Funding Decision in Online Marketplace Lending

F.Y. Eric Lam

Hong Kong Institute for Monetary and Financial Research
Hong Kong Academy of Finance
Hong Kong Monetaty Authority

ABFER-BIS-CEPR workshop on
"Fintech and digital currencies"
Sep 26, 2019

Disclaimer

The views expressed in this paper are those of the authors, and do not necessarily reflect those of the Hong Kong Monetary Authority, Hong Kong Academy of Finance, Hong Kong Institute for Monetary and Financial Research, its Council of Advisers, or the Board of Directors.

Online marketplace lending

Fintech using internet and web/mobile apps to match individual borrowers and lenders

- Peer-to-peer unsecured personal loans
- No geographical, time, and licensing restrictions
- Enormous volume of participants $\rightarrow$ emerging debt market (Havrylchyk, Mariotto, Rahim, Verdier (2019))
- Implications for investment opportunities and financial inclusion, etc.

Research objectives

Induce predictive representations of funding decision (listing outcome, funded amount and credit pricing) from 28MM+ LendingClub loan listings (Jan2014–Dec2018) using the simple decision tree algorithm

1. Loan application features and funding decision
2. Lending preference and decision of representative investor
3. Impact of monetary policy regime on funding decision

Related studies I

Our novetly: examine rejected listings together with funded ones to analyze funding decision in online marketplace lending

- Voluntary disclosure and cost of borrowing (Michels (2012))
- Credit risk and loan performance (Emekter, Tu, Jirasakuldech, Lu (2015))
- Screening and adverse selection (Vallee, Zeng (2018))
- Borrower's maturity choice and private information (Hertzberg, Liberman, Paravisin (2018))

Related studies II

- Quantitative easing and lender risk taking (Chu, Deng (2018))
- Fintech lenders tend not to target borrowers with traditional credit constraints (Di Maggio, Yao (2018))
- LendingClub lenders provide credit to areas that tend to be underserved by traditional banks and areas with adverse economic condition (Jagtiani, Lemieux (2018))
- Fintech credit around the world and policy implications (Claessens, Frost, Turner, Zhu (2018))
- Assessing credibility with alternative data improves financial inclusion and reduces price of credit (Jagtiani, Lemieux (2019))

LendingClub

- US online peer-to-peer lending founded in 2006
- Listed on the NYSE (LC); Russell 2000 constituent
- Largest quarterly loan origination among personal-focused lenders in US between Q2 20013 and Q2 2018 (Dixit (2018))
- Unsecured personal loans US$1,000–40,000; 36 months or 60 months maturity
- Loans listed on www.lendingclub.com (See Jagtiani, Lemieux (2019) for loan application process summary)
- Revenue from service fees on lenders and origination fees on borrowers

Raw data files

Sequence of files on 2,260,701 funded loan listings

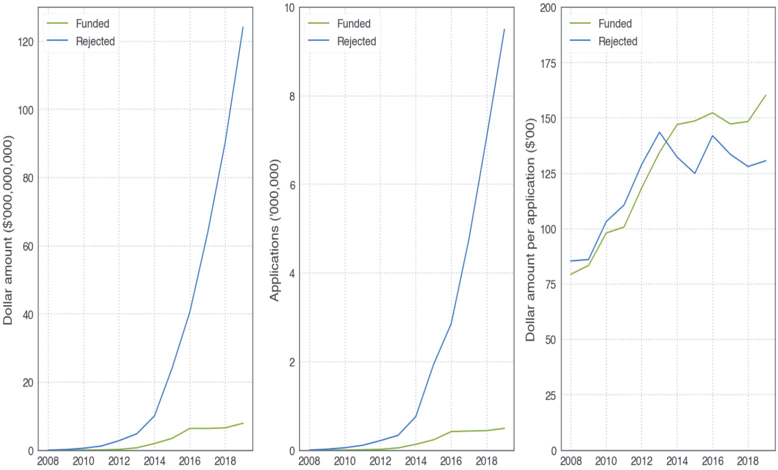
- Total funding of US$34 billions
- June 2007 – December of 2018 (monthly freq)

Sequence of files on 27,648,741 rejected loan listings

- Total request of US$363 billions
- 26 May 2007 – end of 2018 (daily freq)

Merge by the month of a year

Listing volume over time



Sample of loan listings

- Jan2014 – Dec2018
- 28,162,260 loan listings (94% of the listings in the raw files)
- Loan titles of rejected listings were not standardized prior to 2014; impossible to properly idenfity loan purposes
    - 60,326 individual titles, 15,241 missing titles for listings in and before 2012; 15,536 individual titles for listings in 2013
- 27 common titles and 640 individual ones for listings in 2014; titles are more or less standardized since 2015
- No winsorization and other exclusions

Requested amount

- 0.01% of the loan listings request $[\$0, \$1,000)$; 0.69% of the listings request $> \$40,000$

1. Encode ordinal dummy variable: requested amount is acceptable ($[\$1,000, \$40,000]$) vs. otherwise
2. Requested amount as a numerical feature; proxy for investment profitability and/or risk to lenders

Borrowing purpose

- Standard text preprocessing on listing titles or purposes
  - Lowercasing, removal of underscores and beginning/ending whitespaces, and handling misspellings, inflected words, derived word forms, non-canonical word forms, and special characters
- Encode into six ordinal categories
  - Unclassified, consolidate existing debt, consumption expenditure (nondurable goods and service), capital investment (durable goods and service), education, and business
  - Encoding of $0, 1, \ldots, 5$ to proxy for productivity of proposed funding

Debt-to-income ratio

- Monthly debt payments on total debt obligations, excluding mortgage and the loan currently requested via LendClub, divided by self-reported monthly income

1. Encode ordinal feature; proxy for leverage level and uncertatinty to lenders
   - 0 when $[0, 0.4)$; upper bound is a healthy level suggested by LendingClub
   - 1 when $[0.4, 1.0)$: high existing leverage
   - 2 when $[1.0, \infty)$: already insolvency
   - 3 for invalid or missing record
2. Debt-to-income ratio as a numerical feature (invalid records are replaced by -1)

Employment length

- Categorical variable that has 12 levels: unknown, $<$ 1 year, 1 year, 2 years, ..., 9 years, $>$ 10 years
- Follow this ordering to create ordinal feature to proxy for borrower's income stability

Spatial information

- 3-digit zip codes and states of the addresses of the borrowers

1. Four categorical dummy features based on special zip codes
   - US military
   - Government at Washington DC
   - Internal Revenue Service (IRS)
   - Parcel Return Service (PRS) by United Parcel Service (UPS)
2. Two ordinal dummy features based on zip codes and states; proxy for geographic uncertainty
   - Zip code provided is not in use in US
   - Zip code provided does not match state provided
3. 3-digit zip code as a nominal categorical feature: 0, 1, . . . , 999, and 1000 to encode missing zip code; proxy for dispersion in economic condition

Temporal information

- Monthly periodicity and quarterly periodicity; possible for funding decision to have seasonal components
- Loan issuance month or quarter for funded listings
- Application month or quarter for rejected listings; assume decision made in the same month or quarter of application

Loan grade

- LendingClub assigns loan grade to the listings, but records only available on funded listings; credit rating not available on any listing
- LendingClub-assigned loan subgrade has 35 levels: A1, A2, A3, A4, A5, B1, B2, B3, B4, B5, . . . , G5
  - Follow this ordering to create ordinal feature to proxy for credit risk

Listing outcome, funding amount and credit pricing

- 7.21% of loan listings and 8.24% of requested amount are funded
    - 10.86% and 12.66% in 2014 vs. 4.956% and 6.01% in 2018
- Credit pricing on funded listing: interest rate of loan minus risk-free rate
    - Daily yields on 3-year and 5-year US Treasury bonds (Bloomberg) are aggregated into monthly averages, then matched to loans by issue date (month/year) and maturity
    - Credit risk premium: 0.025–0.301 p.a., average $= 0.114$, standard deviation $= 0.049$; high premium reflects unsecured nature

Simple decision tree I

- Nonparametric; flexible function mapping: low bias or reducible modeling error
- Work for classification (funding outcome) and regression (funding amount, credit pricing), practical
- Recursively partition data (binary) to produce nonlinear tree structure resembling sequential decision making process (hierarchical interactions)
  - Tractable (visualize, interpret)

Simple decision tree II

- Require many observations and we have
- # observations $\gg$ # listing features; little concern for curse of dimensionality
- Provide means to examine relative importance of listing features
- Khandani, Kim, Lo (2010) use machine learning algorithm to predict consumer default risk in credit card, achieving high out-of-sample classification accuracy

Workflow

Per prediction problem

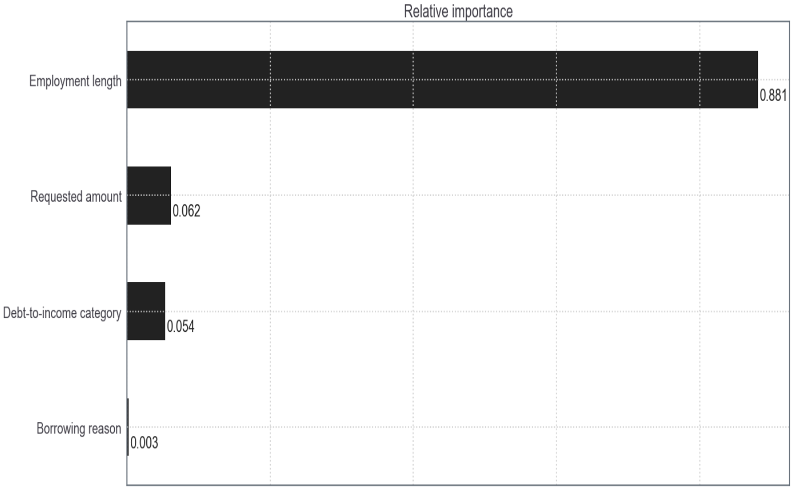
- 70% train-test split; induction learned from train set
- Grid search over 288 specifications using 5 fold cross validation
  - Select listing feature subset $(2 \times 2 \times 4 \times 3)$
  - Tune hyperparameters (2×3)
  - Try to get low variance or reducible out-of-sample error
- In unseen test set, examine out-of-sample performance and relative feature importance of optimal predictive representation

Results I

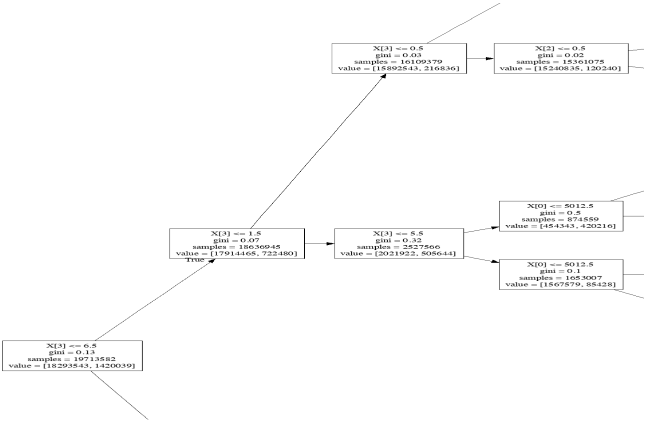
Predicting funding outcome

- Requested amount, borrowing purpose, categorical debt-to-income ratio, employment length
- 95.47% cross validation accuracy
- 95.46% out-of-sample accuracy
- "Always not getting funded" classifer: 92.79% accuracy

Results II


Relative importance
Employment length
Requested amount
Debt-to-income category
Borrowing reason
0.881
0.062
0.054
0.003

Results III


X[3] <= 6.5
gini = 0.13
samples = 19713582
value = [18293543, 1420039]
X[3] <= 1.5
gini = 0.07
samples = 18636945
value = [17914465, 722480]
True
X[3] <= 0.5
gini = 0.03
samples = 16109379
value = [15892543, 216836]
X[2] <= 0.5
gini = 0.02
samples = 15361075
value = [15240835, 120240]
X[3] <= 5.5
gini = 0.32
samples = 2527566
value = [2021922, 505644]
X[0] <= 5012.5
gini = 0.5
samples = 874559
value = [454343, 420216]
X[0] <= 5012.5
gini = 0.1
samples = 1653007
value = [1567579, 85428]

Results IV

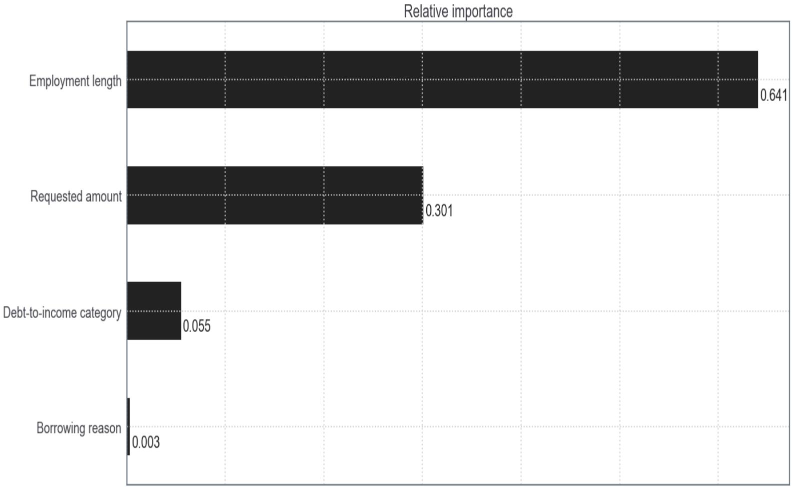
- E.g., for borrower with employment history > 5 years (5.78%), funding chance is 64.79%
  - Low existing leverage boots likelihood to 67.21%
  - Otherwise, drops to 13.80%
- Borrower with unknown employment length or employment length of 5 years or less: 3.88%
- Borrowing purpose does not matter much except when borrower has employment length > 6 years but existing leverge is not low

Results V

Predicting funding amount

- Same features
- 50.64% cross validation adjusted R^2
- 50.54% out-of-sample adjusted R^2
- Requested amount stated in a loan listing is predetermined by the borrower using information not thoroughly disclosed; lenders cannot negotiate and face a binary decision

Results VI

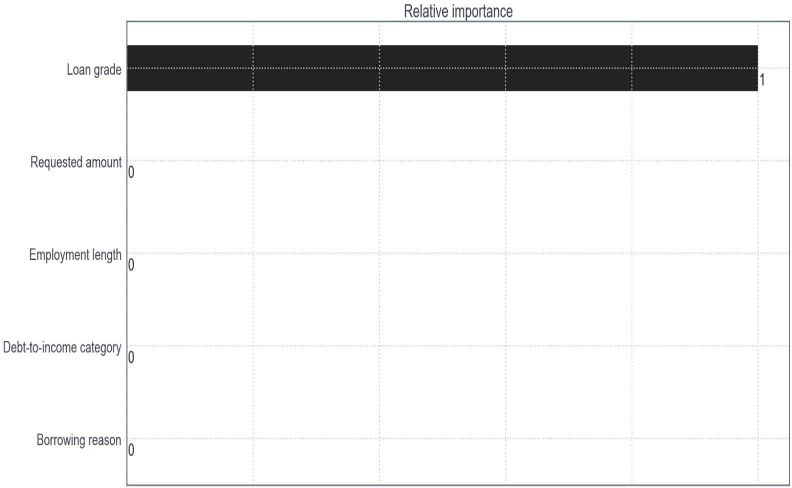

Relative importance
Employment length
0.641
Requested amount
0.301
Debt-to-income category
0.055
Borrowing reason
0.003

Results VII

Predicting credit pricing

- Same features
- 4.67% cross validation adjusted R^2
- 4.62% out-of-sample adjusted R^2
- Adding LendingClub loan grade
- 96.51% cross validation adjusted R^2
- 96.51% out-of-sample adjusted R^2

Results VIII



Nonconventional monetary policy regime dummy as an additional feature

- Time series dummy is one prior to the first Fed rate hike at 17 Dec 2015 and zero otherwise
- Predictive representations and relative feature importance remain almost the same
- Dummy has relative importance 0%–1.5%

Repeat analyses on subsamples splitted by calendar year or by month of year

- Results are largely similar; findings are robust

Conclusion

- 28mm+ recent loan listings on LendingClub, one of the world's largest online marketplace
- Robust predictive representations of funding decision using tree-based machine learning
- Borrower's employment length is the key factor in the preference of the lenders particiapting in this Fintech peer-to-peer lending platform
- Requested amount and existing leverage are secondary
- Credit pricing chraged on a funded listing fully depends on LendingClub loan grade
- Monetary policy regime has little impact on funding decision in this market