LGD

Douglas Beserra Pinheiro[[1]](#footnote-2)\*

*Paketá Crédito*

**RESUMO**

*xxxxx.*

*xxxxxx*

*xxxxx*

*xxxxxx*

**Palavras-chave**: *IPO*, *IPO*, estabilização de preços, *aftermarket short covering*

**Classificação JEL**: G24

1. Introdução

Xxxxxxxxxxxxxxx xxxxxxxxxxxxx xxxxxxxxxxxxxxxx xxxxxxxxxxxxxxxxxxxxxx xxxxxxxx.

FALAR DO LGD, SUA IMPORTÂNCIA E DESAFIOS.

ESTUDOS NO BRASIL

ESTUDOS NO MUNDO E O USO DE ML

EXISTEM POUCOS ESTUDOS

Este artigo está organizado da seguinte forma. A Seção 2 detalha o funcionamento do processo de estabilização. A Seção 3 descreve os dados e variáveis. A Seção 4 apresenta as hipóteses e a metodologia empregada. A Seção 5 apresenta e discute os resultados empíricos. Finalmente, a Seção 6 conclui este artigo.

2. Machine Learning e Crédito (Algo assim)

Segundo

2000) reporta que os *ASC* chega a ser realizado ao longo de meses.

3. Metodologia

**Explicar os diferentes tipos de LGD (work-out) -> tem naquele paper do BNDES**

**Descrever Hipóteses**

4. Dados e Variáveis

Esta seção descreve a fonte de informação utilizada e apresenta uma discussão sumária dos dados e tratamentos.

4.1. Dados

A base com as informações dos empréstimos foi obtida do Kaggle[[2]](#footnote-3) e possui 141 variáveis e 2.925.493 empréstimos realizadas entre 2007 e 2020 pelo *Lending Club*. Utilizando o mesmo critério de Zhou et al. (2018) foram selecionados todos os empréstimos na condição de baixados por inadimplência, o que normalmente ocorre 120 dias após o vencimento da operação. No entanto, diferente de Zhou et al. (2018) e para melhorar a apuração do LGD foi incluído o custo de recuperação cobrado pelo *Lending Club* dos investidores, que representa em média 17,5% da recuperação observada. O LGD, portanto, foi definido considerando o seguinte critério:

|  |  |  |  |
| --- | --- | --- | --- |
| *LGD = 1 –* [ | *(Valor Recuperado – Custo de Cobrança)* | ] | (1) |
| *(Valor do Empréstimo – Principal Pago)* |

A figura 1 apresenta a quantidade de operações baixadas por inadimplência, o que soma 363.309 empréstimos. O LGD médio gira em torno de 90%, porém a quantidade de registros com LGD igual a 100% (ou seja, aqueles casos em que não se observa nenhuma recuperação) aumenta significativamente após 2018 em função do menor período de apuração da recuperação, e é significativamente menor que a média anterior a 2014, portanto os empréstimos realizados antes de 2014 e após 2017 foram excluídos da análise. Também foram excluídas 31 operações cujo valor principal do empréstimo foi integralmente pago, impedindo o cálculo do LGD como proposto aqui, o que resultou em uma amostra final de 266.515 observações no período entre 2014 e 2017. Apenas 17 observações atípicas apresentaram LGD < -0,3 e foram mantidas limitando o valor nesse nível.

**Figura 1**

**Quantidade de Empréstimos e LGD Médio por Ano**

Gráfico, Histograma

Descrição gerada automaticamente

A variável LGD refere-se àquela apresentada na equação 1, já a variável LGD=1 refere-se à proporção de empréstimos que não apresentaram recuperação, ou seja, quando o LGD é igual a 100%.

Zhou et al. (2018) excluem operações cujos clientes não tiveram renda verificada e operações com informações faltantes, reduzindo a amostra em 87%. No caso da renda verificada não há diferença relevante no LGD observado e para os casos de informações faltantes, além de serem pouco representativos também consideramos essas situações no desenvolvimento do estudo, portando todas foram todas mantidas.

Quanto ao custo do dinheiro no tempo não foi considerado pois não há a informação de data de inadimplência e data de recuperação, além disso, o efeito não deve ser relevante pois o work-out do LGD se estabiliza em média 2 anos após a data da concessão, conforme observado no gráfico 1.

A figura 2 apresenta o LGD distribuído nos diferentes anos da amostra. Observa-se uma concentração da distribuição à direita, o que é esperado para um produto de varejo sem garantias associadas, que acarreta um elevado nível de perda após a ocorrência da inadimplência. Além disso, não há variação relevante entre os diferentes anos.

**Figura 2**

**Distribuição do LGD**

Interface gráfica do usuário

Descrição gerada automaticamente com confiança baixa

4.2. Variáveis

Variáveis com baixa representatividade ou com impedimento do uso por aspectos conceituais foram excluídas. Das 52 exclusões 31 decorreram de baixo preenchimento (preenchimento de até 2,7%), 6 têm o mesmo preenchimento para todos os registros, 7 referem-se a dados futuros e/ou compõem o cálculo do LGD e as últimas 8 apresentaram deficiências diversas. Após a criação de algumas variáveis auxiliares e transformação em categorias binárias de algumas variáveis categóricas a base de dados ficou composta por 112 variáveis para a aplicação do modelo de Auto-ML, incluindo o LGD.

No processo de modelagem tradicional, inicialmente são selecionadas informações que possuem sentido econômico sobre o evento analisado. Optamos aqui por utilizar o mesmo conjunto de variáveis de Zhou et al. (2018) apresentadas na tabela 1, mostrando inclusive o efeito esperado sobre o LGD. Além delas, foram incluídas informações sobre o comportamento de crédito do cliente, saldo do contrato no momento da inadimplência e o uso declarado dos recursos, todas detalhadas na tabela 1.

|  |  |  |
| --- | --- | --- |
| **Tabela 1** | | |
| **Informações utilizadas na Modelagem Tradicional** | | |
| **Variável** | **Definição** | **Sinal esperado** |
| *loan\_amnt (1)* | Valor do empréstimo | (+) |
| *term\_c (1)* | Assume 1 quando o prazo do empréstimo é de 60 meses e 0 quando o prazo é 36 meses. | (+) |
| *int\_rate\_n (1)* | Taxa anual de juros do empréstimo | (+) |
| *grade (1)* | Rating de crédito definido na concessão, separado em 7 classes de A (menor risco de inadimplência) a G (maior risco) | (-) |
| *ead* | Valor do saldo do empréstimo no momento da inadimplência | (+) |
| *ead\_inc* | Valor do saldo do empréstimo no momento da inadimplência dividido pela renda anual do devedor. | (+) |
| *ead\_loan* | Valor do saldo do empréstimo no momento da inadimplência dividido pelo valor contratado. | (+) |
| *emp\_length\_n (1)* | Anos de emprego do devedor, limitado a 0,5 se for menor que 1 ano e 10 se for maior. | (-) |
| *home\_c (1)* | Classifica a moradia em (0) Alugada adotada como referência, (1) Financiada, (2) Outros e (3) Própria. | (-) |
| *annual\_inc (1)* | Renda anual do devedor | (-) |
| *dti (1)* | Comprometimento mensal de renda (parcela/salário) excluindo o empréstimo que está sendo analisado e eventual crédito imobiliário contratado. | (+) |
| *delinq\_2yrs (1)* | Número de atrasos acima de 30 dias nos últimos 2 anos | (+) |
| *mths\_since\_last\_delinq* | Número de meses desde que ocorreu o último atraso > 30 dias pelo devedor | (-) |
| *mths\_since\_last\_delinq\_null* | Indicador assume 1 quando não há histórico de atraso do devedor | (-) |
| *pub\_rec* | Número de registros negativos do devedor no mercado | (+) |
| *inq\_last\_12m* | Número de consultas de crédito do devedor nos últimos 12 meses | (+) |
| *purpose* | propósito do empréstimo: carro, cartão de crédito, consolidação de dívidas, educação, reforma, casa, compra de alto valor, médico, mudança, energia renovável, férias, casamento, outros e pequenos negócios, adotada como referência. | Não se aplica |
| *fico\_range\_high (low)* | Limite superior (inferior) do escore Fico (2) do devedor. | (-) |
| **Variável** indica o nome da variável utilizada na avaliação tradicional. **Definição** traz uma breve explicação sobre a construção da variável. **Sinal Esperado** indica se o efeito esperado da variável sobre o LGD é positivo (aumenta) ou negativo (diminui). **(1)** Variáveis também utilizadas por Zhou *et al.* (2018). **(2)** <https://www.fico.com/en/products/fico-score> | | |

As variáveis loan\_amnt e annual\_inc foram divididas por 1000 e como são relativamente assimétricas, para as regressões foi adotado o logaritmo natural dos seus valores, o mesmo feito com a variável *int\_rate\_n*. Outras transformações são possíveis, mas isso estava fora do escopo desse artigo. Na árvore de decisão elas foram utilizadas sem tratamento.

Além disso, foram criadas as variáveis ead, ead\_inc e ead\_loan que não fazem parte do conjunto original de informações, no entanto elas são muito importantes pois mostram o saldo devedor do cliente em termos absolutos, em função da renda e em função do valor originalmente emprestado e são apuradas no momento do default. Importante destacar que quanto maior o ead e o ead em função da renda mais difícil se torna a recuperação do crédito face à maior representatividade da dívida. No caso do ead\_loan, um maior valor significa que o default ocorreu próximo do momento da concessão, o que também reflete uma dificuldade maior do devedor e, portanto, uma menor expectativa de recuperação. Dessa forma é esperado que essas informações contribuam na estimativa do LGD e

Todas as demais informações derivam da base de dados original e alguns tratamentos complementares foram: 1) censura na variável ead\_inc no valor 30 que impactou apenas 89 observações, 2) censura na variável dti em 200 que impactou 56 observações, 3) censura na variável annual\_inc para valores menores de 100 e maiores de 1.000.000 que impactou 87 observações, 4) os campos nulos foram substituídos por zero e 5) foram criadas 5 variáveis complementares que indicavam essas substituições, um exemplo é a *mths\_since\_last\_delinq* onde um maior significa que o devedor apresentou um atraso há muito tempo, o que é um bom indicativo e relação àqueles que inadimpliram recentemente, enquanto o 0 (zero) significa que não há atraso no histórico, o que também é um bom indicativo. Dessa forma o efeito é não linear e a variável binária complementar *mths\_since\_last\_delinq\_null* busca capturar esse efeito.

Os sinais esperados da tabela 1 são discutidos na seção de resultados apenas no caso daquelas que compuseram as equações finais.

5. Resultados

.

6. Discussão e conclusão

A possibilidade

Referências

Zhou, G., Zhang, Y. e Luo S. (2018). P2P Network Lending, Loss Given Default and Credit Risks, *Sustainability* **10**(4): 1010.

URL: <https://doi.org/10.3390/su10041010>

1. \*Paketá Crédito, São Paulo, Brasil. Email: [douglas.pinheiro@paketa.com.br](mailto:douglas.pinheiro@paketa.com.br) [↑](#footnote-ref-2)
2. Utilizar o link <https://www.kaggle.com/code/ztrimus/loan-repayment-prediction/data?select=Loan_status_2007-2020Q3.gzip> para acesso à base e ao dicionário de dados. [↑](#footnote-ref-3)