

ANÁLISE DE FATORES DE RISCO EM FINANCIAMENTOS PEER-TO-PEER

Renato Martins Chavez

Resumo

As empresas que intermedeiam financiamentos Peer-to-Peer oferecem uma classificação individual de risco para cada tomador de crédito que auxilia os credores a avaliar o risco dos empréstimos. Porém, outras informações pessoais dos tomadores de crédito estão disponíveis nos pedidos de empréstimo. Meu objetivo é verificar se a combinação dessas informações ajuda a identificar o risco de empréstimos não serem pagos quando comparadas ao uso da classificação gerada pela empresa de forma isolada e com isso diminuir a assimetria de informação entre credores e tomadores de crédito. A base de dados disponibilizada pela plataforma P2P do Lending Club dos anos de 2012 a 2018 foi utilizada com um total de aproximadamente um milhão de observações em uma regressão Logit, testando cinco modelos diferentes de forma a verificar qual combinação de informações disponíveis apresenta a melhor capacidade de distinguir empréstimos pagos de não pagos. O modelo que combina todas as variáveis analisadas no estudo apresenta o melhor poder de predição, com uma AURoc de 0,688. Também se observou que nem todas as categorias da variável do objetivo do empréstimo são significativas para prever se o empréstimo será pago. A partir do resultado dos modelos pode-se inferir que as informações utilizadas no estudo tem significância para a classificação de empréstimos pagos dos não pagos, porém o maior poder explicativo vem da própria classificação gerada pelo Lending Club, pois a diferença do poder explicativo do modelo Logit simples com apenas a variável *grade* como independente é pequena quando comparamos ao melhor modelo que usa todas as variáveis.

Introdução

Com o desenvolvimento da Internet surgiu uma alternativa no mercado tradicional de empréstimos composto pelas instituições financeiras tradicionais, os modelos de financiamento Peer-to-Peer (P2P). Trata-se de uma plataforma online em que indivíduos credores e tomadores de crédito podem negociar diretamente através de uma empresa que apenas realiza o intermédio entre as transações. Este modelo de negócio oferece diversas vantagens tanto para os credores, que recebem um retorno maior de seus investimentos comparado a investimentos conservadores como poupança e renda fixa, quanto para os tomadores de crédito que pagam juros menores comparados aos de instituições financeiras tradicionais. Isto é possível pois as empresas P2P sofrem com menores custos por não serem sujeitas aos requerimentos de bancos de capitais, não coletarem depósitos, e não necessitarem de custos com ativos e com funcionários especificamente para a análise de risco de crédito. (SERRANO;GUTIÉRREZ;LÓPEZ, 2015)

Este risco de crédito é uma preocupação constante para plataformas de financiamento pois não é simples quantificar o risco individual de não receber os pagamentos dos tomadores de crédito. Grandes agências bancárias conseguem lidar melhor com este problema devido a enormes fundos de investimento que permitem uma grande diversificação de portfólio e acesso a muitas informações dos clientes, diminuindo a assimetria de informação em que o credor desconhece a responsabilidade e capacidade de pagamento de um determinado tomador de crédito tão bem quanto ele próprio. A responsabilidade da análise de risco de crédito nas plataformas P2P é feita pelo próprio credor conforme suas preferências de risco. Porém os credores não costumam ser especialistas em análise de risco e nem possuem a grande quantidade de informações disponíveis ao banco para diminuir a assimetria de informação, fazendo com que seja necessário a existência de mecanismos que os auxiliem nessa tomada de decisão.

A empresa analisada nesse estudo, o Lending Club, auxilia os credores através de um algoritmo que classifica os tomadores de crédito quanto ao seu risco, essa classificação vai de A1 que representa o menor risco e por tanto tem o menor retorno médio esperado a G5 que representa o maior risco e o maior retorno médio. Além disso, os tomadores de crédito precisam listar algumas informações pessoais como idade, gênero, estado de trabalho, se possui casa própria, salário e afins. Neste contexto, entender quais informações disponíveis representam o risco de crédito e

diminuem a assimetria de informação se torna relevante para a saúde financeira dos investidores e da empresa.

O objetivo deste estudo é o de entender quais informações listadas pelos tomadores de crédito têm significância para quantificar seu risco na plataforma P2P do Lending Club. Através de uma regressão Logit binária com o status do empréstimo (pago / não pago) como variável dependente, serão analisadas as informações listadas pelos tomadores de crédito como variáveis independentes afim de observar quais delas tem significância para o risco de o empréstimo não ser pago, e será construído um modelo para prever esse risco.

Referencial teórico e hipótese de pesquisa

A primeira plataforma comercial online P2P apareceu em 2005, e vem crescendo globalmente desde então.(BACHMANN; FUNK, 2011; RUIQIONG; JUNWEN, 2014) Apenas o Lending Club gerou 44 bilhões em empréstimos no período de 2012 até 2018, uma porcentagem pequena comparada as grandes instituições financeiras, porém que demonstra uma demanda crescente da população sobre serviços alternativos de empréstimos.

Devido ao crescimento dessas plataformas apareceram três tópicos importantes de pesquisa. O estudo sobre os motivos pelo qual esses serviços surgiram, os fatores que explicam quais pedidos de crédito tem sucesso e o risco de crédito, e por último o desempenho de um empréstimo P2P dado um certo nível de risco. (EMEKTER *et al.*, 2015)

Este estudo está inserido dentro do segundo tópico de pesquisa, especificamente em quais fatores explicam o risco crédito. Existem alguns estudos que pesquisam os determinantes do risco de crédito em plataformas P2P. Liu *et al.* (2018) demonstram que o uso de um modelo Logit no mercado P2P chinês diminui a assimetria de informação usando as características do empréstimo, taxa de juros e detalhes pessoais dos tomadores de crédito. Emekter *et al.* (2015) descobrem que a pontuação FICO, revolving line e o endividamento são importantes para a análise de risco de crédito através do uso de uma regressão Logit. Serrano, Gutiérrez e López (2015) analisam os determinantes do risco de crédito através de análise univariada e de sobrevivência, e usando uma regressão logística concluem que apesar da pontuação FICO ser o fator mais importante para o risco de crédito, a precisão do modelo melhora

com a adição de outras variáveis. Kumar *et al.* (2016) e Jin e Zhu (2015) utilizam algoritmos de auto aprendizagem para selecionar os determinantes de risco de crédito e melhorar o poder de precisão. Zhang (2018) utiliza da regressão linear múltipla para identificar os atributos que contribuem para o risco de crédito. Iyer *et al.* (2009) mostra que os credores conseguem avaliar com 45% melhor precisão o risco de crédito usando as informações listadas pelos tomadores de crédito do que usando apenas a pontuação de crédito como informação.

Devido a importância de análise de risco de crédito para os credores alocarem corretamente seus investimentos e da assimetria de informação que existe entre indivíduos, a seguinte hipótese de pesquisa será estudada:

H1. É possível melhorar a identificação do risco de crédito através de uma combinação das informações disponíveis.

Dados

Este artigo usa os dados públicos de empréstimos do Lending Club do período de 2012 até 2018. As variáveis utilizadas nos modelos foram selecionadas através da revisão de estudos anteriores realizados por Emekter *et al.* (2015), Liu *et al.* (2018), Pokorná e Sponer (2016), Serrano, Gutiérrez e López (2015), Kumar *et al.* (2016), Jin e Zhu (2015) e Zhang (2018). Estas variáveis são descritas na tabela 1.

A variável dependente *loan_status* foi tratada de acordo com a metodologia utilizada por Emekter *et al.* (2015) para se tornar binária e ser utilizável em uma regressão Logit, conforme descrito na tabela 1.

As seguintes variáveis apresentaram problemas com outlier e foram tratadas usando Winsorização nos percentis 1% e 99%: *Annual_inc*, *Dti*, *Installment*, *Mths_since_last_delinq*, *revol_bal* e *revol_util*.

Todas as variáveis falharam no teste de normalidade, por isso foi usado o método de Spearman de autocorrelação para testar problemas de multicolinearidade entre variáveis contínuas, os resultados são apresentados na tabela 2. O único problema de multicolinearidade surge entre *loan_amnt* e *Installment*, que apresenta uma

colinearidade quase perfeita de 0.96, que é esperado pois conforme o tamanho do empréstimo aumenta o pagamento mensal aumenta na mesma proporção.

A tabela 3 apresenta a porcentagem de empréstimos bons e ruins de cada variável discreta, que demonstra a importância da variável *Grade*, conforme a classificação diminui aumentando o risco a porcentagem de empréstimos bons diminui, e uma pequena variação entre as outras variáveis.

A tabela 4 apresenta a estatística descritiva das variáveis contínuas, é interessante notar que todas as variáveis têm uma distribuição assimétrica a direita, com exceção da variável *Revol_util*.

Estudo Empírico

Do ponto de vista do investidor, o importante é entender como classificar o risco de cada pedido de empréstimo disponível, nesse sentido a plataforma do Lending Club oferece uma classificação de risco para auxiliar o investidor que vai de A1 até G5, sendo A1 o menor risco e G5 o maior risco. Para avaliar se outras informações disponíveis impactam o risco dos empréstimos foram criados modelos usando as informações disponíveis dos tomadores de crédito além da classificação oferecida pelo Lending Club.

A tabela 5 apresenta o resultado dos modelos utilizando a regressão Logit. O primeiro modelo foi feito usando apenas a variável categórica de classificação do Lending Club, *grade*. Todas as variáveis se mostraram significativas, e assim como o esperado, conforme a classificação diminui o risco aumenta pois o coeficiente é maior e positivo, aproximando o resultado da variável dependente *loan_status* de 1, que representa os empréstimos classificados como ruins. O R^2 é adequado apenas para regressões lineares, devido a isso foram usadas duas medidas de Pseudo R^2 apropriadas para a regressão Logit, o R^2 de Nagelkerke e Mcfadden. Observa-se que o poder explicativo do modelo é baixo com R^2 de Nagelkerke de 8,4% e Mcfadden de 5,92%. Em problemas de classificação como o deste artigo, em que o objetivo do modelo é prever os empréstimos com bons pagadores como 0 e pagadores ruins como 1, uma medida muito utilizada de performance de predição é o AUROC (Area Under the ROC Curve). De forma simplificada, a ROC Curve é uma curva de probabilidade e a AUC (Area

under the curve) representa a performance do modelo. Uma AUC de 0.5 não tem nenhum poder de predição, pois a chance de prever um resultado entre 0 e 1 é naturalmente de 50%, quanto mais a AUC se aproxima de 1 maior seu poder de predição, ou seja, prever os resultados 1 como 1 e os resultados 0 como 0. Um modelo com poder de predição razoável tem uma AUC superior ou igual a 0.7. No modelo 1 temos uma AUC de 0,669, indicando que apesar de o modelo ter algum poder de predição, ele não é muito alto.

Os modelos 2 e 3 incrementam ao modelo 1 todas as variáveis contínuas e a discreta *emp_length*, sendo que o 2 usa a variável *installment* e o 3 usa a variável *loan_amnt* pois são variáveis com alta correlação. Nota-se que nesses modelos os pseudo R^2 e AUC aumentam ligeiramente, porém o AUC continua abaixo de 0,7. O modelo 4 usa a variável principal *Grade* junto com as variáveis discretas, e ele apresenta resultados ligeiramente piores ao modelo com as variáveis contínuas. O modelo 5 usa todas as variáveis disponíveis e usa do *loan_amnt* ao invés do *installment*, e apresenta o melhor resultado de todos os modelos, porém a diferença é pequena e sua AUC ainda não chega aos 0,7, indicando que a variável mais importante é de fato a classificação *grade* oferecida pelo Lending Club, e as outras variáveis apesar de oferecerem um poder explicativo quanto a qualidade do empréstimo, não aumentam muito a qualidade do modelo.

Conclusão

O presente estudo realizou uma análise dos fatores que podem influenciar o risco dos investimentos além da classificação de risco oferecida pela empresa na plataforma P2P conhecida como Lending Club, através do uso da regressão Logit, com o objetivo de melhorar a identificação do risco de crédito.

Foi observado que o modelo que usa todas as variáveis explica melhor o risco de crédito e que todas as variáveis utilizadas são significativas, com exceção da variável *purpose* em que suas únicas categorias significativas são:

Debt_consolidation, *Home_improvement*, *House*, *Other*, *Renewable_energy* e *small_business*.

Verificou-se que apesar de uma combinação de todas as variáveis utilizadas no estudo melhorarem o poder explicativo do modelo, esta melhora é sutil e o maior poder explicativo se dá pela própria classificação oferecida pelo Lending Club.

Dada à grande heterogeneidade dos dados, futuros estudos podem categorizar os dados por diferentes critérios como faixas de renda com o objetivo de observar diferentes influências das variáveis para diferentes perfis de tomadores de crédito, diminuindo a assimetria de informação entre credor e tomador de crédito.

Referências bibliográficas

BACHMANN, A.; FUNK, B. Journal of Internet Banking and Commerce. n. January 2011.

EMEKTER, R. et al. Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending. **Applied Economics**, v. 47, n. 1, p. 54–70, 2015.

IYER, R. et al. Screening Peers Softly: Inferring the quality of small borrowers. p. 1-62, 2009.

JIN, Y.; ZHU, Y. A data-driven approach to predict default risk of loan for online Peer-to-Peer (P2P) lending. Fifth International Conference on Communication Systems and Network Technologies, p. 609–613, 2015.

KUMAR, V.. et al. Credit Risk Analysis in Peer-to-Peer Lending System. **IEEE International Conference on Knowledge Engineering and Applications (ICKEA)**, p. 193–196, 2016.

LIU, Y. et al. Can Listing Information Indicate Borrower Credit Risk in Online Peer-to-Peer Lending ? **Emerging Markets Finance and Trade**, v. 54, n. 13, p. 2982–2994, 2018.

POKORNÁ, M.; SPONER, M. Social Lending and Its Risks. **Procedia - Social and Behavioral Sciences**, v. 220, n. March, p. 330–337, 2016.

RUIQIONG, G. A. O.; JUNWEN, F. An Overview Study on P2P Lending. v. 8, n. 2, p. 14–18, 2014.

SERRANO-CINCA, C.; GUTIÉRREZ-NIETO, B.; LÓPEZ-PALACIOS, L.

Determinants of Default in P2P Lending. p. 1–22, 2015.

ZHANG, Y. Research on Credit Scoring Mechanism of P2P Lending Platform. v. 181, n. Icsshe, p. 883–886, 2018.

Anexos

Tabela 1 – Descrição das variáveis do estudo

| Variável | Definição |
|------------------------|--|
| <i>Dependente</i> | |
| loan_status | Estado atual de pagamento do empréstimo, sendo que empréstimos totalmente pagos e com pagamentos em dia são classificados como bons (0) e empréstimos não pagos e atrasados são classificados como ruins (1). |
| <i>Contínuas</i> | |
| Annual_inc | A renda anual declarada pelo tomador de crédito. |
| Dti | A proporção entre o total de pagamentos mensais devidos pelo tomador de crédito, excluindo hipoteca e o empréstimo pedido no Lending Club, dividido pela sua renda mensal auto declarada. |
| Installment | O pagamento mensal que será devido pelo tomador de crédito caso o empréstimo seja realizado. |
| Int_rate | Taxa de juros do empréstimo. |
| Mths_since_last_delinq | O número de meses desde a última inadimplência do tomador de crédito. |
| Revol_bal | Total de crédito rotativo disponível. |
| Revol_util | Taxa de crédito rotativo usada pelo tomador de crédito em relação ao total de crédito rotativo disponível. |
| Loan_amnt | Valor total do empréstimo. |
| <i>Discretas</i> | |
| Emp_length | Anos de trabalho do tomador de crédito, com valores possíveis entre 0 a 10. |
| Grade | Classificação de risco do tomador de crédito feita pelo Lending Club, com possíveis valores de A a G, com A sendo o menor risco e G o maior risco. |
| Home_Ownership | O estado atual de moradia do tomador de crédito, com possíveis valores: Casa própria, alugada ou em hipoteca. |
| Purpose | Uma categoria com o objetivo do empréstimo preenchida pelo tomador de crédito. As categorias são: Cartão de crédito, Consolidação de dívida, Melhorias para casa, Casa, Compra grande, Despesas médicas, Mudança, Energia renovável, Pequeno empreendimento, Férias, Casamento e outros. |
| Term | O número de pagamentos do empréstimo, os valores são de 36 ou 60 meses. |

Fonte: Próprio autor.

Tabela 2 – Correlação das variáveis contínuas

| | Annual_inc | Dti | Installment | Int_rate | loan_amnt | Mths_since _last_delinq | Revol_bal | Revol_util |
|------------------------|------------|--------|-------------|----------|-----------|----------------------------|-----------|------------|
| Annual_inc | 1 | -0.248 | 0.420 | -0.124 | 0.442 | -0.062 | 0.383 | 0.093 |
| Dti | -0.248 | 1 | 0.039 | 0.174 | 0.031 | 0.012 | 0.230 | 0.155 |
| Installment | 0.420 | 0.039 | 1 | 0.135 | 0.96 | -0.017 | 0.418 | 0.133 |
| Int_rate | -0.124 | 0.174 | 0.135 | 1 | 0.12 | -0.049 | -0.039 | 0.210 |
| loan_amnt | 0.442 | 0.031 | 0.96 | 0.12 | 1 | -0.008 | 0.432 | 0.118 |
| Mths_since_last_delinq | -0.062 | 0.012 | -0.017 | -0.049 | -0.008 | 1 | 0.010 | 0.008 |
| Revol_bal | 0.383 | 0.230 | 0.418 | -0.039 | 0.432 | 0.010 | 1 | 0.442 |
| Revol_util | 0.093 | 0.155 | 0.133 | 0.210 | 0.118 | 0.008 | 0.442 | 1 |

Fonte: Próprio autor.

Tabela 3 – Estudo das variáveis discretas

| Empréstimos bons (%) | | |
|-----------------------|------|------|
| Variável | Sim | Não |
| <i>Emp_length</i> | | |
| 1 | 84.5 | 15.5 |
| 2 | 85.2 | 14.8 |
| 3 | 85.3 | 14.7 |
| 4 | 85.4 | 14.6 |
| 5 | 85.4 | 14.6 |
| 6 | 85.6 | 14.4 |
| 7 | 85.1 | 14.9 |
| 8 | 84.4 | 15.6 |
| 9 | 85 | 15 |
| 10 | 86.4 | 13.6 |
| <i>Grade</i> | | |
| A | 95.7 | 4.3 |
| B | 90.6 | 9.4 |
| C | 84.6 | 15.4 |
| D | 78.4 | 21.6 |
| E | 70 | 30 |
| F | 62.3 | 37.7 |
| G | 57.2 | 42.8 |
| <i>Home_Ownership</i> | | |
| Own | 85.5 | 14.5 |
| Mortgage | 87.3 | 12.7 |
| Rent | 83.2 | 16.8 |
| <i>Purpose</i> | | |
| Credit_card | 88.4 | 11.6 |
| Debt_consolidation | 84.5 | 15.5 |
| Home_improvement | 87 | 13 |
| House | 85.4 | 14.6 |
| Major_purchase | 86.2 | 13.8 |
| Medical | 85.5 | 14.5 |
| Moving | 82.5 | 17.5 |
| Other | 85.8 | 14.2 |
| Renewable_energy | 81 | 19 |
| Small_business | 78.2 | 21.8 |
| Vacation | 87 | 13 |
| Wedding | 84.8 | 15.2 |
| <i>Term</i> | | |
| 36 Months | 87.8 | 12.2 |
| 60 Months | 79.9 | 20.1 |

Fonte: Próprio autor.

Tabela 4 – Estatística descritiva das variáveis contínuas

| | Min | 1ª Quartil | Mediana | Média | 3ª Quartil | Max | Desvio padrão |
|------------------------|--------|------------|---------|--------|------------|--------|---------------|
| Annual_inc | 4766 | 50000 | 70000 | 76653 | 100000 | 135234 | 32344.45 |
| Dti | -0.45 | 11.79 | 17.49 | 18.08 | 24 | 35.43 | 8.332 |
| Installment | 7.61 | 258.44 | 384.43 | 434.01 | 598.83 | 848.07 | 226.956 |
| Int_rate | 0.0531 | 0.10 | 0.129 | 0.134 | 0.16 | 0.309 | 0.047 |
| loan_amnt | 1000 | 8000 | 13000 | 15130 | 20000 | 40000 | 9085.99 |
| Mths_since_last_delinq | 0 | 16 | 31 | 34.33 | 50 | 81.41 | 21.736 |
| Revol_bal | 0 | 5659 | 10589 | 12641 | 18700 | 27921 | 8549.11 |
| Revol_util | 0 | 0.328 | 0.508 | 0.508 | 0.691 | 1.048 | 0.239 |

Fonte: Próprio autor.

Tabela 5 – Resultados dos modelos Logit

| Variável | Modelo 1 | Modelo 2 | Modelo 3 | Modelo 4 | Modelo 5 |
|---------------------------|----------|---------------|---------------|-----------|---------------|
| <i>Grade</i> | | | | | |
| A (Omitida) | | | | | |
| B | 0.847*** | 0.966*** | 0.946*** | 0.812*** | 0.922*** |
| C | 1.411*** | 1.650*** | 1.610*** | 1.346*** | 1.570*** |
| D | 1.825*** | 2.205*** | 2.159*** | 1.747*** | 2.113*** |
| E | 2.266*** | 2.788*** | 2.729*** | 2.165*** | 2.672*** |
| F | 2.615*** | 3.276*** | 3.214*** | 2.493*** | 3.144*** |
| G | 2.827*** | 3.589*** | 3.528*** | 2.702*** | 3.457*** |
| Annual_inc | | -2.934e-06*** | -3.023e-06*** | | -2.477e-06*** |
| Dti | | 0.012*** | 0.012*** | | 0.012*** |
| Emp_Length | | -0.011*** | -0.013*** | -0.007*** | -0.006*** |
| Installment | | 7.045e-04*** | | | |
| Int_rate | | -4.809*** | -4.409*** | | -4.455*** |
| Loan_amnt | | | 1.705e-05*** | | 1.591e-05*** |
| Mths_since_last_delinq | | -8.423e-04*** | -9.116e-04*** | | -0.001*** |
| Revol_bal | | -1.251e-05*** | -1.271e-05*** | | -1.119e-05*** |
| Revol_util | | 0.086*** | 0.103*** | | 0.104*** |
| <i>Home_Ownership</i> | | | | | |
| Mortgage (omitida) | | | | | |
| Own | | | | 0.13*** | 0.118*** |
| Rent | | | | 0.279*** | 0.268*** |
| <i>Term</i> | | | | | |
| 36 months (omitida) | | | | | |
| 60 months | | | | 0.117*** | 0.069*** |
| <i>Purpose</i> | | | | | |
| Credit_card (omitida) | | | | | |
| Debt_consolidation | | | | 0.078*** | 0.060*** |
| Home_improvement | | | | 0.015 | 0.038** |
| House | | | | -0.182*** | -0.147*** |
| Major_purchase | | | | -0.026 | 0.008 |
| Medical | | | | -0.046. | 0.022 |
| Moving | | | | -0.027 | 0.060. |
| Other | | | | -0.143*** | 0.085*** |
| Renewable_energy | | | | 0.143 | 0.213* |
| Small_business | | | | 0.237*** | 0.282*** |
| Vacation | | | | -0.127*** | -0.029 |
| Wedding | | | | -0.315** | -0.240. |
| Nagelkerke R ² | 0.084 | 0.093 | 0.093 | 0.089 | 0.096 |
| Mcfadden R ² | 0.0592 | 0.0649 | 0.0647 | 0.0626 | 0.0671 |
| AURoc | 0.669 | 0.681 | 0.681 | 0.678 | 0.684 |

*** Significante a 1%, ** significante a 1%, * significante a 5%, . significante a 10%

Fonte: Próprio autor.