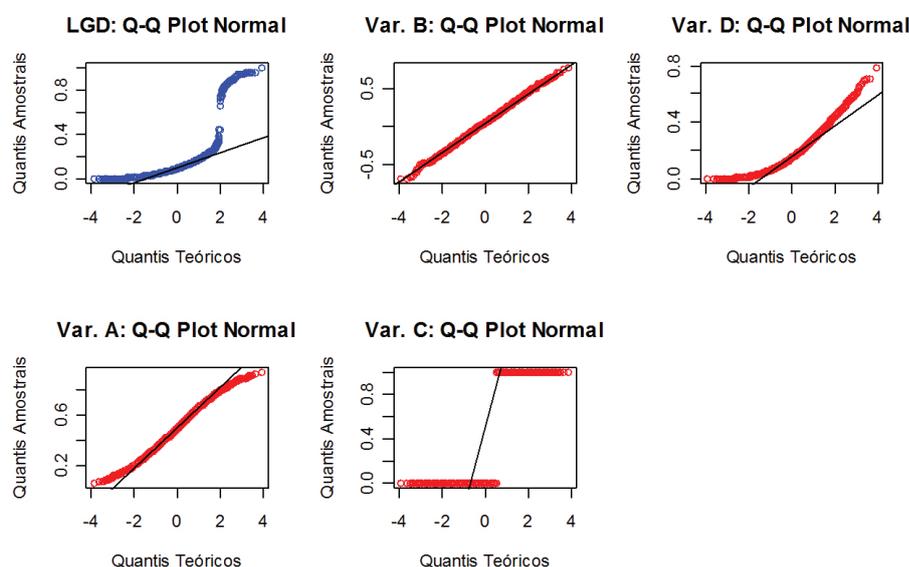
No total, foram geradas oito bases de dados, sendo quatro bases “ruins”, uma de cada tipo de correlação, e quatro bases “boas”, também uma de cada tipo de correlação. As distribuições dos parâmetros *A*, *B*, *C* e *D* na base de dados “boa” para a correlação do tipo 1 assim como a distribuição de LGD são apresentadas na Figura 6. As distribuições das quatro variáveis e de LGD nas outras bases de dados simuladas não serão apresentadas, pois que se assemelham aos histogramas que estão dispostos na sequência.

Figura 6 – Histograma de *A*, *B*, *C*, *D* e LGD



Fonte: Os autores (2013)

Figura 7 – Q-Q Plot de A, B, C, D e LGD



Fonte: Os autores (2013)

Os gráficos Q-Q Plot normais apresentados na Figura 7 mostram que somente a variável *B* se adapta à distribuição normal, conforme esperado, e as outras variáveis possuem distribuições diferentes da normal. A Tabela 2 resume as estatísticas descritivas das variáveis *A*, *B*, *C* e *D* na base de dados “boa” para a correlação do tipo 1 e de LGD. Para as outras bases de dados simuladas, os valores não serão apresentados, uma vez que se assemelham aos resultados para a correlação do tipo 1.

Tabela 2 – Estatísticas descritivas de A, B, C, D e LGD

Estatísticas	A	B	C	D	LGD
Média	0,498484	0,048505	0,294200	0,168630	0,128072
Erro padrão	0,001523	0,001997	0,004557	0,001108	0,001342
Mediana	0,498410	0,048716	0,000000	0,147913	0,099300
Modo	0,320364	0,277671	0,000000	0,369981	0,041900
Desvio padrão	0,152325	0,199693	0,455705	0,110847	0,134196
Variância da amostra	0,023203	0,039877	0,207667	0,012287	0,018009
Curtose	-0,475696	0,026618	-1,184112	0,948789	19,160714
Assimetria	-0,010788	-0,004987	0,903396	0,980680	4,037200
Intervalo	0,876709	1,458774	1,000000	0,780396	0,996200
Mínimo	0,056274	-0,687998	0,000000	0,000854	0,002800
Máximo	0,932983	0,770776	1,000000	0,781250	0,999000
Contagem	10000	10000	10000	10000	10000

Fonte: Os autores (2013)

Com relação às estatísticas descritivas apresentadas, verifica-se que a média de LGD é próxima de 0,1, indicando que a maior concentração de observações localiza-se na parte esquerda da distribuição, conforme pode ser verificado no histograma de LGD na Figura 6. Nota-se também que o desvio-padrão de LGD é maior que a média, e o erro padrão possui valor próximo de 0.

A Tabela 3 apresenta as matrizes de correlação das variáveis explicativas. Constata-se que, ao contrário do que deveria ocorrer, as dependências predefinidas na Tabela 1 são diferentes das realizadas após a

combinação entre as variáveis independentes e a variável resposta. Essa pode ser considerada uma das limitações do procedimento que, ao combinar os parâmetros explicativos e LGD, perde parte da dependência predefinida entre as variáveis *B*, *C* e *D* pertencentes à cópula.

Tabela 3 – Matrizes de correlação

Correlação		Base de dados "boa"				Base de dados "ruim"			
		A	B	C	D	A	B	C	D
Tipo 1	A	1.000	-0.702	-0.562	-0.680	1.000	-0.318	-0.261	-0.308
	B	-0.702	1.000	0.570	0.673	-0.318	1.000	0.331	0.404
	C	-0.562	0.570	1.000	0.604	-0.261	0.331	1.000	0.348
	D	-0.680	0.673	0.604	1.000	-0.308	0.404	0.348	1.000
Tipo 2	A	1.000	-0.693	-0.566	-0.685	1.000	-0.353	-0.274	-0.346
	B	-0.693	1.000	0.583	0.683	-0.353	1.000	0.335	0.421
	C	-0.566	0.583	1.000	0.618	-0.274	0.335	1.000	0.335
	D	-0.685	0.683	0.618	1.000	-0.346	0.421	0.335	1.000
Tipo 3	A	1.000	-0.718	-0.559	-0.701	1.000	-0.335	-0.261	-0.320
	B	-0.718	1.000	0.584	0.709	-0.335	1.000	0.331	0.409
	C	-0.559	0.584	1.000	0.618	-0.261	0.331	1.000	0.331
	D	-0.701	0.709	0.618	1.000	-0.320	0.409	0.331	1.000
Tipo 4	A	1.000	-0.709	-0.568	-0.697	1.000	-0.332	-0.247	-0.317
	B	-0.709	1.000	0.568	0.695	-0.332	1.000	0.331	0.400
	C	-0.568	0.568	1.000	0.620	-0.247	0.331	1.000	0.343
	D	-0.697	0.695	0.620	1.000	-0.317	0.400	0.343	1.000

Fonte: Os autores (2013)

Os resultados da regressão linear nos dois modelos utilizados são apresentados nas Tabelas 4 e 5. Verifica-se que tanto na base de dados "boa" quanto na base "ruim" todas as variáveis se mostraram significativas. Além disso, os sinais dos coeficientes da regressão linear indicaram que as correlações preestabelecidas entre as variáveis e LGD se realizaram, ou seja, o coeficiente de *A* foi positivo e os coeficientes de *B*, *C* e *D* foram negativos. Vale lembrar que os coeficientes da regressão linear não gerarão os valores de LGD, e sim valores que serão transformados em LGD.

A homocedasticidade dos modelos foi verificada mediante testes de Breusch-Pagan. Nesse procedimento, os resultados sugerem que, ao nível de 0,1% de significância, a homocedasticidade deve ser rejeitada, vale dizer, não se pode garantir que a variância dos termos de erros da função de regressão é constante. Adicionalmente, tanto a análise visual por meio de QQ-Plot quanto o teste de Kolmogorov-Smirnov para investigação dos resíduos da regressão indicam que os erros não seguem uma distribuição normal, podendo comprometer as estimativas dos coeficientes e intervalos de confiança.

É importante destacar ainda que o processo de estimação do modelo de LGD baseado em regressão linear múltipla sugerido por Gupton e Stein (2005), apesar de simples e direto, pode sofrer limitações. Em particular, pela própria construção das variáveis explicativas da LGD, existe multicolinearidade e, por exemplo, sinais e magnitudes dos coeficientes da regressão podem se tornar bastante sensíveis à cobertura da amostra (FARRAR; GLAUBER, 1967).

Tabela 4 – Resultados para a base “boa”

Correlação		Hamerle et Al. (2009)				KMV (Gupton e Stein, 2005)			
		Coef.	Erro Pad	Estat. T	Valor-p	Coef.	Erro Pad	Estat. T	Valor-p
Tipo 1	Const	-2,627	0,042	-63,204	0,0%	1,663	0,051	32,460	0,0%
	A	1,726	0,065	26,452	0,0%	1,247	0,080	15,500	0,0%
	B	-1,776	0,050	-35,823	0,0%	-1,635	0,061	-26,760	0,0%
	C	-0,185	0,019	-9,943	0,0%	-0,476	0,023	-20,710	0,0%
	D	-1,745	0,089	-19,598	0,0%	-2,699	0,110	-24,600	0,0%
	<b>R<sup>2</sup></b>	<b>0,597</b>				<b>0,554</b>			
Tipo 2	Const	-2,563	0,042	-60,610	0,0%	1,659	0,052	31,780	0,0%
	A	1,718	0,065	26,330	0,0%	1,373	0,081	17,050	0,0%
	B	-1,557	0,049	-31,480	0,0%	-1,353	0,061	-22,160	0,0%
	C	-0,283	0,019	-15,190	0,0%	-0,529	0,023	-22,980	0,0%
	D	-1,985	0,097	-20,550	0,0%	-3,019	0,119	-25,330	0,0%
	<b>R<sup>2</sup></b>	<b>0,599</b>				<b>0,556</b>			
Tipo 3	Const	-2,555	0,044	-57,660	0,0%	1,732	0,054	32,170	0,0%
	A	1,795	0,068	26,310	0,0%	1,319	0,083	15,910	0,0%
	B	-1,257	0,052	-24,000	0,0%	-1,154	0,064	-18,140	0,0%
	C	-0,213	0,019	-11,210	0,0%	-0,513	0,023	-22,180	0,0%
	D	-2,549	0,102	-24,900	0,0%	-3,465	0,124	-27,860	0,0%
	<b>R<sup>2</sup></b>	<b>0,582</b>				<b>0,552</b>			
Tipo 4	Const	-2,777	0,042	-66,790	0,0%	1,427	0,051	28,050	0,0%
	A	2,014	0,067	29,990	0,0%	1,627	0,082	19,800	0,0%
	B	-1,248	0,051	-24,690	0,0%	-1,108	0,062	-17,910	0,0%
	C	-0,165	0,019	-8,810	0,0%	-0,436	0,023	-18,980	0,0%
	D	-1,762	0,069	-25,440	0,0%	-2,436	0,085	-28,730	0,0%
	<b>R<sup>2</sup></b>	<b>0,594</b>				<b>0,558</b>			

Fonte: Os autores (2013)

Os valores de  $R^2$  foram sempre superiores no modelo de Hamerle, Knapp e Wildenauer (2009), que utiliza a transformação do tipo Logit, em comparação ao modelo de Gupton e Stein (2005), que emprega a transformação Beta. Assim, conclui-se que em todas as bases de dados analisadas o primeiro modelo se ajustou melhor às regressões do que o segundo, ou seja, suas variáveis independentes conseguiram explicar um percentual maior da variação de LGD.

Outro ponto observado é que houve somente pequenas variações no desempenho de ambos os modelos quando comparados os resultados obtidos nas bases de dados formadas com cada tipo de correlação. Dessa forma, pode-se dizer que a dependência entre as variáveis *B*, *C* e *D* não impactou os resultados dos modelos nas bases simuladas neste trabalho.

Tabela 5 – Resultados para a base “ruim”

Correlação		Hamerle et Al. (2009)				KMV (Gupton e Stein, 2005)			
		Coef.	Erro Pad	Estat. T	Valor-p	Coef.	Erro Pad	Estat. T	Valor-p
Tipo 1	Const	-1,985	0,034	-59,250	0,0%	1,951	0,040	48,590	0,0%
	A	0,907	0,054	16,770	0,0%	0,887	0,065	13,680	0,0%
	B	-1,341	0,043	-30,960	0,0%	-1,434	0,052	-27,610	0,0%
	C	-0,491	0,018	-26,790	0,0%	-0,635	0,022	-28,870	0,0%
	D	-2,725	0,078	-34,810	0,0%	-3,127	0,094	-33,320	0,0%
	<b>R<sup>2</sup></b>	<b>0,439</b>				<b>0,414</b>			
Tipo 2	Const	-2,244	0,036	-61,700	0,0%	1,710	0,043	39,420	0,0%
	A	1,336	0,058	23,170	0,0%	1,302	0,069	18,940	0,0%
	B	-1,227	0,045	-27,030	0,0%	-1,351	0,054	-24,950	0,0%
	C	-0,337	0,019	-18,070	0,0%	-0,464	0,022	-20,860	0,0%
	D	-2,751	0,087	-31,570	0,0%	-3,228	0,104	-31,060	0,0%
	<b>R<sup>2</sup></b>	<b>0,402</b>				<b>0,382</b>			
Tipo 3	Const	-2,270	0,034	-66,860	0,0%	1,701	0,041	41,080	0,0%
	A	1,403	0,054	26,040	0,0%	1,326	0,066	20,190	0,0%
	B	-1,479	0,043	-34,800	0,0%	-1,523	0,052	-29,390	0,0%
	C	-0,375	0,018	-21,060	0,0%	-0,533	0,022	-24,520	0,0%
	D	-2,724	0,083	-33,000	0,0%	-3,156	0,101	-31,340	0,0%
	<b>R<sup>2</sup></b>	<b>0,462</b>				<b>0,418</b>			
Tipo 4	Const	-2,440	0,033	-73,230	0,0%	1,521	0,040	38,050	0,0%
	A	1,402	0,055	25,340	0,0%	1,349	0,066	20,320	0,0%
	B	-1,670	0,044	-38,140	0,0%	-1,714	0,053	-32,600	0,0%
	C	-0,329	0,018	-17,890	0,0%	-0,479	0,022	-21,750	0,0%
	D	-1,539	0,059	-26,140	0,0%	-1,954	0,071	-27,660	0,0%
	<b>R<sup>2</sup></b>	<b>0,431</b>				<b>0,404</b>			

Fonte: Os autores (2013)

## 5 CONCLUSÃO

O acordo de Basileia II tem exigido das instituições financeiras uma melhor gestão de seu risco de crédito, dando flexibilidade para a utilização de modelos mais sofisticados. A modelagem do parâmetro LGD é essencial para os bancos que pretendem utilizar modelos internos no cálculo de seu capital regulatório. O presente trabalho apresentou uma análise comparativa entre o modelo LossCalc, da Moody's KMV e descrito por Gupton e Stein (2005) e o modelo descrito por Hamerle et al (2009), com a utilização de uma metodologia que pode ser adaptada pelas instituições financeiras. Além disso, utilizaram-se procedimentos diversos para a simulação de bases de dados pelo método de Monte Carlo. Ressalta-se que a relação de dependência entre as variáveis simuladas é obedecida por meio de mecanismos baseados em cópulas e em matrizes biestocásticas.

Tendo em vista a escassez de dados de LGD armazenados pelas instituições financeiras no Brasil, o uso de técnicas de simulação torna-se fundamental para que propriedades de modelos sejam testadas e avaliadas. Apesar de resultados empíricos que confrontem a perda em carteiras reais não terem sido estudados, os modelos desenvolvidos neste trabalho permitem que a instituição possa identificar alternativas de modelagem e de estimação da LGD. Neste contexto, a pesquisa com carteiras simuladas, baseada em Gupton e Stein (2005), e em Hamerle, Knapp e Wildenauer (2009), possibilita avaliar, a priori, a adequação de modelos de LGD, mesmo que as instituições ainda não tenham à disposição um banco de dados histórico extenso que viabilize estudos empíricos.

Destaca-se que o estudo realizado envolveu diversas etapas. Primeiramente, a simulação de portfólios exigiu a criação de observações de LGD e de quatro variáveis explicativas via Monte Carlo, seguindo a distribuição inicialmente definida para os parâmetros. A dependência entre algumas variáveis explicativas foi contemplada usando-se cópulas gaussianas. Em seguida, foi necessário combinar os valores simulados de LGD e das quatro variáveis, usando matrizes de probabilidade de junção entre quintis. A técnica para essa combinação baseou-se em matrizes biestocásticas, seguindo-se Hlawatsch e Ostrowski (2011). Finalmente, diversas bases simuladas foram escolhidas aleatoriamente para realizar estimativas de LGD com utilização dos modelos de Hamerle, Knapp e Wildenauer (2009), e de Gupton e Stein (2005). O primeiro deles obteve os melhores resultados em todos os casos.

Em particular, visto que as bases de dados foram simuladas e criadas por intermédio de matrizes de probabilidade para a junção entre as variáveis explicativas e LGD, os valores apresentados de  $R^2$  são muito superiores aos encontrados em trabalhos semelhantes com bases de dados reais. Bellotti e Crook (2012), por exemplo, reportaram em seu trabalho um valor de  $R^2$  igual a 0,11 em estudo realizado com base em 55.000 cartões de crédito no Reino Unido, entre os anos de 1999 e 2005, utilizando diversas variáveis explicativas e realizando regressões lineares com/pelo método dos mínimos quadrados ordinários.

Já Dermine e Carvalho (2005) apresentaram pseudo  $R^2$  igual a 0,13 nas estimativas de LGD feitas em bases de dados reais, sendo analisados 374 empréstimos corporativos de um banco europeu entre os anos de 1995 e 2000, novamente utilizando uma regressão linear pelo método dos mínimos quadrados ordinários nos valores transformados da variável dependente. Vale salientar que o  $R^2$  avalia a qualidade do ajuste da linha de regressão, ou seja, mede a proporção da variação total da variável dependente explicada pelo modelo de regressão. A acurácia e o desempenho dos modelos de previsão não foram avaliados em testes fora-da-amostra e fora-do-tempo, dadas as limitações decorrentes das bases de dados utilizadas.

Algumas limitações da metodologia de simulação podem ser elencadas, como por exemplo, o alto ajuste observado dos modelos relativamente à reta de regressão, em virtude do tipo de construção das bases. Como sugestão, as combinações entre as variáveis explicativas e LGD poderiam ser feitas por outros métodos em vez da matriz estocástica dupla utilizada. Outra limitação ocorre novamente na junção entre as variáveis e LGD, já que as correlações entre as variáveis  $B$ ,  $C$  e  $D$  não são totalmente mantidas na forma originalmente definida.

Para a criação de dependência, uma sugestão em trabalhos futuros seria a utilização de relações não lineares entre as variáveis explicativas, isto é, a criação de cópulas não gaussianas, visto que, em dados reais, é provável que as relações existentes entre as variáveis não sejam lineares. Outra sugestão seria a utilização da metodologia descrita em bases de dados reais de instituições financeiras com muitas ou poucas operações em inadimplência, para verificar o comportamento de variáveis não simuladas nas estimativas de perdas em portfólios de crédito. Adicionalmente, outras técnicas de estimação podem ser investigadas, em substituição à análise de regressão linear múltipla utilizada no presente estudo.

Finalmente, as instituições financeiras poderiam utilizar este trabalho como base metodológica para sua gestão de risco de crédito, visto que apresenta importantes ferramentas de simulações de dados e, ao mesmo tempo, propõe modelos que podem ser utilizados na mensuração das perdas em portfólios de

crédito. Além disso, os bancos poderiam estimar de forma mais precisa a LGD de suas carteiras, parâmetro essencial no cálculo dos requisitos mínimos de capital.

## LOSS GIVEN DEFAULT ESTIMATES IN SIMULATED CREDIT PORTFOLIOS

### ABSTRACT

Basel II allows banks to use internal models as a basis for the calculation of minimum capital requirements based on the level of exposure to credit risk. Among the main components of the models, there are estimates of probability of default (PD), loss given default (LGD) and exposure at default (EAD). This study aims to investigate mechanisms of LGD estimation based on regression models and logit and Beta transformations. Considering that the Brazilian market is still in an incipient stage in the analysis of LGD and that the availability of recovery rate data is restricted, we will simulate credit portfolios using Monte Carlo techniques. The dependence between the explanatory variables and LGD, the fraction of losses in case of default, is modeled by Gaussian copulas and doubly stochastic matrices. The results suggest that multivariate regression analysis, using different transformation functions, can properly identify the variables that explain the LGD. More particularly, the results indicate that the model developed by Hamerle, Wildenauer and Knapp (2009) for estimation of LGD has better performance than the performance of the model discussed by Gupton and Stein (2005).

**Keywords:** Credit risk. Loss given default. Basel accord. Doubly stochastic matrix. Gaussian copula.

### REFERÊNCIAS

ACHARYA, V.; BHARATH, S.; SRINIVASAN, A. Understanding the recovery rates on defaulted securities. **CEPR Discussion Papers**, v. 20, n. 10, Oct. 2003. Disponível em: <<http://www.cepr.org/pubs/dps/DP4098>>. Acesso em: 30 jul. 2013.

ALTMAN, E. Default recovery rates and LGD in credit risk modeling and practice: an updated review of the literature and empirical evidence. In: JONES, S.; HENSHER, D. A. (Ed.). **Advance in credit risk modeling and corporate bankruptcy prediction**, Cambridge: Cambridge University Press, 2008. p.175-206.

ALTMAN, E. I.; KISHORE, V. M. Almost everything you wanted to know about recoveries on defaulted bonds. **Financial Analysts Journal**, Charlottesville, VA, v. 52, n. 6, p. 57-64, Nov./Dec. 1996.

ALTMAN, E.; KALOTAY, E. **A flexible approach to modeling ultimate recoveries on defaulted loans and bonds**. [2010]. Disponível em: <[http://pages.stern.nyu.edu/~ealtman/FlexibleRecovery\\_v1.1.pdf](http://pages.stern.nyu.edu/~ealtman/FlexibleRecovery_v1.1.pdf)>. Acesso em: 30 jul. 2013.

ARATEN M.; JACOBS JR., M.; VARSHNEY, P. Measuring LGD on commercial loans: an 18-year internal study. **The RMA Journal**, Philadelphia, PA, v. 86, n.8, p. 96-103, May 2004.

BAIXAULI, J.; ALVAREZ, S. On the accuracy of loss-given-default prediction intervals. **Journal of Risk Finance**, Bingley, UK, v. 10, n. 2, p. 131-141, 2009.

BANCO CENTRAL DO BRASIL (BACEN). **Relatório de Economia Bancária e Crédito 2007**. [2008]. Disponível em: <[http://www.bcb.gov.br/Pec/Depep/Spread/relatorio\\_economia\\_bancaria\\_credito2007.pdf](http://www.bcb.gov.br/Pec/Depep/Spread/relatorio_economia_bancaria_credito2007.pdf)>. Acesso em: 30 jul. 2013.

BANCO CENTRAL DO BRASIL (BACEN). **Relatório de Economia Bancária e Crédito 2010**. [2011]. Disponível em: <<http://www.bcb.gov.br/Pec/Depep/Spread/REBC2010.pdf>>. Acesso em: 30 jul. 2013.

- BANCO CENTRAL DO BRASIL (BACEN). **Relatório de Economia Bancária e Crédito 2011**. [2012]. Disponível em: <[http://www.bcb.gov.br/pec/depep/spread/REBC\\_2011.pdf](http://www.bcb.gov.br/pec/depep/spread/REBC_2011.pdf)>. Acesso em: 31 jul. 2013.
- BASEL COMMITTEE ON BANKING SUPERVISION (BCBS). **Basel III: a global regulatory framework for more resilient banks and banking systems - revised version**. [2011]. Disponível em: <<http://www.bis.org/publ/bcbs189.htm>>. Acesso em: 30 jul. 2013.
- BELLOTTI, T.; CROOK, J. Loss given default models incorporating macroeconomic variable for credit cards. **International Journal of Forecasting**, Philadelphia, PA, v. 28, n. 1, p. 171-182, 2012.
- BENNETT, R.; CATARINEU, E.; MORAL, G. Loss given default validation. In: Basel Committee on Banking Supervision (Org.). **Studies on the Validation of Internal Rating Systems: working paper n. 14**. Basel, Switzerland: Bank for International Settlements, 2005. p. 60-93. Disponível em: <[http://www.bis.org/publ/bcbs\\_wp14.pdf](http://www.bis.org/publ/bcbs_wp14.pdf)>. Acesso em: 30 jul. 2013.
- CAREY, M. Credit risk in private debt portfolios. **Journal of Finance**, New Jersey, v. 53, n. 4, p. 1363-1387, 1998.
- CASELLI, S.; GATTI, S.; QUERCI, F. The sensitivity of the loss given default rate to systematic risk: new empirical evidence on bank loans. **Journal of Financial Services Research**, New York, v. 34, p. 1-34, 2008.
- DERMINE, J.; CARVALHO, C. Bank loan losses-given-default: a case study. **Journal of Banking & Finance**, Philadelphia, PA, v. 30, p. 1219-1243, 2005.
- DIERICK, F. et al. The new Basel capital framework and its implementation in the European Union. **Occasional Paper Series, European Central Bank**, Frankfurt, n. 42, 2005. Disponível em: <<http://www.ecb.int/pub/pdf/scops/ecbocp42.pdf>>. Acesso em: 30 jul. 2013.
- DULLMANN, K.; TRAPP, M. Systematic risk in recovery rates: an empirical analysis of U.S. corporate credit exposures. **EFWA Basel Paper**, Frankfurt, n. 2, 2004.
- FARRAR, D.; GLAUBER, R. Multicollinearity in regression analysis: the problem revisited. **The Review of Economics and Statistics**, Cambridge, v. 49, p. 92-107, 1967.
- FRYE, J. Depressing recoveries. **Risk**, London, v. 13, p. 108-111, Nov. 2000.
- GORDY, M.; JONES, D. **Capital allocation for securitizations with uncertainty in loss prioritization**. [2002]. Disponível em: <[https://www.bis.org/publ/bcbs\\_wp11gj.pdf](https://www.bis.org/publ/bcbs_wp11gj.pdf)>. Acesso em: 31 jul. 2013.
- GUPTON, M.; STEIN, R. **LossCalc V2: dynamic prediction of LGD**. [2005]. Disponível em: <[http://www.defaultrisk.com/pp\\_recov\\_63.htm](http://www.defaultrisk.com/pp_recov_63.htm)>. Acesso em: 31 jul. 2013.
- HAMERLE, A.; KNAPP, M.; WILDENAUER, N. Modelling loss given default: a “point in time-approach”. In: ENGELMANN, B.; RAUHMEIER, R. (Eds). **The Basel II risk parameters: estimation, validation, and stress testing**. Berlin: Springer, 2009. cap. 8.
- HAYDEN, E.; PORATH, D. Statistical Methods to Develop Rating Models. In: ENGELMANN, B.; RAUHMEIER, R. (Eds). **The Basel II risk parameters: estimation, validation, and stress testing**. Berlin: Springer, 2009. cap. 1.
- HLAWATSCH S.; OSTROWSKI, S. Simulation and estimation of loss given default. **Journal of Credit Risk**, London, v. 7, n. 3, p. 39-73, 2011.
- HU, Y.; PERRAUDIN, W. **The dependence of recovery rates and defaults**. [2002]. Disponível em: <[http://www.riskcontrollimited.com/research\\_papers/Hu\\_Perraudin\\_ref6\\_1.pdf](http://www.riskcontrollimited.com/research_papers/Hu_Perraudin_ref6_1.pdf)>. Acesso em: 31 jul. 2013.
- HUANG, X.; OOSTERLEE, C. Generalized Beta regression models for random loss given default. **Journal of**

**Credit Risk**, London, v. 7, n. 4, p. 45-70, 2011.

IVANOVA, V. **LGD-rating for a portfolio of retail loans**. [2004]. Disponível em: <<http://ibrarian.net/navon/page.jsp?paperid=11699093&searchTerm=loan+rate>>. Acesso em: 31 jul. 2013.

JACOBSON, T.; LINDE, J.; ROSZBACH, K. Internal rating systems, implied credit risk and the consistency of bank's risk classification policies. **Journal of Banking & Finance**, Philadelphia, PA, v. 30, p. 1899-1926, 2006.

JORION, P. **Value at Risk**. New York: McGraw-Hill, 2001.

MIU, P.; OZDEMIR, B. Basel requirement of downturn LGD: modeling and estimating PD & LGD correlations. **Journal of Credit Risk**, London, v. 2, n. 2, p. 43-68, 2006.

MIU, P.; OZDEMIR, B. **Basel II implementation: a guide to developing and validating a compliant, internal risk rating system**. New York: McGraw-Hill, 2009.

NELSEN, R. **An introduction to Copulas**. 2nd ed. New York: Springer, 1999.

ONORATO, M.; ALTMAN, E. An integrated pricing model for defaultable loans and bonds. **European Journal of Operational Research**, Philadelphia, PA, v. 163, p.65 – 82, 2005.

PESARAN, M. et al. Macroeconomic dynamics and credit risk: a global perspective. **Journal of Money, Credit & Banking**, Malden, MA, v. 38, p. 1211-1261, 2006.

PETER, C. Estimating loss given default: experiences from banking practice. In: ENGELMANN, B.; RAUHMEIER, R. (Eds). **The Basel II risk parameters: estimation, validation, and stress testing**. Berlin: Springer, 2009. p.151-184.

PYKHTIN, M. Unexpected recovery risk. **Risk**, London, v. 16, p. 74-78, 2003.

ROSCH, D.; SCHEULE, H. A multi-factor approach for systematic default and recovery risk. **Journal of Fixed Income**, New York, v. 15, n. 2, p. 63-75, 2005.

SCHUERMANN, T. **What do we know about loss given default?** [2004]. Disponível em: <<http://fic.wharton.upenn.edu/fic/papers/04/0401.pdf>>. Acesso em: 31 jul. 2013.

SILVA, A.; MARINS, J.; NEVES, M. **Loss given default: um estudo sobre perdas em operações prefixadas no mercado brasileiro**. Brasília: Banco Central do Brasil, 2009. Disponível em: <<http://www.bcb.gov.br/pec/wps/port/wps193.pdf>>. Acesso em: 31 jul. 2013.

SKLAR, A. Random variables, joint distribution functions, and copulas. **Kybernetika**, Praga, República Tcheca, v. 9, p. 449-460, 1973.

STEPHANOU, C.; MENDOZA, J. C. **Credit risk measurement under Basel II: an overview and implementation issues for developing countries**. [2005]. Disponível em: <<http://elibrary.worldbank.org/content/workingpaper/10.1596/1813-9450-3556>>. Acesso em: 31 jul. 2013.

WAGSTER, J. D. Impact of the 1988 Basle Accord on international banks. **Journal of Finance**, New Jersey, v. 51, n. 4, p. 1321-1346, 1996.

YAN, J. Enjoy the joy of copulas: With a Package copula. **Journal of Statistical Software**, Los Angeles, v. 21, n. 1, p. 1-21, 2007.