



Revista ADM.MADE

Revista do Mestrado em Administração e
Desenvolvimento Empresarial - Universidade
Estácio de Sá

Revista ADM.MADE, Rio de Janeiro, ano 15, v.19, n.2, p.1-20, maio/agosto, 2015

Revista do Mestrado em Administração e Desenvolvimento Empresarial da Universidade

Estácio de Sá – Rio de Janeiro (MADE/UNESA). ISSN: 2237-5139

Conteúdo publicado de acesso livre e irrestrito, sob licença Creative Commons 3.0.

Editor científico: Marco Aurélio Carino Bouzada

Modelos Estatísticos para Previsão do LGD de Empréstimos de Varejo

*Natália Cordeiro Zaniboni*¹

*Alessandra de Ávila Montini*²

*Alcides Carlos de Araújo*³

Artigo recebido em 20/07/2014 e aprovado em 25/05/2015. Artigo avaliado em *double blind review*.

¹ Mestre em Administração de Empresas pela Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo (FEA/USP). Endereço: Av. Professor Luciano Gualberto, 908 - sala G 157 - Butantã, São Paulo - SP, 05508-010. Email: nzaniboni@usp.br.

² Doutora em Administração de Empresas pela Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo (FEA/USP). Professora do Programa de Pós-Graduação em Administração da FEA-USP. Endereço: Av. Professor Luciano Gualberto, 908 - sala G 157 - Butantã, São Paulo - SP, 05508-010. Email: amontini@usp.br.

³ Mestre em Administração de Empresas pela Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo (FEA/USP). Endereço: Av. Professor Luciano Gualberto, 908 - sala G 157 - Butantã, São Paulo - SP, 05508-010. Email: alcides.carlos@usp.br.

Modelos Estatísticos para Previsão do LGD de Empréstimos de Varejo

A concessão de crédito gera riscos financeiros que podem ocasionar a inadimplência e a quebra de grandes instituições financeiras. Diversos modelos de previsão de inadimplência foram criados para o gerenciamento do risco de crédito. Estes modelos estatísticos vêm sendo constantemente aprimorados com as exigências do Acordo de Basiléia 2, que apresenta o capital requerido para risco de crédito, composto por três componentes de risco: probabilidade de descumprimento (PD), exposição no momento do descumprimento (EAD) e perda financeira dado o descumprimento (LGD). A modelagem estatística do componente LGD se mostrou complexa e sua distribuição tem difícil ajuste. Esse estudo objetivou propor uma técnica de modelagem e previsão do LGD em duas fases: na primeira fase, é prevista a probabilidade de o LGD ser alto (maior que 50%). Na segunda fase, o LGD é predito por meio do LGD médio de cada grupo (alto LGD e baixo LGD). O diferencial dessa técnica é a proposta de utilização da média na segunda fase. Os resultados indicam que, na primeira fase, uma árvore de decisão é mais adequada que uma regressão logística, pois classificou corretamente 91%, ante 87% do modelo de regressão logística. O segmento e a EAD afetam o LGD. O modelo final proposto foi comparado com um modelo de regressão linear e apresentou melhor ajuste, com menor soma dos quadrados dos erros.

Palavras-chave: Basiléia; *Loss Given Default*; Regressão Logística; Árvore de Decisão; Risco de Crédito.

Keywords: Basel; Loss Given Default; Logistic Regression; Decision Tree; Credit Risk.

Statistical Models on Forecasting LGD for Retail Loans

Credit creates financial risks that may result in large financial institutions default and bankruptcy. Several default prediction models were created for credit risk management. These statistical models have been constantly refined with Basel 2 requirements, which presents the required capital for credit risk, that consists of three risk components: PD (probability of default), EAD (exposure at default) and LGD (loss given default). Statistical modeling of component LGD has been complex and its distribution is difficult to adjust. This study aimed to propose a simple method for modeling and predicting LGD in two phases: at the first phase, logistic regression and decision tree models were constructed and at the second phase LGD was predicted using descriptive statistics. Results indicate that, at first stage, decision tree is more appropriate as it correctly classified 91%, compared to 87% for logistic regression model. Segment and exposure at default (EAD) affect the LGD. The proposed final model was compared with a linear regression model and showed better fit with lower sum of squared errors.

1. INTRODUÇÃO

A disponibilidade do crédito estimula a economia, pois possibilita o consumo das pessoas físicas e o desenvolvimento das empresas. Teorias econômicas indicam que o avanço do crédito pode explicar o desenvolvimento de alguns países (CAOUILLE et al., 1998). Por outro lado, a concessão de crédito gera riscos financeiros que podem ocasionar a inadimplência. A quebra de diversas instituições financeiras, muitas vezes, está associada com essa inadimplência. Com a globalização, a inadimplência interna de determinado país pode proporcionar reflexo no sistema financeiro de diversos países (MILEO et al., 2013).

A correta decisão no gerenciamento de risco de crédito é essencial para as instituições financeiras, pois a inadimplência de poucas operações pode consumir o ganho de muitas operações lucrativas (STEINER et al., 1999). Nos EUA, o setor de crédito ao consumidor cresceu aproximadamente 107% entre 2000 e 2013 (US\$1,5 bilhão para US\$3,1 bilhões), enquanto a taxa de inadimplência passou de 0,96% a 2,88% no mesmo período (FEDERAL RESERVE BANK, 2014). A crise financeira de 2008 foi gerada por financiamentos imobiliários a clientes de alto risco (em inglês, *subprime*) nos EUA e acarretou crise financeira no mundo inteiro, incluindo falências de grandes instituições financeiras como Lehman Brothers, AIG e Merrill Lynch. Essa crise aumentou a atenção de agências reguladoras, de governos e de instituições financeiras com relação ao gerenciamento e à mensuração de risco de crédito.

O gerenciamento de crédito é um grande desafio das instituições financeiras em todo o mundo. Diversos modelos de mensuração de risco de crédito foram criados para o gerenciamento das operações de crédito na concessão, na precificação e na recuperação, e possibilitam, ao banco, conceder créditos de forma a minimizar o risco de inadimplência e a precificar de forma adequada. A grande competitividade entre as instituições financeiras torna esse gerenciamento um diferencial (STEINER et al., 1999).

Os modelos estatísticos de previsão de inadimplência vêm sendo constantemente aprimorados com as exigências do Acordo de Basiléia 2, desenvolvido pelo Comitê de Basiléia em 2004 com o objetivo principal de padronizar as formas de cálculo do capital mínimo requerido (SANTOS, 2001). Esse capital requerido é a reserva financeira necessária para uma instituição fazer frente às suas exposições de risco, e é calculado através das previsões de inadimplência das operações de crédito. O acordo possibilita aos bancos desenvolverem metodologias e calcularem suas próprias estimativas de risco para cálculo deste capital requerido, que, geralmente, são feitas por meio de modelos estatísticos de previsão de inadimplência.

O capital requerido para risco de crédito é composto por três componentes de risco: a probabilidade de descumprimento de um contrato de crédito (PD), o valor devido no momento do descumprimento (EAD), e a perda financeira caso haja o descumprimento (LGD). O descumprimento é definido, geralmente, como 90 dias de atraso das parcelas de um contrato de crédito. O componente LGD mensura a perda financeira de um contrato de crédito após o descumprimento, e considera a perda do principal do contrato, a perda decorrente dos custos de oportunidade e os custos relacionados às despesas do processo de cobrança (SILVA et al., 2008).

As primeiras pesquisas publicadas relacionadas ao LGD, que se iniciaram na década de 1990 com Asarnow e Edwards (1995), concentraram-se na análise da distribuição do LGD ou nos fatores que influenciam a variabilidade do LGD. Pesquisas posteriores, a partir de 2008, concentraram-se nas técnicas estatísticas de modelagem do LGD. A maior parte destas pesquisas abordam contratos de crédito a grandes empresas ou contratos de crédito imobiliário (TONG et al., 2013).

Os primeiros trabalhos propuseram a modelagem direta do LGD (BASTOS, 2010; BELLOTTI; CROOK, 2008; GUPTON; STEIN, 2002; HARTMANN-WENDELS et al., 2014; LEOW;

MUES, C, 2012; ZHANG; THOMAS, L, 2012), porém a precisão das estimativas foi limitada devido à natureza bimodal do LGD e à falta de características preditivas (TONG et al., 2013).

Pesquisas posteriores (LEOW; MUES, 2012; THOMAS et al., 2010; TONG et al., 2013) propuseram a modelagem do LGD em duas fases. Inicialmente, o LGD é reclassificado como binário (ou seja, um alto ou um baixo LGD) e modelado por meio de distribuições binárias. Em seguida, o valor do LGD é previsto a partir de um modelo linear para cada classificação binária. Nessa fase, os modelos propostos também tiveram pequena precisão ou a distribuição ajustada não se apresentou adequada à modelagem.

Esse estudo objetivou propor uma técnica de modelagem e de previsão do LGD em duas fases. Na primeira fase, um modelo prevê a probabilidade de o LGD ser alto – ou seja, maior que 50%. Na segunda fase, o LGD é predito por meio do LGD médio de cada grupo – ou seja, alto LGD e baixo LGD.

O diferencial dessa técnica é a proposta de utilização da média na segunda fase. Um modelo logístico binário e uma árvore de decisão são apresentados para a modelagem da probabilidade de alto LGD na primeira fase do modelo, e os resultados obtidos são comparados. Os valores preditos pela técnica proposta de dois estágios são comparados aos resultados de um modelo de regressão linear simples.

O trabalho contribui para a literatura brasileira de risco de crédito nos seguintes aspectos: (1) propõe uma técnica ainda não apresentada e de simples aplicação para modelagem do LGD, em que o diferencial é a proposta de utilização da média na segunda fase, evitando possíveis problemas de ajuste ou acurácia na segunda fase de modelagem (LOTERMAN et. al., 2012); (2) utiliza todos os tipos de contratos de crédito do varejo, e não somente crédito imobiliário ou de grandes empresas; e (3) utiliza base de dados composta por informações de diversas instituições financeiras no mundo, diferentemente de pesquisas anteriores.

A seção 2 aborda a revisão bibliográfica, a seção 3 apresenta o método utilizado, a seção 4 expõe os resultados encontrados, e a seção 5 traz as considerações finais do trabalho.

2. REFERENCIAL TEÓRICO

O referencial teórico deste trabalho é dividido em três subseções: 1) definição e forma de cálculo do LGD; 2) estudos que se concentram em identificar fatores que afetam o LGD; e 3) estudos que propõem ou que comparam modelos estatísticos para previsão do LGD.

2.1. Cálculo do LGD

Os métodos de cálculo do LGD são três (CHALUPKA; KOPECSNI, 2009):

- 1) Market LGD: calculado através dos preços dos títulos e de empréstimos negociáveis após o descumprimento.
- 2) Workout LGD: calculado através do fluxo de caixa resultante do período de recuperação a partir da data do descumprimento, descontado a valor

presente na data do descumprimento. É o método mais comumente utilizado (SILVA et al., 2008).

- 3) Implied Market LGD: calculado através de preços de ativos de risco, mas ainda não em descumprimento, usando um modelo teórico de precificação de ativos.

A expressão (1) apresenta o cálculo do LGD utilizando a metodologia Workout LGD.

$$\text{LGD} = 1 - \frac{\text{Pagamentos} - \text{Custos}}{\text{EAD}}, \quad (1)$$

em que os pagamentos são compostos pelos valores recuperados após o descumprimento trazidos a valor presente na data do descumprimento, os custos são os custos de recuperação trazidos a valor presente na data do descumprimento, e o EAD é o valor devido no momento do descumprimento.

Para exemplificar o cálculo do LGD, suponha-se um contrato que teve o pagamento atrasado e cujo valor devido, no momento do descumprimento, é R\$100,00 (EAD). Após período de esforço de recuperação de crédito, o banco recuperou do cliente pagamento de R\$ 20 e teve custo de R\$10. Assim, o valor do LGD resultante é de 90%, apresentado na expressão (2), e indica que a instituição financeira perdeu 90% do valor devido do contrato.

$$\text{LGD} = 1 - \frac{20 - 10}{100} = 1 - 0,10 = 0,9. \quad (2)$$

2.2. Fatores que influenciam o LGD

Os estudos sobre modelos de LGD se iniciaram na década de 1990, com Asarnow e Edwards (1995), que analisaram o comportamento do LGD durante 24 anos do banco Citibank nos EUA. Eles mostraram que existe diferença dos LGDs entre operações estruturadas e operações comerciais e industriais. As operações com garantias apresentam LGD médio de 12% e as operações sem garantia apresentam LGD médio de 35%.

Hurt e Felsovalyi (1998) também analisaram o comportamento da taxa de perda do banco Citibank, porém com foco na América Latina, mostrando que o valor devido no contrato de crédito afeta o LGD. A distribuição do LGD apresentou dois picos, sendo um com baixo valor de LGD (13,2%) e outro com alto valor de LGD (53,8%).

Gupton et al. (2000), apresentaram um estudo com 181 empréstimos bancários dos EUA, apontando que a média dos LGDs dos empréstimos com garantia é 52% e que a média do LGD dos empréstimos sem garantia é 69%.

Gupton e Stein (2002) propuseram modelo denominado LossCalc para previsão do LGD, desenvolvido com 1.800 títulos, empréstimos e ações preferenciais em

descumprimento, abrangendo mais de 900 empresas públicas e privadas dos EUA. Utilizaram, como variáveis explicativas, tipo de dívida e senioridade, estrutura de capital da empresa, indicadores da indústria e fatores macroeconômicos.

Schuermann (2004) também obteve uma distribuição para o LGD com dois picos. A presença e a qualidade da garantia, a indústria da empresa, o grau de subordinação do título e a situação econômica são fatores que afetam o LGD - sendo que, no caso da indústria da empresa, foi indicado que empresas de ativos tangíveis têm maior recuperação.

Querci (2005) estudou o comportamento de 15.827 empréstimos em um banco comercial Italiano. Para empréstimos de varejo, a média do LGD foi 53,3%, enquanto que, para pequenas e médias empresas (PME), a média foi 48,4%. Também obteve a distribuição de dois picos para o LGD.

Dermine e Carvalho (2006) desenvolveram estudo com 371 operações de crédito de empresas de um banco em Portugal. O valor devido, a garantia e o setor da empresa influenciam o LGD, que apresentou uma distribuição de dois picos.

Para produtos de cartão de crédito na Grã-Bretanha, Belotti e Crook (2008) identificaram que relacionamento, renda, número de cartões de crédito, tempo no endereço, tipo de emprego, escores de crédito, valor devido e região de residência podem influenciar o LGD.

Chalupka e Kopecsni (2008) estudaram o LGD de empresas de uma instituição financeira da República Checa. Observou-se também a distribuição com dois picos. Saldo devedor, tempo de recuperação, tipo de indústria e tempo de empresa foram variáveis significativas no modelo.

Silva et al. (2008) estudaram 9.557 operações registradas no Sistema de Informações de Crédito do Banco Central do Brasil das modalidades cheque especial, conta garantida e capital de giro. Também observaram a distribuição de dois picos do LGD. O valor devido, o tempo de relacionamento do cliente, a presença de garantia, o porte do cliente, o *rating*, a presença de renegociações e o setor industrial da empresa afetam o valor do LGD. Foi recomendada a análise do LGD visando o segmento dos contratos.

Qi e Yang (2009) estudaram empréstimos imobiliários de diversas seguradoras residenciais nos EUA e mostraram que o LGD destes empréstimos pode ser explicado pelo percentual emprestado em relação ao valor total do imóvel, pela depreciação do imóvel, pelo valor emprestado, pelo tipo de cliente, pelo propósito do financiamento (se é uma compra ou um refinanciamento), pelo imóvel servir, ou não, para moradia, e pelo tempo decorrido desde o início do empréstimo.

Park e Bang (2014), utilizando dados de empréstimos imobiliários na Coreia, estudaram fatores que afetam o LGD desses contratos. Identificaram que as hipotecas geram valores de LGD baixos (5-10%), enquanto o LGD de créditos subordinados varia entre 30% e 50%. O valor da garantia, e as características do cliente e do processo de leilão de venda do imóvel afetam o LGD de empréstimos imobiliários.

A Tabela 1 apresenta um resumo dos estudos relacionados aos fatores que afetam o LGD.

Tabela 1 - Resumo dos estudos de LGD

Estudo	Fatores que afetam o LGD	Valores Médios de LGD		Distribuição do LGD	Local
Asarnow e Edwards (1995)	Tipo de operação	Com Garantia	12%	-	EUA
		Sem Garantia	35%		
Hurt e Felsovalyi (1998)	Exposição	32%		Dois Picos	América Latina
Gupton et al. (2000)	Garantia	Com Garantia	52%	-	EUA
		Sem Garantia	69%		
Gupton e Stein (2002)	- Tipo de dívida e senioridade	32%		-	EUA
	- Estrutura de capital da empresa				
	- Indicadores da indústria				
	- Fatores macroeconômicos				
Schuermann (2004)	- Presença e a qualidade da garantia	40%		Dois Picos	EUA
	- Indústria da empresa				
	- Grau de subordinação do título				
	- Situação do contrato				
Querci (2005)	Código do cliente	Varejo	53,3%	Dois Picos	Itália
		PME	48,4%		
Dermine e Carvalho (2006)	- Valor do empréstimo	29%		Dois Picos	Portugal
	- Garantia				
	- Setor da empresa				
	- Tempo de relacionamento				
	- Renda				
Belotti e Crook (2008)	- Número de cartões de crédito	-		-	Grã-Bretanha
	- Tempo no endereço				
	- Tipo de emprego				
	- Escores de crédito ou comportamental				
	- Saldo devedor				
	- Região de residência				

Chalupka e Kopecsni (2008)	- Exposição - Tempo de recuperação - Tipo de indústria - Tempo de empresa	52%	Dois Picos	República Checa
Silva et al. (2008)	- Saldo devedor - Tempo de relacionamento do cliente - Presença de garantia - Porte e <i>rating</i> do cliente - Presença de renegociações - Setor industrial da empresa - Segmento do contrato	47%	Dois Picos	Brasil
Qi e Yang (2009)	- <i>Loan-to-value</i> - Depreciação do imóvel - Valor emprestado - Tipo de cliente - Propósito do financiamento - Moradia ou não - Tempo decorrido desde o início do empréstimo	-	-	EUA
Park e Bang (2014)	- Tipo da garantia - Valor da garantia - Cliente - Processo de leilão - Mercado imobiliário	Hipoteca Subordinado	5-10% 30-50%	Unimodal Coréia

Fonte: Elaborado pelos autores

2.3. Modelos estatísticos de previsão do LGD

Belotti e Crook (2008) compararam modelos de mínimos quadrados ordinários, regressão tobit e árvore de decisão para previsão do LGD com dados de cartão de crédito na Grã-Bretanha, sendo que os modelos de mínimos quadrados ordinários apresentaram resultados mais precisos.

Bastos (2010) comparou os modelos de regressão fracional e árvore de regressão não paramétrica utilizando dados de empréstimos a PMEs de um banco de Portugal. Obteve que a árvore de regressão apresentou estimativas mais próximas dos valores reais para

horizontes de recuperação curtos, enquanto a regressão fracional apresentou melhores resultados para horizontes de recuperação longos.

Thomas et al. (2010) utilizaram uma árvore de decisão na previsão do LGD em empréstimos de uma instituição financeira no Reino Unido, abrangendo período de 1989 a 2004. Também obtiveram uma distribuição com dois picos. A árvore de decisão proposta é um processo de modelagem em dois estágios:

(1) o LGD foi reclassificado como binário, em que $Y_i = 0$ se $LGD \leq 0$ e $Y_i = 1$ se $LGD > 0$, e um modelo de regressão logística foi utilizado. A exposição do contrato, o histórico de atrasos, o tempo em que o cliente vive no mesmo endereço e se há outro requerente do crédito foram significativos no modelo.

(2) se o contrato foi classificado como $LGD > 0$ na primeira etapa, o valor do LGD foi previsto a partir de modelo de regressão linear múltipla, de regressão beta, de transformação log-normal e de regressão linear weight of evidence (WOE). O histórico de atrasos, o score do cliente na concessão do crédito, a exposição do contrato e o tempo até o descumprimento foram as variáveis significativas. O maior R^2 do segundo modelo obtido foi 0,22 para o método WOE.

Zhang e Thomas (2012) compararam o modelo de regressão linear múltipla com a análise de sobrevivência na previsão do LGD utilizando empréstimos sem garantia e empréstimos por meio de cartão de crédito de uma instituição financeira no Reino Unido, abrangendo período de 1987 a 2003. A regressão linear obteve melhores resultados que a análise de sobrevivência tanto para os casos em que o LGD foi modelado diretamente, quanto nos casos de modelagem em dois estágios.

Loterman et al. (2012) compararam técnicas de modelagem do LGD já estudadas na literatura. Utilizaram tanto modelos de um estágio, como regressão linear múltipla, quanto modelos de dois estágios, que combinam mais de uma técnica, totalizando 24 técnicas. Indicaram que grande parte da variabilidade do LGD não é explicada, e que modelos não lineares, como redes neurais, alcançam melhores resultados que modelos lineares tradicionais.

Leow e Mues (2012) propuseram um modelo de dois estágios, que cria um modelo de regressão logística para a probabilidade de alta perda e um modelo de regressão linear para a severidade da perda (ou para a perda com a venda dos imóveis). A distribuição da perda não apresentou normalidade, premissa importante para a modelagem de regressão, porém foi aproximada com objetivo de previsão do LGD. O modelo proposto foi mais adequado para a previsão do LGD que um modelo de regressão linear múltipla.

Tong et al. (2013) propuseram modelagem em dois estágios em que a probabilidade de um LGD igual a zero e o valor do LGD são previstos. O método é aplicado em dados de contratos residenciais de um banco do Reino Unido. O modelo proposto (gama inflada de zeros) obteve bom ajuste.

Hartmann-Wendels et al. (2014) utilizaram dados de empresas alemãs de *leasing* para analisar diferentes abordagens de estimação do LGD. Usando a média histórica e regressão linear simples como *benchmarks*, compararam modelos de mistura finitos híbridos

(FMMs), árvore de decisão e árvores de regressão. Indicaram que árvores de decisão são mais robustas se o tamanho da amostra é suficientemente grande.

A Tabela 2 apresenta um resumo dos estudos que focalizam nas técnicas de modelagem do LGD.

Tabela 2 - Técnicas de modelagem do LGD

Autores	Técnicas utilizadas	Técnica mais adequada	Local
Belotti e Crook (2008)	- Modelos de mínimos quadrados ordinários - Regressão tobit - Árvore de decisão	Modelos de mínimos quadrados ordinários	Grã-Bretanha
Bastos (2010)	- Regressão fracional - Árvore de regressão	Árvore de regressão	Portugal
Thomas et al. (2010)	- Modelo em dois estágios: Estágio 1. Modelo de regressão logística Estágio 2. Regressão linear	Modelo em dois estágios	Reino Unido
Zhang e Thomas (2012)	- Regressão linear - Análise de sobrevivência	Regressão linear	Reino Unido
Loterman et. al. (2012)	- Regressão linear - Regressão beta - Regressão robusta - Regressão de cumeieira - Regressão por <i>splines</i> - Redes neurais - Árvore de regressão	Redes neurais	Diversos países
Leow e Mues (2012)	- Modelos de dois estágios Estágio 1. Regressão logística Estágio 2. Regressão linear - Regressão linear	Modelos de dois estágios	Reino Unido
Tong et al. (2013)	- Regressão beta - Gama inflada de zeros - Regressão tobit	Gama inflada de zeros	Reino Unido
Hartmann-Wendels et al. (2014)	- FMMs - Árvore de decisão - Árvore de regressão	Árvore de decisão	Alemanha

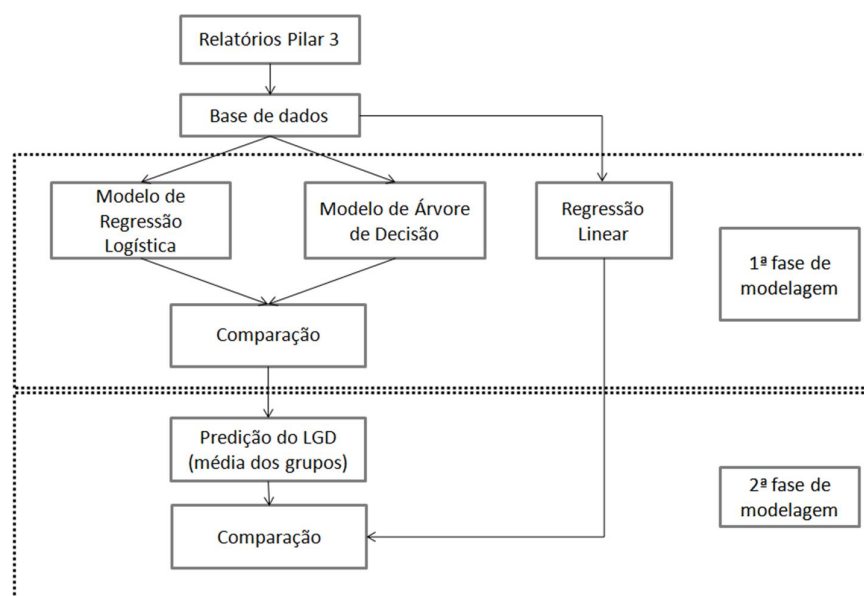
Fonte: Elaborado pelos autores

3. MÉTODO

O método utilizado é dado por: (1) construção da base de dados a partir dos Relatórios de Pilar 3 de diversas instituições financeiras, (2) construção e comparação de dois modelos para a primeira fase da estimação, ou seja, a previsão da probabilidade de alto LGD (regressão logística e árvore de decisão), (3) estimação do valor do LGD por meio da sua média grupo de alta e baixa probabilidade de alto LGD e (4) comparação do modelo final proposto (combinação das duas fases) com o método de regressão linear.

Na Figura 1 é apresentado o resumo do processo de pesquisa.

Figura 1 - Esquema do método



Fonte: Elaborado pelos autores

3.1. Base de Dados

Pelo Acordo de Basiléia 2, as instituições devem se submeter a um pilar denominado Disciplina de Mercado, em que participantes do mercado podem acessar informações-chave sobre o capital, sobre exposições a risco, sobre processos de gerenciamento de riscos e sobre a adequação do capital das instituições. Estas instituições geralmente publicam um relatório de gerenciamento de riscos anual para todo o mercado.

No presente trabalho analisaram-se 12 relatórios de gerenciamento de riscos de instituições financeiras (Australia and New Zealand Banking Group, Barclays, Commonwealth, Credit Suisse, Danske Bank, Deutsche Bank, National Australian Bank, NIBC, RBS, Santander, SNS Bank's e Westpac) no período de dezembro de 2008 a março de 2011.

Os dados utilizados se referem às exposições de varejo de grupos homogêneos de risco. A base de dados foi composta por 214 grupos homogêneos de risco de contratos oriundos desses oito países (Alemanha, Austrália, Dinamarca, Escócia, Espanha, Holanda,

Inglaterra e Suíça). As variáveis selecionadas foram PD, LGD, EAD (padronizada em dólares) e segmento. O segmento é uma classificação definida pelo Acordo de Basileia 2, que divide os contratos como empréstimos a PMEs, financiamento imobiliário, créditos rotativos e outros empréstimos.

Utilizou-se o método de Thomas et al. (2010) na definição da variável resposta do modelo: $Y_i = 1$ para grupos com LGD maior ou igual a 50% (grupo de alta perda dado o descumprimento) e $Y_i = 0$ para grupos com LGD menor que 50% (grupo de baixa perda dado o descumprimento).

É essencial separar contratos que obtiveram alta perda (alto LGD) e baixa perda (baixo LGD) para a assertividade da previsão do LGD, pois os fatores que afetam a recuperação de crédito são diferentes para os grupos (GÜRTLER; HIBBELN, 2013; THOMAS, L et al., 2012).

3.2. Regressão Logística

O primeiro modelo utilizado para explicar a probabilidade de $Y_i = 1$ é o modelo de regressão logística binária. A regressão logística binária consiste em relacionar uma variável resposta binária com variáveis explicativas, que podem ser categóricas, contínuas ou discretas.

Segundo Hosmer e Lemeshow (1989) a função ideal para modelar casos binários é a função logito. O modelo estima a probabilidade de $Y_i = 1$, neste caso estima-se a probabilidade de alto LGD, que é apresentada na expressão (3)

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k}} \quad (3)$$

em que $\beta_0, \beta_1, \dots, \beta_k$ são os parâmetros do modelo relacionados às k variáveis explicativas, $\pi(x)$ é a probabilidade estimada de alto LGD, e X_k é a k -ésima variável explicativa.

Como os erros do modelo não possuem distribuição normal e a variância não é constante, o método de estimação dos parâmetros deve ser o método da máxima verossimilhança. Os estimadores dos parâmetros obtidos pelo método de maximização da função de verossimilhança (máxima verossimilhança) são eficientes e produzem estimativas com distribuição normal assintótica.

3.3. Árvore de Decisão

A árvore de decisão é método que tem, como objetivo, classificar as observações em grupos com base no perfil. Existem diversos métodos desenvolvidos para a elaboração da árvore de decisão (WITTEN; FRANK, 1999). Breiman et al. (1984) elaboraram o método Classification and Regression Trees (CART), que testa as subdivisões de nós possíveis e elege a melhor divisão baseada na capacidade de discriminação de acordo com indicadores (Gini, Entropia ou Paridade).

Pode-se obter o perfil associado ao grupo de alta perda dado o descumprimento (LGD) também por meio de uma árvore de decisão. Uma árvore de decisão é construída de forma recursiva. Inicia-se com o conjunto de dados, e este é dividido com base em uma das

variáveis independentes, formando um subconjunto homogêneo em relação ao LGD. Esse processo é repetido até que se formem vários subconjuntos homogêneos com LGD semelhantes entre si.

O resultado da árvore é formado por uma sequência de grupos divididos de acordo com a probabilidade de alta perda dado o descumprimento. A variável com maior importância em relação à variável resposta compõe o primeiro nó. Os demais nós são formados sucessivamente de acordo com a importância da variável. Cada combinação de decisões da árvore resulta em uma probabilidade do evento da variável resposta.

3.4. Comparação de Modelos

Para comparar os resultados dos modelos foi obtida a matriz de assertividade do modelo de regressão logística e do modelo de árvore de decisão. A matriz de assertividade apresenta a quantidade de observações que o modelo fez a previsão de forma correta e a quantidade de observações que o modelo fez a previsão incorreta.

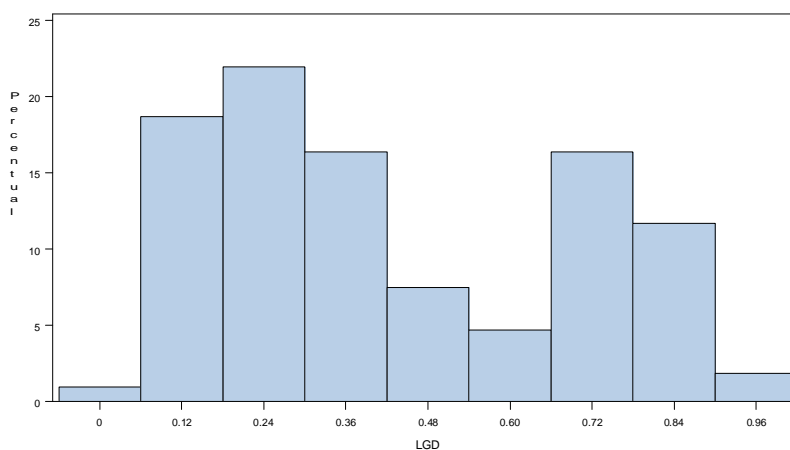
4. RESULTADOS

Nessa seção são apresentados os resultados do estudo, que abrangem a análise exploratória da base de dados, o modelo de regressão logística obtido, a árvore de decisão obtida e o comparativo entre os modelos.

4.1. Análise Exploratória

Notou-se a existência de dois picos para o LGD (ver Figura 2), conforme já observado na literatura (HURT; FELSOVALYI, 1998; QUERCI, 2005; SCHUERMANN, 2004). Há um grupo com baixo valor de LGD (inferior a 50%) e um grupo de alto valor de LGD (superior a 50%).

Figura 2 - Distribuição do LGD



Fonte: Elaborado pelos autores

Esse comportamento justifica a escolha dos métodos de modelagem, que utilizam o LGD como resposta a uma variável binária (regressão logística e árvore de decisão). Como existem duas concentrações, um grupo com LGD menor que 50% e um grupo com LGD maior que 50%, esse valor (50%) será utilizado para definição da variável resposta binária ($Y_i = 1$ para grupos com LGD maior ou igual a 50%, e $Y_i = 0$ para grupos com LGD menor que 50%). Têm-se 136 grupos homogêneos de risco alocados ao grupo $Y_i = 0$ e 78 alocados ao grupo de $Y_i = 1$.

A Tabela 3 apresenta a média, o desvio padrão e o coeficiente de variação para as variáveis PD (%) e EAD (milhões de dólares) nos dois grupos de LGD. Nota-se que a média da variável PD é semelhante para os dois grupos. A média da exposição (EAD) é maior no grupo de baixo LGD.

Tabela 3 - Média da PD e EAD para as categorias de LGD

Grupo	PD (%)			EAD (milhões de dólares)		
	Média	Desvio-Padrão	Coeficiente de Variação	Média	Desvio-Padrão	Coeficiente de Variação
Baixo LGD	7,4	15,2	205,41	19.283	38.753	200,97
Alto LGD	8,8	17	193,18	4.967	10.134	204,43

Fonte: Elaborado pelos autores

A Tabela 4 apresenta o total de grupos homogêneos para a classificação de segmento Basileia nos dois grupos de LGD. A maior parte dos grupos homogêneos de risco classificados como empréstimos imobiliários e PME obtiveram baixo LGD, indicado menor risco nesses segmentos. Nota-se também que a maior parte dos grupos homogêneos de risco classificado como crédito rotativo obteve alto LGD, indicando o maior risco referente a este segmento.

Tabela 4 - Quantidade de grupos de cada segmento Basileia para as categorias de LGD

Grupo	Empréstimos Imobiliários	Outros	Crédito Rotativo	PME
Baixo LGD	75	29	17	15
Alto LGD	1	31	41	5

Fonte: Elaborado pelos autores

A Tabela 5 apresenta uma estatística descritiva do LGD por país. Dado as diferenças de regulação entre os países, pode haver algum país com maior LGD. Dentro de cada grupo (baixo LGD e alto LGD) os quartis são semelhantes entre si, a mediana está próxima da média e o desvio-padrão é aproximadamente 11% para ambos os grupos. Dadas estas características, consideraremos a média de todos os países no trabalho.

Tabela 5 – Estatísticas descritivas do LGD por país

Estatística	Baixo LGD	Alto LGD
1º Quartil	20,01%	58,27%
2º Quartil	25,06%	72,19%
3º Quartil	27,05%	75,46%
Média	24,75%	74,92%
Desvio Padrão	11,34%	10,52%

Fonte: Elaborado pelos autores

4.2. Regressão Logística

Para o modelo de regressão logística, as variáveis EAD e segmento foram significativas para estimar a probabilidade de alto LGD, considerando um nível descritivo de 0,05. A Tabela 6 apresenta os coeficientes estimados e o nível descritivo associado a essas variáveis. A variável PD não discrimina o LGD, pois apresentou um nível descritivo de 0,5634. Pelos coeficientes do modelo nota-se que o segmento imobiliário tem menor probabilidade de alto LGD. Percebe-se também que, se o grupo tem valor de EAD menor do que 175 milhões de dólares, ele tem menor chance de obter um alto LGD.

Tabela 6 - Variáveis, coeficientes estimados e nível descritivo da regressão logística

Variável	Coeficientes Estimados	Nível Descritivo
Constante	1,0762	0,0005
EAD < 175 milhões de dólares	-1,2210	0,0282
EAD ≥ 175 milhões de dólares	0,0000	.
Segmento Imobiliário	-5,3652	<0,0001
Segmento PME	-1,9908	0,0010
Segmento Outros	-0,8759	0,0277
Segmento Rotativo	0,0000	.

Fonte: Elaborado pelos autores

A Tabela 7 apresenta o Odds Ratio para o segmento. Nota-se que o segmento imobiliário é o menos arriscado no que se refere à probabilidade de alto LGD, seguido por PME, por outros e, por fim, por crédito rotativo. Quando o contrato pertence ao segmento imobiliário, a chance de ter alto LGD em relação a baixo LGD é 0,005 vezes menor que os contratos do segmento rotativo.

Tabela 7 - Odds Ratio para o segmento

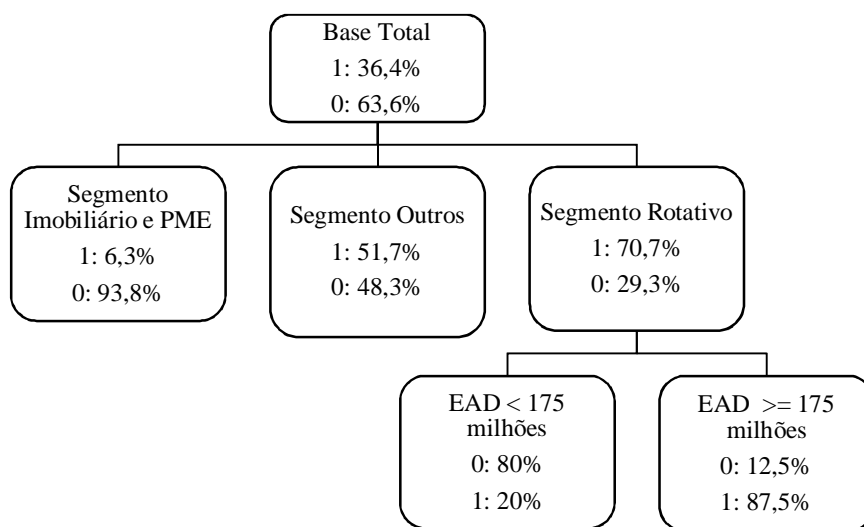
Descritivo	ODDS
Segmento Imobiliário vs. Rotativo	0,005
Segmento PME vs. Rotativo	0,137
Segmento Outros vs. Rotativo	0,416

Fonte: Elaborado pelos autores

O modelo faz a previsão correta para 79% das observações, sendo que, dos grupos que apresentaram baixo LGD, o modelo classificou corretamente 74% das observações, e, dos grupos que apresentaram alto LGD, o modelo classificou corretamente 87% das observações.

4.3. Árvore de Decisão

A árvore de decisão obtida na modelagem é apresentada na Figura 3. Verifica-se que o segmento é a variável mais influente na classificação do LGD. O segmento imobiliário e PME apresentam probabilidade 93,8% de obter baixo LGD. O segmento outros apresenta probabilidade 48,3% de obter baixo LGD e o segmento rotativo deve ser combinado com a variável EAD. Quando a variável EAD apresentar valor menor que 175 milhões de dólares dentro do segmento rotativo, há uma probabilidade de 20% de obter baixo LGD. Porém se o valor do EAD for maior ou igual a 175 milhões de dólares a observação apresenta 87,5% de probabilidade de obter baixo LGD.

Figura 3 - Árvore de Decisão

Fonte: Elaborado pelos autores

A árvore de decisão faz a previsão correta para 79% das observações, sendo que, dos grupos que apresentaram baixo LGD, o modelo classificou corretamente 71% das observações, e, dos grupos que apresentaram alto LGD, o modelo classificou corretamente 91% das observações.

4.4. Comparativo

A Tabela 8 apresenta o comparativo das classificações resultantes do modelo de regressão logística e do modelo de árvore de decisão. Nota-se que ambos os modelos têm assertividade semelhante. O modelo de árvore de decisão obtém vantagem na previsão de alto LGD, classificando corretamente 91,0% destas observações, contra 87,2% do modelo logístico. O modelo logístico obtém vantagem quando na previsão baixo LGD e classifica corretamente 74,3% das observações.

Tabela 8 - Comparativo de assertividade dos modelos logístico e árvore de decisão

Realizado	Previsto			
	Regressão Logística		Árvore de Decisão	
	Baixo LGD	Alto LGD	Baixo LGD	Alto LGD
Baixo LGD	101 74,3%	35 25,7%	97 71,3%	39 28,7%
Alto LGD	10 12,8%	68 87,2%	7 9,0%	71 91,0%
Total	79,0%		78,5%	

Fonte: Elaborado pelos autores

4.5. Previsão do LGD

A segunda fase da modelagem do LGD é a atribuição da média de cada grupo ($LGD \leq 50\%$ ou $LGD > 50\%$) como previsão para o LGD. Este método é vantajoso por ter simples aplicação e não necessita de ajuste de distribuição - problema apresentado em trabalhos que optam por uma metodologia de modelagem estatística nessa fase (THOMAS, L et al., 2010; LEOW; MUES, C, 2012).

A Tabela 9 apresenta os valores preditos de LGD.

Tabela 9 - Previsão do LGD para os grupos de baixo e alto LGD

Grupo	LGD Predito
Baixo LGD ($LGD \leq 50\%$)	24,75%
Alto LGD ($LGD > 50\%$)	74,92%

Fonte: Elaborado pelos autores

O LGD predito por meio do método proposto obteve uma soma dos quadrados dos erros (SQE) de 9,91. Para comparar o método proposto com uma metodologia tradicional de modelagem, foi construída uma regressão linear simples, e foi obtida uma soma dos

quadrados dos erros (SQE) de 13,82. Portanto, o método proposto se apresentou mais adequado para modelagem do LGD.

5. CONSIDERAÇÕES FINAIS

Esse estudo teve, como objetivo, propor uma técnica de modelagem e de previsão do LGD, de simples aplicação, em duas fases: na primeira fase, um modelo prevê a probabilidade de o LGD ser alto (maior que 50%) por meio de uma árvore de decisão ou de uma regressão logística. Na segunda fase, o LGD é predito por meio da média de cada grupo: alto LGD (maior que 50%) e baixo LGD (menor que 50%). O diferencial dessa técnica é a proposta de utilização da média na segunda fase. O método foi testado em uma base de dados de 12 instituições financeiras e foram utilizadas as variáveis EAD, PD e segmento Basileia (empréstimos imobiliários, crédito rotativo, PME e outros).

O segmento do contrato (imobiliário, crédito rotativo, outros ou PME) e o tamanho da exposição no momento do descumprimento (EAD) são fatores influentes na probabilidade de alto LGD (1ª fase), confirmando resultados de Asarnow e Edwards (1995), de Hurt e Felsovalyi (1998), de Gupton, Gates e Carty (2000), de Schuermann (2004), de Dermine e Carvalho (2006), de Belotti e Crook (2008), de Chalupka e Kopecsni (2008) e de Silva, Marins e Neves (2009). Ambos os modelos obtiveram aproximadamente 79% de acerto total, sendo que o modelo de árvore de decisão obteve vantagem quando se trata do grupo de alto LGD, e o modelo logístico obteve vantagem quando se trata do grupo de baixo LGD. Isso sugere que o modelo indicado depende do foco gerencial de uso da previsão do LGD.

A segunda fase consistiu em prever o LGD por meio da média dos grupos de alto LGD e de baixo LGD, considerando os resultados do modelo da primeira fase. Os resultados finais do método proposto (combinação das duas fases) se configuraram adequados, pois apresentaram menor soma dos quadrados dos erros quando comparados com os resultados do método de regressão linear simples.

Para futuros estudos recomenda-se estudar a influência do uso dos modelos na gestão na perda, pois esses são obrigatórios para candidatura de Basileia. Conforme as instituições diferenciam melhor seus clientes na recuperação de crédito, por exemplo, podem tornar essa recuperação mais eficiente e obter menor perda. Um modelo de LGD baseado na política interna de recuperações de crédito é eficiente na sua previsão (HAN; JANG, 2013).

Como limitações do trabalho, a base de dados utilizada considera diversos países com média de LGD diferentes entre si, e pode-se, em trabalhos futuros, avaliar se fatores afetam os países de maneira diferente.

Referências

ASARNOW, E.; EDWARDS, D. Measuring Loss on Defaulted Bank Loans: A 24-Year Study. **Journal of Commercial Lending**, v. 77, n. 7, p. 11-23, 1995.

Revista ADM.MADE, Rio de Janeiro, ano 15, v.19, n.2, p.1-20, maio/agosto, 2015.



- BASTOS, J. A. Forecasting bank loans loss-given-default. **Journal of Banking & Finance**, v. 34, n. 10, p. 2510-2517, out 2010.
- BELLOTTI, T.; CROOK, J. **Modelling and estimating Loss Given Default for credit cards**. Credit Research Centre Working Paper, n. 8, Edinburgo, 2008.
- BREIMAN, L.; FRIEDMAN, J.; OLSHEN, R.; STONE, C. **Classification and regression trees**. Monterey: Chapman and Hall, 1ª ed. 1984.
- CAOQUETTE, J. B.; ALTMAN, I.; NARAYANAN, P. **Gerenciando o risco de Crédito: o grande desafio para os mercados financeiros globais**. Wiley John & Sons, 1ª ed. 1998.
- CHALUPKA, R.; KOPECSNI, J. Modeling Bank Loan LGD of Corporate and SME Segments : A Case Study*. **Czech Journal of Economics and Finance**, v. 59, n. 4, p. 360-382, 2009.
- DERMINE, J.; CARVALHO, C. N. Bank loan losses-given-default: A case study. **Journal of Banking & Finance**, v. 30, n. 4, p. 1219-1243, abr 2006.
- FEDERAL RESERVE BANK. **Data download program home: consumer credit**. Disponível em: <<http://www.federalreserve.gov/datadownload/default.htm>>. Acesso em: 8 fev. 2014.
- GUPTON, G.; GATES, D.; CARTY, L. V. **Bank-Loan Loss Given Default**. Nova York, 2000.
- GUPTON, G.; STEIN, R. **LossCalc: Model for Predicting Loss Given Default**. Nova York, 2002.
- GÜRTLER, M.; HIBBELN, M. Improvements in loss given default forecasts for bank loans. **Journal of Banking & Finance**, v. 37, n. 7, p. 2354-2366, jul 2013.
- HAN, C.; JANG, Y. Effects of debt collection practices on loss given default. **Journal of Banking & Finance**, v. 37, n. 1, p. 21-31, jan 2013.
- HARTMANN-WENDELS, T.; MILLER, P.; TÖWS, E. Loss given default for leasing: Parametric and nonparametric estimations. **Journal of Banking & Finance**, v. 40, p. 364-375, mar 2014.
- HOSMER, D.; LEMESHOW, S. **Applied Logistic Regression**. John Wiley & Sons, 2ª ed. Nova York, 1989.
- HURT, L.; FELSEVALYI, A. Measuring Loss on Latin American Defaulted Bank Loans: A 27-Year Study of 27 Countries. **Journal of Lending & Credit Risk Management**, v. 81, n. 2, p. 41-46, 1998.
- LEOW, M.; MUES, C. Predicting loss given default (LGD) for residential mortgage loans: A two-stage model and empirical evidence for UK bank data. **International Journal of Forecasting**, v. 28, n. 1, p. 183-195, jan 2012.
- LOTERMAN, G.; BROWN, I.; MARTENS, D.; MUES, CHRISTOPHE; BAESENS, B. Benchmarking regression algorithms for loss given default modeling. **International Journal of Forecasting**, v. 28, n. 1, p. 161-170, jan 2012.
- MILEO, R.; KIMURA, H.; KAYO, E. K. Análise do Modelo CreditRisk + em uma amostra de portfólio de crédito. **Revista Contemporânea de Economia e Gestão**, v. 11, n. 1, p. 103-116, 2013.
- PARK, Y. W.; BANG, D. W. Loss given default of residential mortgages in a low LTV regime: Role of foreclosure auction process and housing market cycles. **Journal of Banking & Finance**, v. 39, p. 192-210, fev 2014.
- QI, M.; YANG, X. Loss given default of high loan-to-value residential mortgages. **Journal of Banking & Finance**, v. 33, n. 5, p. 788-799, maio 2009.
- QUERCI, F. Loss Given Default on a medium-sized Italian bank's loans: an empirical exercise. **European Financial Management Association**. Milão, 2005.
- SANTOS, J. Bank Capital Regulation in Contemporary Banking Theory: A Review of the Literature. **Financial Markets, Institutions & Instruments**, v. 10, n. 2, p. 41-84, 2001.
- SCHUERMANN, T. **What Do We Know About Loss Given Default?** In: SHIMKO, D. (Ed.). Credit Risk: Models and Management. UK Risk Book, 2ª ed. Londres, 2004.

SILVA, A.; MARINS, J.; NEVES, M. **Loss Given Default: Um Estudo sobre Perdas em Operações Prefixadas no Mercado Brasileiro**. XXXII Encontro da ANPAD. Rio de Janeiro, 2008.

STEINER, M.; CARNIERI, C.; KOPITKE, B.; NETO, P. Sistemas especialistas probabilísticos e redes neurais na análise do crédito bancário. **Revista de Administração**, v. 34, n. 3, p. 56-67, 1999.

THOMAS, L.; MATUSZYK, A.; MOORE, A. Comparing debt characteristics and LGD models for different collections policies. **International Journal of Forecasting**, v. 28, n. 1, p. 196-203, jan 2012.

THOMAS, L.; MUES, C.; MATUSZYK, A. Modelling LGD for unsecured personal loans: decision tree approach. **Journal of the Operational Research Society**, v. 61, n. 3, p. 393-398, 14 out 2010.

TONG, E. N. C.; MUES, CHRISTOPHE; THOMAS, LYN. A zero-adjusted gamma model for mortgage loan loss given default. **International Journal of Forecasting**, v. 29, n. 4, p. 548-562, out 2013.

WITTEN, I.; FRANK, E. **Data mining: practical machine learning tools and techniques with Java implementations**. 1ª ed. Nova York: Morgan Kaufmann, 1999. p. 371.

ZHANG, J.; THOMAS, L. Comparisons of linear regression and survival analysis using single and mixture distributions approaches in modelling LGD. **International Journal of Forecasting**, v. 28, n. 1, p. 204-215, jan 2012.