Funding Decision in Online Marketplace Lending *

F.Y. Eric Lam
Hong Kong Institute for Monetary and Financial Research
Hong Kong Academy of Finance
Units 1005–1011, 10/F, One Pacific Place
88 Queensway, Hong Kong
Tel: (852) 2597-0802
Email: cfylam@hkma.gov.hk

September 15, 2019

Abstract

This study analyzes more than 28 million recent loan listings on LendingClub, one of the world's largest online marketplace lending platform. Using tree-based machine learning, we develop robust predictive representations of funding decision in this Fintech peer-to-peer lending plaftorm. We find that a borrower's employment length is the main factor in the preference of lenders making funding decision. Requested amount and the existing leverage of a borrower are secondary in lenders' preference. The credit pricing charged on a funded listing fully depends on the loan grade assigned by LendingClub. Monetary policy regime has little impact on funding decision in this platform.

Keywords: Financial Technology, Fintech Lending, P2P Lending, Peer-to-peer Lending

---

* 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. All errors are mine.

1 Introduction

Online marketplace lending is a financial technology that utilizes the internet and web and mobile applications as a platform to match individual borrowers and lenders. Such online peer-to-peer platform typically focuses on personal loans that are unsecured, thus the proposed and funded loan amounts per loan application are not large. Yet, the enormous volume of participants and loan listings has led this lending channel into an emerging debt market (see, e.g., Havrylchyk, Mariotto, Rahim, and Verdier (2019) [7]). This digital credit channel does not have geographical, time and licensing restrictions on its participating lenders and borrowers, at least within the jurisdiction the online platform operates. Not only non-institutional lenders can expand their investment opportunities, individuals who might find it difficult to participate in traditional loan markets can also enhance their funding capabilities.

This study analyzes more than 28 million recent loan listings, from January 2014 to December 2018, on LendingClub, the world's largest online consumer lending platform operating in the US.[1] The raw data sets contain listings that are funded and those that are not. We use the simple decision tree algorithm to induce predictive representations of funding decision (listing outcome, funded amount and credit pricing) from the loan listings.[2] This machine learning algorithm is easy to visualize hence to interpretable. There are three research objectives. We first explore the association between loan application features and funding decision. The lending preference and decision of a representative investor is then investigated. Finally, we examine the impact of monetary policy regime on funding decision.

This paper is related to several other studies on online marketplace lending. Michels (2012) [12] analyzes data from Prosper.com, a peer-to-peer lending website, and shows that voluntary disclosure reduces cost of borrowing. Emekter, Tu, Jirasakuldech, and Lu (2015) [6] investigate peer-to-peer credit risk and subsequent loan performances. Vallee and Zeng (2018) [13] model screening and adverse selection in a marketplace lending platform and use data from LendingClub to test their predictions. Hertzberg, Liberman, and Paravisin (2018) [8] examine the private information revealed from screening LendingClub borrowers by their maturity choices. Chu and Deng (2019) [2] argue that quantitative easing encourages investors fund riskier peer-to-peer loans using the Prosper.com data from 2007 to 2013. On the one hand, Di Maggio and Yao (2018) [4] show that there is little evidence on Fintech lenders targeting borrowers who are credit constrained by traditional banks. On the other hand, Jagtiani and Lemieux (2018) [9] show that LendingClub's consumer lending provides credit to areas that tend to be underserved by traditional banks and areas with adverse economic condition. Claessens, Frost, Turner, and Zhu (2018) [3] study Fintech credit around the world and discuss policy implications of the technology. More recently, Jagtiani and Lemieux (2019) [10] find that LendingClub uses alternative data to evaluate borrower credibility and this improves financial inclusion and reduces the price of credit. The novelty of our study is that it examines rejected listings together with funded ones to examine funding decision in online marketplace lending.

[1] According to Dixit (2018) [5], LendingClub has the largest quarterly loan origination among personal-focused lenders in US between Q2 20013 and Q2 2018.

[2] Khandani, Kim, and Lo (2010) [11] use machine learning algorithm to predict consumer default risk in credit-card-holder delinquencies and the technique achieves high out-of-sample classification accuracy.

The rest of the paper is organized as follows. Section 2 provides the background of LendingClub and describes the raw data files. Section 3 presents exploratory data analysis on the loan listing features constructed from the raw data. Section 4 presents our machine learning workflow and reports the predictive representations and auxiliary results induced from the decision tree algorithm. Section 5 investigates the impact of monetary policy regime on funding decision. Section 6 examines model dynamics and the robustness of the findings. Section 7 concludes the study.

2 Sample of peer-to-peer loan listings

2.1 LendingClub

We analyze peer-to-peer loan listings from LendingClub, a US firm founded in 2006. The firm operates one of the world's largest online peer-to-peer lending platform. It is listed on the New York Stock Exchange (NYSE: LC) and is a constituent stock of the Russell 2000 Index. The platform allows borrowers to request unsecured personal loans between US$1,000 and US$40,000. The loan period is either 36 months or 60 months. Lenders inquire about the listed loans on the LendingClub website (www.lendingclub.com) and choose loans that they want to invest according to the background information of the borrowers, the requested amounts, and the borrowing purposes, etc. See Jagtiani and Lemieux (2019) [10] for a summary of the loan application process. LendingClub generates revenue by charging service fees on lenders and origination fees on borrowers. Lenders receive interest payments as investment returns.

2.2 Raw data files

LendingClub provides information on loan listings that are funded and loan listings that are rejected in two sequences of files. The first sequence is data files on 2,260,701 funded listings from June 2007 to December of 2018, involving a total of US$34 billions. The second sequence is data files containing 27,648,741 rejected listings from 26 May 2007 to the end of 2018, requesting a total of US$363 billions. The funded listings and the rejected listings are merged by the month of a year.

Figure 1 shows the dollar amount funded and the dollar amount rejected together with the number of funded loan listings and the number of rejected listings over time. It also displays the time series of the dollar amount per funded listing and that of the dollar amount per rejected listing. The volume of loan listings, in terms of either the sum of the dollar amount funded and the dollar amount rejected or the sum of the number of funded listings and the number of rejected listings, has dramatically increased over the past decade. On the one hand, this upsurge is mostly driven by the increase in the number of listings, especially the enormous increase in the number of rejected ones. This suggests that a fast growing number of borrowers has turned to this channel to seek financing but it is less likely for a listing to be funded over time. On the other hand, the relatively modest increase in the dollar amount per listing plays a much lesser role in the upsurge of listing volume.

2.3 Sample data

The data used in our empirical analysis starts from Jan 2014 and ends in December 2018. Our sample begins from Jan 2014 because the loan titles of rejected listings were not standardized prior to 2014. For example, there are 60,326 individual titles and 15,241 missing titles for listings in and before 2012. There are 15,536 individual titles for listings in 2013. Due to high idiosyncratic noise in the titles, it is impossible to properly identify the borrowing purposes for these listings. The loan titles for listings in 2014 are much less erratic with 27 common titles and 640 individual ones. The titles have become more or less standardized since 2015. Our sample contains 28,162,260 loan listings, which makes up 94% of the listings in the raw data files.

3 Exploratory data analysis

To learn the predictive representations of listing outcome (fund versus reject), funding amount, and credit pricing for funded listings, we construct features from the intersection of the data columns in the two sequences of data files.[3]

3.1 Requested amount

LendingClub adverstises loans between \$1,000 and \$40,000. In our sample, 0.01% of the loan listings request \$0 to less than \$1,000. There are also listings requesting more than \$40,000 (0.69% of the listings). We create ordinal dummy variable from `loan_amnt` in the files containing funded listings and from `Amount Requested` in the files containing rejected listings. The dummy is encoded to one when the requested amount is acceptable or in the interval $[\$1,000, \$40,000]$ and to zero otherwise. We also use the requested amount itself as a numerical feature. This feature can proxy for investment profitability and/or risk to lenders.

3.2 Borrowing purpose

We perform standard text preprocessing on `purpose` in the files containing funded listings and on `loan titles` in the files containing rejected listings [4]. The loan purposes or titles are categorized into six borrowing purposes [5]. These ordinal categories are unclassified, consolidate existing debt, consumption expenditure (nondurable goods and service), capital investment (durable goods and service), education, and business. We then create an ordinal feature following this order with the encoding of 0, 1, . . . , 5 to proxy for the productivity of the proposed funding.

[3]For a funded listing, funding amount equals requested amount in the data. Funding amount is set to zero for a rejected listing.

[4]The strings are first lowercased, followed by removal of underscores and beginning and ending whitespaces. For non-standard titles in 2014 and 2015, we manually handle misspellings, inflected words, derived word forms, non-canonical word forms, and special characters.

[5]The encoding takes into account the temporal changes in loan titles that can be selected from the drop down menu on LendingClub's website.

3.3 Debt-to-income ratio

This ratio is defined as the monthly debt payments on total debt obligations, excluding mortgage and the loan currently requested via LendClub, divided by self-reported monthly income. We create an ordinal feature from `dti` in the files containing funded listings and `Debt-To-Income Ratio` in the files containing rejected listings. The encoding is 0 when the ratio in the range $[0, 0.4)$, where the upper bound is a healthy level of leverage recommended by LendingClub. The encoding is 1 when the existing leverage of the borrower is considered to be high (ratio $\in [0.4, 1.0]$), is 2 when the borrower is already insolvent (ratio $\in (1.0, \infty)$), and is 3 when the record is invalid, as indicated by -1 or 999, or when the record is missing. Higher encoded value can imply higher leverage level and uncertatinty to lenders. Alternatively, we use the debt-to-income ratio itself as a numerical feature (invalid records are replaced by -1).

3.4 Employment length

The variable `emp_length` in the files containing funded listings and the variable `Employment Length` in the files containing rejected listings are categorical. This variable has 12 levels ranging from unknown, less than 1 year, 1 year, 2 years, ..., 9 years, to more than 10 years. We follow this ordering to create an ordinal feature with the encoding of 0, 1, ..., 11 to proxy for borrower's income stability.

3.5 Spatial information

The data files provide the 3-digit zip codes and states of the addresses of the borrowers. Based on `zip_code` and `addr_state` in the files containing funded listings and `Zip Code` and `State` in the files containing rejected listings, we create four categorical dummy features and two ordinal dummy features. The first four spatial features indicate whether a borrower's zip code is associated with the US military, whether the zip code is associated with the government at Washington DC, whether the zip code is associated with the Internal Revenue Service (IRS), and whether the zip code is associated with the Parcel Return Service (PRS) by the United Parcel Service (UPS). These features indicate whether a zip code is special. The other two spatial features indicate whether the zip code provided by the borrower is actually not in use in the US and whether the zip code provided does not match the state provided.[6] These features can proxy for geographic uncertainty. Alternatively, we use the 3-digit zip code itself as a nominal categorical feature to proxy for dispersion in economic condition. This feature has the encoding of 0, 1, ..., 1000. The value 1000 encodes missing zip code. We do not employ one hot encoding due to high cardinality.

3.6 Temporal information

Funding decision might be seasonal. Therefore, we create periodicity features using the month or the quarter a listing is funded according to the loan issuance date. For rejected listing, we use the month or the quarter of the loan application. This assumes that the decision on a rejected listing is made within the month or the quarter of the application.

[6] The supplementary information for analyzing zip codes is obtained from en.wikipedia.org/wiki/List_of_ZIP_Code_prefixes.

The nominal encoding for monthly periodicity is 0, 1, ..., 11 while the nominal encoding for quarterly periodicity is 0, 1, 2, 3.

3.7 Loan grade

Listing outcome, funding amount and credit pricing might depend on credit assessment (e.g., the FICO score), which reflects information about a borrower's track record, financial condition and creditbility. However, credit rating is not available from the data files. Although loan grades assigned by LendingClub are available in the files containing funded listings, they are not available in the files containing rejected ones. Consequently, we are not able to utilize loan grade as a feature for predict listing outcome and funding amount. We are able to use loan grade as an additional feature in predicting credit pricing. In this analyisis, we assume loan grades are known to investors before the credit pricing on funded listing is determined.

Loan grade is available as the LendingClub-assigned loan grade (`grade`) and as the LendingClub-assigned loan subgrade (`sub_grade`). The categorical loan grade is an alphabetical character ranging from A to G while the finer loan subgrade is a string taking 35 possible values A1, A2, A3, A4, A5, B1, B2, B3, B4, B5, ..., G5. We use the finer subgrade, which presumably is more informative. We follow the given ordering to create an ordinal feature with the encoding of 0, 1, 2, ..., 34 to proxy for credit risk.

3.8 Listing outcome, funding amount and credit pricing

In our sample, 7.21% of loan listings and 8.24% of requested amount are funded. These proportions have dropped from 10.86% and 12.66% in 2014 to 4.956% and 6.01% in 2018. Such reductions are consistent with the patterns in Figure 1. We measure the credit pricing on a funded listing as the difference between the interest rate charged on the loan (`int_rate` in the files containing funded listings) and the risk-free rate proxied by the US Treasury bond yield. Daily yields on 3-year and 5-year US Treasury bonds are obtained from Bloomberg. In view of the loan issue dates (`issue_d`) being stated in month/year, the daily bond yields are aggregated into monthly averages. The monthly risk-free rates are then matched to interest rates of the funded loans by loan issue date and maturity. The resulting credit risk premium ranges from 0.025 to 0.301 p.a., with an average of 0.114 and a standard deviation of 0.049. The premium are rather high and these reflect the unsecured nature of the loans.

3.9 Bivariate analytics

We group the sample of loan listings by a listing feature and examine funding decision of the groups. Table 1 presents the number of listings, the requested amount, the proportion funded in terms of number of listings or requested amount, and the average credit pricing of funded listings. Panel A shows that the requested amount must be in the advertised interval ($[\$1,000, \$40,000]$) for the listing to have a chance of receiving funding. 0.65% of the listings and 5.62% of the requested amount are not funded due to violation of this restriction. For listings with acceptable requested amounts, the correlation between the numerical requested amount and listing outcome is 0.068. This somewhat indicates that requested amount tends to proxy for investment profitability rather than risk to lenders.

Panel B shows that 79.03% of the listings do not present clear motivation or are associated with debt consolidation. Listings for debt consolidation have the highest funding rate, in both number of listings and requested amount, followed by listings with unclassified purpose. Listings for other reasons have lower funding rate, with listings for education being the least likely to be funded. The average credit pricing is higher for loans funded for education or business. The pricing for other purposes are somewhat similar. It seems that listings with more productive intentions are not favored more.

Panel C shows that low leverage borrowers are much more likely to receive funding. Borrowers with unfavorable leverage conditions and those with invalid or missing leverage information still can but rarely be funded. For listings with valid debt-to-income ratios presented, the correlation between the numerical ratio and listing outcome is -0.009. This suggest that the relation between existing leverage and listing outcome can be nonlinear. The average credit pricing for borrowers with high leverage is about 0.017 p.a. higher than that for borrowers with low leverage. It seems that existing leverage condition plays a rather limited role in pricing.

Panel D shows that while some borrowers with unknown employment length receive funding, borrowers with employment length less than 1 year rarely get funded. Beyond these, funding rate is strictly increasing in employment length of borrower. It seems that employment length plays a an important role in determining listing outcome and funding amount. Borrowers with stabler income are more preferred by lenders. However, this listing feature is a lot less important for credit pricing.

Panel E shows that borrower's location might matter for funding decision. On the one hand, listings with zip codes associated the PRS have slightly higher funding rate than otherwise. On the other hand, listings with zip codes associated with the Government or the IRS have lower funding rate than otherwise. Listings which do not have valid zip codes or have zip codes that conflicts with the states provided have lower funding rate than otherwise. It seems that listings with high geographic uncertainty are less favored by lenders.

The grouping using the nominal zip code feature (Figure 2 and Figure 3) reveals that listing volume and funding rate vary significantly across location. New York (Main 3), NY has the highest funding rate by number of listings (15.43%) and this location has 1,244 listings requesting $28.14 millions. Washington (Parcel Return), DC has the highest funding rate by requested amount (18.86%) but this location only has 17 listings requesting $0.28 millions. Other than zip codes associated with the IRS, the unincorporated territory San Juan (West), PR, and a few invalid zip codes, Fort Dodge, IA, Creston, IA, Carroll, IA, and Burlington, IA have the lowest funding rate by number of listings (0.00%) as well as by requested amount (0.00%). These four locations have 351 listings requesting $5.40 millions. Figure 4 shows that there is some variation in the average credit pricing across location. There are occassional spikes in pricing. Fresno, CA (IRS), Des Moines, IA (N-Z), Washington, DC (Government 4) and two unknown locations with zip codes not in use (862 and 929) are the top five places with average credit risk premium being 0.172 p.a. or higher.

Figure 5 shows that funding rate has dropped substantially since the end of 2015. Although the volume of loan listings has grown enormously (see Figure 1), the declining funding rate has kept the growth of funded listings at a comparatively modest rate. The credit pricing has also decreased. This might be due to improvement in quality of funded listings or simply due to the hikes of US Treasury yields during monetary normalization. Panel F and Panel G of Table 1 show that funding rate might be higher at the beginning of a year. Yet, the seasonality in funding decision seems to be rather weak.

4 Machine learning workflow and results

This section reports the findings from our main analysis. We induce predictive representations of funding decision from the large sample of loan lisings. Since the bivariate analysis presented in the previous section shows that the association between loan listing feature and financing decision can be nonlinear, we employ the simple decision tree algorithm (see, e.g., Breiman, Friedman, Stone, and Olshen (1984)). This nonparametric learning algorithm can tackle both classification problem (predicting listing outcome) and regression problem (predicting funding amount and credit pricing). This algorithm recursively partitions data. One of the main output from this algorithm is a nonlinear tree structure that resembles a sequential decision making process and is easy to visualize hence to interpret. This algorithm does not impose strong restriction on the form of the function mapping listing features to funding decision. Such flexibility enables low bias or low reducible modeling error even in handling complex problems. One drawback of this method is that it requires large amount of data. As we have a large sample of listings and there are much more observations than listing features, the drawback is of little concern for us, so is the curse of dimensionality. Since this method is not affected by the scales of variables, the numeric listing features do not have to be standardized; the categorical features and the ordinal features do not have to be one hot encoded. This method also provides the means to examine the relative importance of the listing features in each of the prediction tasks.

For each prediction task, our workflow begins by randomly allocating 70% of the entire sample into a training set and the rest into a test set.[7] The test set is reserved for merely evaluating the performance of the learned predictive representation. The induction of the tree structure is performed on the training set. We also select subsets of listing features and tune hyperparameters of the learning algorithm along with the induction. For feature selection, we consider numerical versus categorical version of requested amount, numerical versus categorical version of the debt-to-income ratio, special zip dummy versus geographic uncertainty dummy versus all 3-digit zip codes versus exlusion of spatial informtion, and monthly periodicity versus quarterly periodicity versus exclusion of temporal information.[8] While keeping other listing featues intact, we end up considering 48 ($2 \times 2 \times 4 \times 3$) feature sets. For hyperparameter tuning, we consider two information gain criterion for splitting nodes (gini impurity versus entropy) and three maximal tree depth (three versus four versus five). We do not allow higher depth to avoid the algorithm from growing a tree that is excessively big. This can help prevent overfitting, i.e., learning too much of the idiosyncratic noises specific to the training set,

[7]A random seed is used to make our results reproducible.

[8]This allows for a more parsimony feature set that only includes borrower specific information.

and can ensure interpretability of the learned representation. The reason is the same for considering subset of features.

We then undertake grid search over the space containing the 288 specifications ($48 \times 2 \times 3$) using a five fold cross validation. Specifically, we compute an average cross validation score (accuracy for classification and adjusted R^2 for regression) for a specification and then select the specification that has the maximum score.[9] The average cross validation score is a selection measure based on out-of-sample evaluation.[10] The objective of cross validated specification selection is to achieve low variance or low reducible out-of-sample error.[11] The average score for a given specification is obtained as follows. We first randomly split the relevant part of the training set into five equal sized subsamples. We then remove one of the subsamples and train the algorithm according to the specification on the remaining observations. A score for the trained algorithm is computed based on the subsample being left out. After we iterate this process over the subsample being left out, we end up with five scores and we calculate the simple average. Finally, we take the predictive representation learned with the algorithm under the optimal specification to the unseen test set to examine out-of-sample prediction performance and relative feature importance.

4.1 Predicting funding outcome

The feature set containing numerical requested amount, borrowing purpose, categorical debt-to-income ratio, and employment length has the highest cross validation accuracy and is the most parsimony. Spatial features and temporal features are not informative and our hunch is that these features provide little information beyond borrower specific information. The decision tree has 95.47% cross validation accuracy. It achieves 95.46% out-of-sample accuracy in the unseen test set and this indicates that the predictive representation generalizes rather well. As a benchmark, the "always not getting funded" classifer would have an accuracy of 92.79% given the base proportion of listing funded is 7.21% in the full sample.

Figure 6 presents the relative importance of the four listing features based on the test set. Employment length, the proxy for borrower's income stability, is the major factor in funding outcome. Lenders tend to have very high preference on borrower's repayment capability. Requested amount and the debt-to-income ratio are secondary factors while borrowing purpose is almost trivial. The three main features account for more than 99% of the relative importance.

Figure 7 presents the decision tree. In this sequential decision making process, employment length is of the highest concern for lenders. For a borrower with an employment history longer than five years (5.78% of the loan listings), the chance of receiving funding

[9] In case of a tie, the specification with a more parsimonuous feature set is selected to reduce overfitting.

[10] Since the data is unbalanced (only 7% of the listings re funded), we also use the F1 measure as the cross validation score for classification. The F1 measures is the weighted harmonic average of precision and recall. Precision is the percentage of listings that are predicted to receive funding that are in fact funded. Recall is the percentage of funded listings that are predicted to receive funding. The results obatined using this measures are similar.

[11] One of the advantages in using cross validation is that we can evaluate specifications out-of-sample and simultaneously use all the data for training and testing.

is 64.79%. If a borrower has unknown employment length or employment length of five years or less, the likelihood of obtaining funding is only 3.88%. Lenders' preference is heavily biased towards borrowers with longer employment length. Furthermore, for a borrower with unknown employment length or employment length less than one year, the funding likelihood drops to 1.35%. Lenders then consider borrower's existing leverage, typically followed by requested amount. It seems that lenders do not consider borrow reason much. For a borrower with employment length of one to five years, the funding likelihood is 20.00%. Lenders then consider requested amount, typically followed borrower's existing leverage. Borrowing purpose does not matter at all. For a borrower with employment length more than five years, lenders consider the borrower's existing leverage. When the borrower has low leverage, chance of obtaining funding is 67.21%. Otherwise, the funding likelihood drops to 13.80%. Lenders then consider requested amount and, in some cases, borrowing reason.

4.2 Predicting funding amount

The same features in the previous subsection is selected. The decision tree has 50.64% cross validation adjusted R^2 and 50.54% out-of-sample adjusted R^2 in the test set. Prediction of numerical funding amount does not perform as well as prediction of binary funding outcome. A possible reason is that the requested amount stated in a loan listing is predetermined by the borrower and this amount is mostly based on information not thoroughly disclosed in the application. If lenders cannot negotiate on the requested amount, they essentially face a binary decision.

Figure 8 presents the relative importance of the four listing features based on the test set. Employment length is again the major factor in funding amount, but it is not as dominiating as in funding outcome. Requested amount becomes more important in this setting, probably for mechanical reason. The relative importance of the debt-to-income ratio is similar to that in the previous subsection and borrowing purpose remains trivial. The three main features account for more than 99% of the relative importance.

4.3 Predicting credit pricing

Again, the same features in the previous subsection is selected. The decision tree only has 4.67% cross validation adjusted R^2 and 4.62% out-of-sample adjusted R^2 in the test data. Figure 9 presents the relative importance of the four listing features based on the test set. The relative importance is quite different from those in the other two problems. Employment length is trivial, so is the debt-to-income ratio. The dominant factors are requested amount and borrowing purpose.

When loan subgrade is incorporated into the analysis, the results are much more encouraging. The decision tree has 96.51% cross validation adjusted R^2 and 96.51% out-of-sample adjusted R^2 in the unseen test data. Loan grade provides substantial information about credit pricing. Loan subgrade is included in further analysis of credit pricing. Figure 10 presents the relative importance of the five listing features based on the test set. Loan grade is the only factor in credit pricing . Other listing features are not important at all. LendingClub might use these features, together with additional information, in assigning loan grade (see, e.g., Jagtiani and Lemieux (2019)).

5 The impact of monetary policy regime

We introduce a nonconventional monetary policy dummy variable into as an additional feature. The time series dummy is one prior to the first Fed rate hike at 17 Dec 2015 and zero otherwise. This dummy variable itself can be associated with funding decision and can interact with the other features in the decision tree. The results indicate that monetary policy regime does not matter much. The three predictive representations remain almost the same. The decision trees with the monetary policy regime dummy achive almost the same accuracies in predicting funding outcome (95.49% cross validation accuracy and 95.49% out-of-sample accuracy). Figure 11 shows that the monetary policy regime dummy is extremely minor, with relative importance of merely 0.08%. In predicting funding amount, the decision tree has cross validation adjusted R^2 and 50.64% out-of-sample adjusted R^2. Figure 12 shows that the monetary policy regime dummy is almost not important. In predicting credit pricing, the decision tree has cross validation adjusted R^2 and 97.82% and 97.83% out-of-sample adjusted R^2. Figure 13 shows that loan grade remains as the dominant feature, with relative importance of 98.5%, but the monetary policy regime dummy again has a very tiny role, with relative importance of just 1.5%.

6 Model dynamics and robustness checks

We split the sample based calendar year, i.e. 2014, 2015, ..., 2018, and also split the sample based on the month of a year, i.e. January, Feburary, ..., December. We then repeat the analyses on each of these subsamples. The results are largely similar to those obtained from the full sample, suggesting that our findings are robust. In predicting funding outcome, the cross validation accuracies of the decision trees range from 93.62% to 96.52% for subsamples splitted by calendar year and the out-of-sample accuracies range from 93.63% to 95.96%. Employment length is the major factor in all years. Requested amount and the debt-to-income ratio seem to have slightly gained importance in 2018. In predicting funding amount, the cross validation adjusted R^2 range from 46.66% to 55.94% and the out-of-sample adjusted R^2 range from 46.46% to 56.18%. In all years, employment length, requested amount, and the debt-to-income ratio are the major factors while the debt-to-income ratio becomes slightly more important in 2018. In predicting credit pricing, the cross validation adjusted R^2 range from 97.64% to 99.66% and the out-of-sample adjusted R^2 range from 97.76% to 99.64%. Loan grade is the dominant factor in all years.

In predicting funding outcome, the cross validation accuracies of the decision trees range from 94.99% to 95.83% for subsamples splitted by the month of a year and the out-of-sample accuracies range from 93.63% to 95.96%. Employment length is the major factor in all months. Requested amount and the debt-to-income ratio again are the two other factors that matter. In predicting funding amount, the cross validation adjusted R^2 range from 47.30% to 53.30% and the out-of-sample adjusted R^2 range from 46.67% to 53.14%. In all months, employment length, requested amount, and the debt-to-income ratio are the factors that matter. In predicting credit pricing, the cross validation adjusted R^2 range from 96.17% to 97.20% and the out-of-sample adjusted R^2 range from 96.07% to 97.21%. Not surprisingly, loan grade is the dominant factor in all months.

7 Conclusion

This paper studies a large volume of recent loan listings on LendingClub, one of the world's largest online marketplace lending platform. Using the simple decision tree algorithm, we document robust predictive patterns of funding decision in this Fintech lending platform. A borrower's employment length, a proxy for income stability, is unambigiously the main element in lenders' preference. Lenders prefer to fund a borrower with longer employment length. Requested amount and the existing leverage of a borrower are secondary in lenders' preference. The loan grade assigned by LendingClub is the only determinant of the credit pricing charged on a funded loan listing. Furthermore, monetary policy regime has very little impact on the funding decision in this lending platform.

References

[1] Breiman, Leo, Jerome Friedman, Charles J. Stone, and R.A. Olshen, 1984. *Classification and Regression Trees*. Chapman & Hall/CRC.

[2] Chu, Yongqiang, and Xiaoying Deng, 2019. *Monetary Policy and Individual Investors' Risk-taking Behavior: Evidence from Peer-to-peer Lending*. Working Paper.

[3] Claessens, Stijn, Jon Frost, Grant Turner, and Feng Zhu, 2018. *Fintech Credit Markets around the World: Size, Drivers and Policy Issues*. BIS Quarterly Review, September 2018.

[4] Di Maggio, Marco, and Vincent W. Yao, 2018. *Fintech Borrowers: Lax-screening or Cream-skimming?* Working Paper.

[5] Dixit, Nimayi, 2018. *2018 US Digital Lending Market Report*. S&P Global Market Intelligence.

[6] Emekter, Riza, Yanbin Tu, Benjamas Jirasakuldech, and Min Lu, 2015. *Evaluating Credit Risk and Loan Performance in Online Peer-to-peer (P2P) Lending*. Applied Economics 47, 54–70.

[7] Havrylchyk, Olena, Carlotta Mariotto, Talal Rahim, and Marianne Verdier, 2019. *What Has Driven the Expansion of the Peer-to-peer Lending?* Working Paper.

[8] Hertzberg, Andrew, Andres Liberman, and Daniel Paravisin, 2018. *Screening on Loan Terms: Evidence from Maturity Choice in Consumer Credit*. Review of Financial Studies 31, 3532–3567.

[9] Jagtiani, Julapa, and Catharine Lemieux, 2018. *Do Fintech Lenders Penetrate Areas that Are Underserved by Traditional Banks?* Journal of Economics and Business 100, 43–54.

[10] Jagtiani, Julapa, and Catharine Lemieux, 2019. *The Roles of Alternative Data and Machine Learning in Fintech Lending: Evidence from the LendingClub Consumer Platform*. Federal Reserve Bank of Philadelphia Working Paper WP 18-15.

[11] Khandani, Amir E., Adlar J. Kim, and Andrew W. Lo, 2010. *Consumer Credit-risk Models via Machine-learning Algorithms*. Journal of Banking and Finance 34, 2767–2787.

[12] Michels, Jeremy, 2012. *Do Unverifiable Disclosure Matter? Evidence from Peer-to-peer Lending*. Accounting Review 87, 1385–1413.

[13] Vallee, Boris, and Yao Zeng, 2018. *Marketplace Lending: A New Banking Paradigm?* Review of Financial Studies 31, 1939–1982.

Table 1. Characteristics of groups sorted by a listing feature

Number of listings is in '000 and requested amount is in $'000,000. Figure with * is in actual value.

Panel A Requested amount	Too low or too high	Acceptable
Number of listings	182	27,980
Proportion funded	0.000	0.073
Requested amount	21,008	353,127
Proportion funded	0.000	0.087
Average credit pricing	N/A	0.114

Panel B Borrowing purpose	Unclassified	Consolidate existing debt	Consumption expenditure	Capital investment	Education	Business
Number of listings	9,338	12,920	4,098	1,329	99*	478
Proportion funded	0.064	0.089	0.052	0.040	0.020	0.042
Requested amount	119,387	185,209	54,607	7,487	1	7,443
Proportion funded	0.072	0.100	0.055	0.060	0.007	0.045
Average credit pricing	0.10	0.12	0.11	0.12	0.15	0.14

Panel C Debt-to-income ratio	Low leverage	High leverage	Insolvent	Invalid
Number of listings	21,194	3,661	2,134	1,173
Proportion funded	0.094	0.007	0.001	0.002
Requested amount	255,084	55,035	30,304	33,713
Proportion funded	0.119	0.009	0.002	0.001
Average credit pricing	0.114	0.131	0.123	0.119

Table 1 (cont'd)

Panel D Employment length (year)	Unknown	< 1	1	2	3	4	5	6	7	8	9	10+
Number of listings	1,070	21,941	368	347	312	223	2,361	144	123	136	111	103
Proportion funded	0.129	0.008	0.362	0.526	0.521	0.548	0.052	0.614	0.649	0.598	0.641	0.660
Requested amount	12,149	293,961	4,239	4,668	4,173	3,066	27,317	2,158	1,898	2,073	1,648	16,785
Proportion funded	0.138	0.009	0.448	0.570	0.571	0.59	0.067	0.662	0.643	0.608	0.669	0.659
Average credit pricing	0.115	0.114	0.115	0.114	0.114	0.114	0.114	0.114	0.115	0.115	0.114	0.113

Panel E	US military		Government		IRS		PRS		Invalid zip		Mismatch	
Borrower's location	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Number of listings	28,158	4	28,162	629*	28,162	576*	28,162	17*	28,158	4	28,057	105
Proportion funded	0.072	0.080	0.072	0.017	0.072	0.009	0.072	0.118	0.072	0.019	0.072	0.034
Requested amounted	374,070	65	374,125	10	374,126	9	374,135	0	374,074	61	372,437	1,698
Proportion funded	0.082	0.092	0.082	0.015	0.082	0.010	0.082	0.189	0.082	0.021	0.083	0.033
Average credit pricing	0.114	0.106	0.114	0.116	0.114	0.106	0.114	0.138	0.114	0.111	0.114	0.113

Panel F Decision month	1	2	3	4	5	6	7	8	9	10	11	12
Number of listings	1,966	1,671	2,078	2,122	2,74	2,318	2,651	2,637	2,459	2,745	2,659	2,582
Proportion funded	0.077	0.083	0.087	0.077	0.072	0.068	0.073	0.069	0.059	0.075	0.068	0.066
Requested amount	27,532	24,241	29,529	28,759	29,343	29,480	34,386	34,552	32,618	36,653	34,497	32,545
Proportion funded	0.085	0.089	0.093	0.086	0.084	0.081	0.084	0.079	0.069	0.085	0.080	0.079
Average credit pricing	0.115	0.115	0.113	0.114	0.113	0.113	0.117	0.116	0.116	0.113	0.111	0.110

Table 1 (cont'd)

Panel G Decision quarter	1	2	3	4
Number of listings	5,715	6,713	7,747	7,987
Proportion funded	0.082	0.072	0.067	0.070
Requested amount	81,303	87,581	101,556	103,695
Proportion funded	0.089	0.084	0.077	0.081
Average credit pricing	0.114	0.113	0.116	0.112

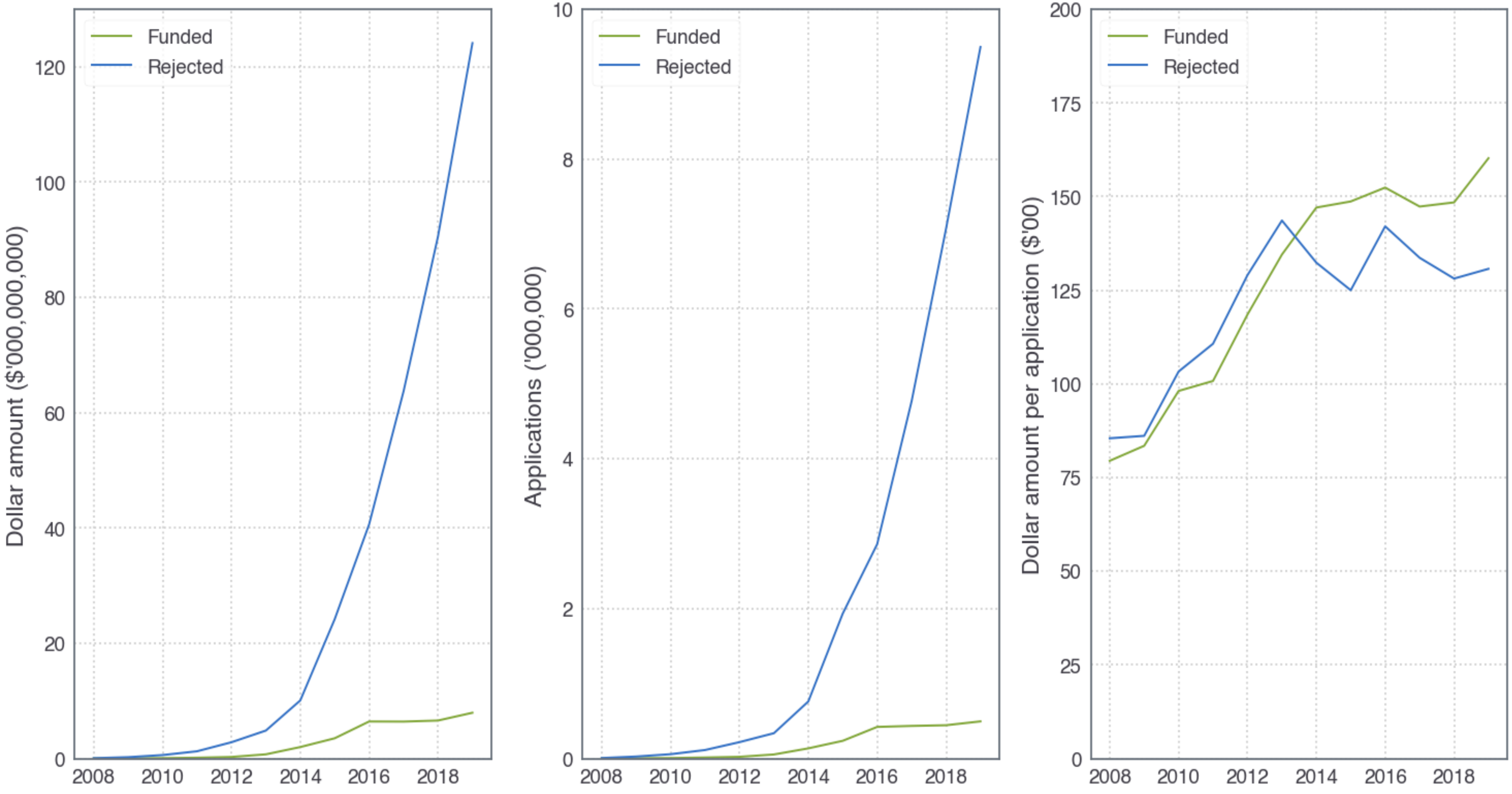

Figure 1: Listing volume over time

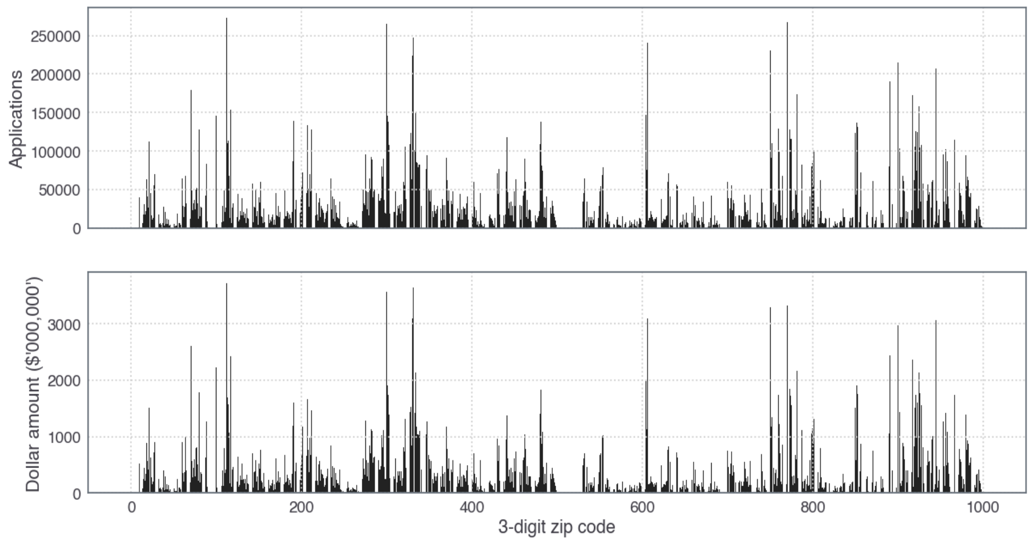

Figure 2: Listing volume across location

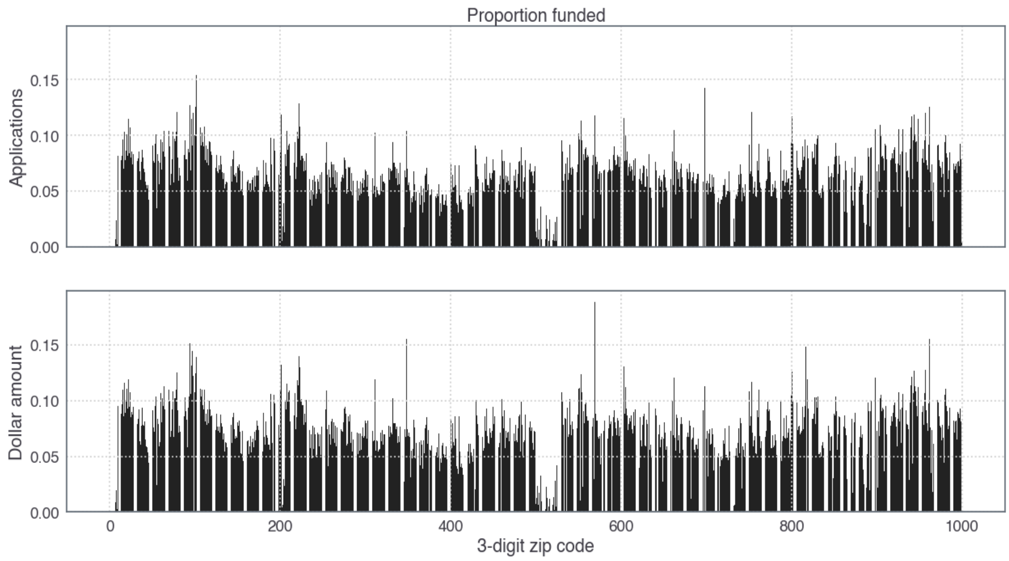

Figure 3: Funding rate across location

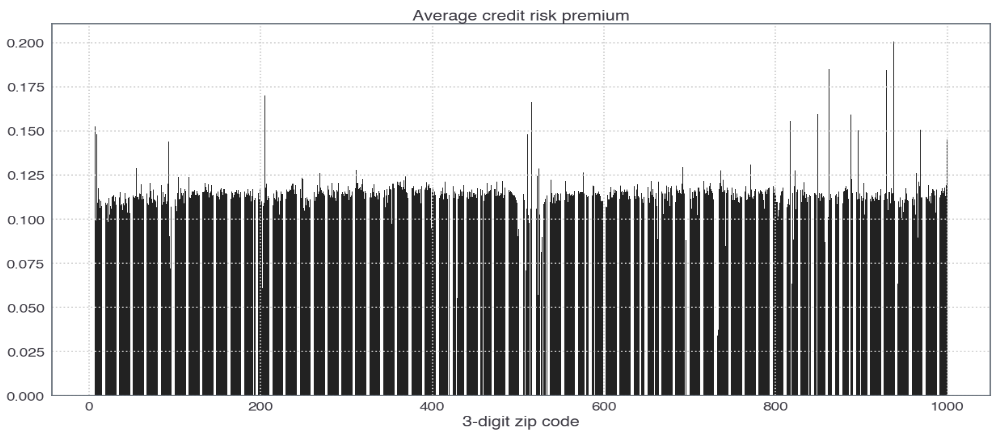

Figure 4: Credit pricing across location

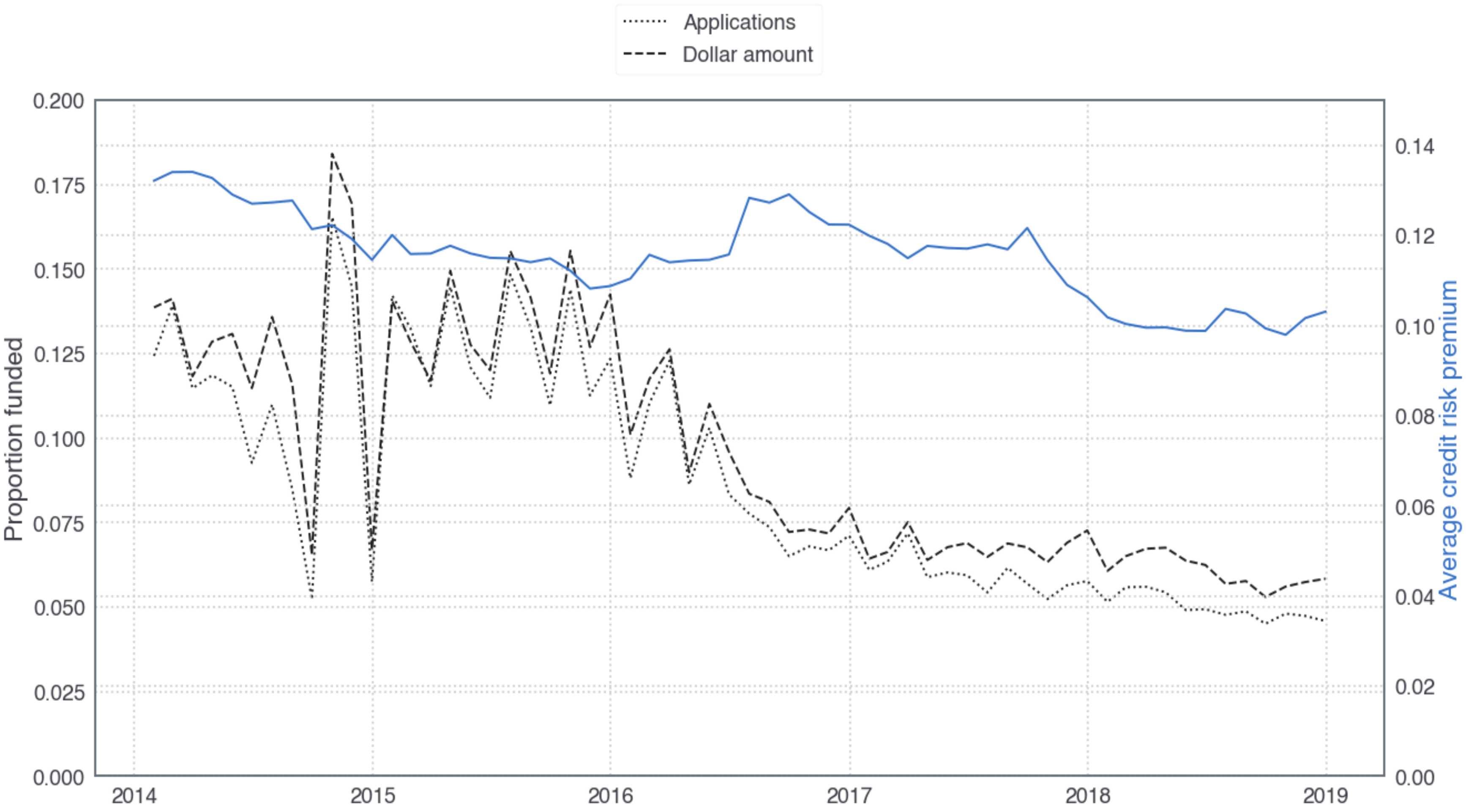

Figure 5: Funding decision across time

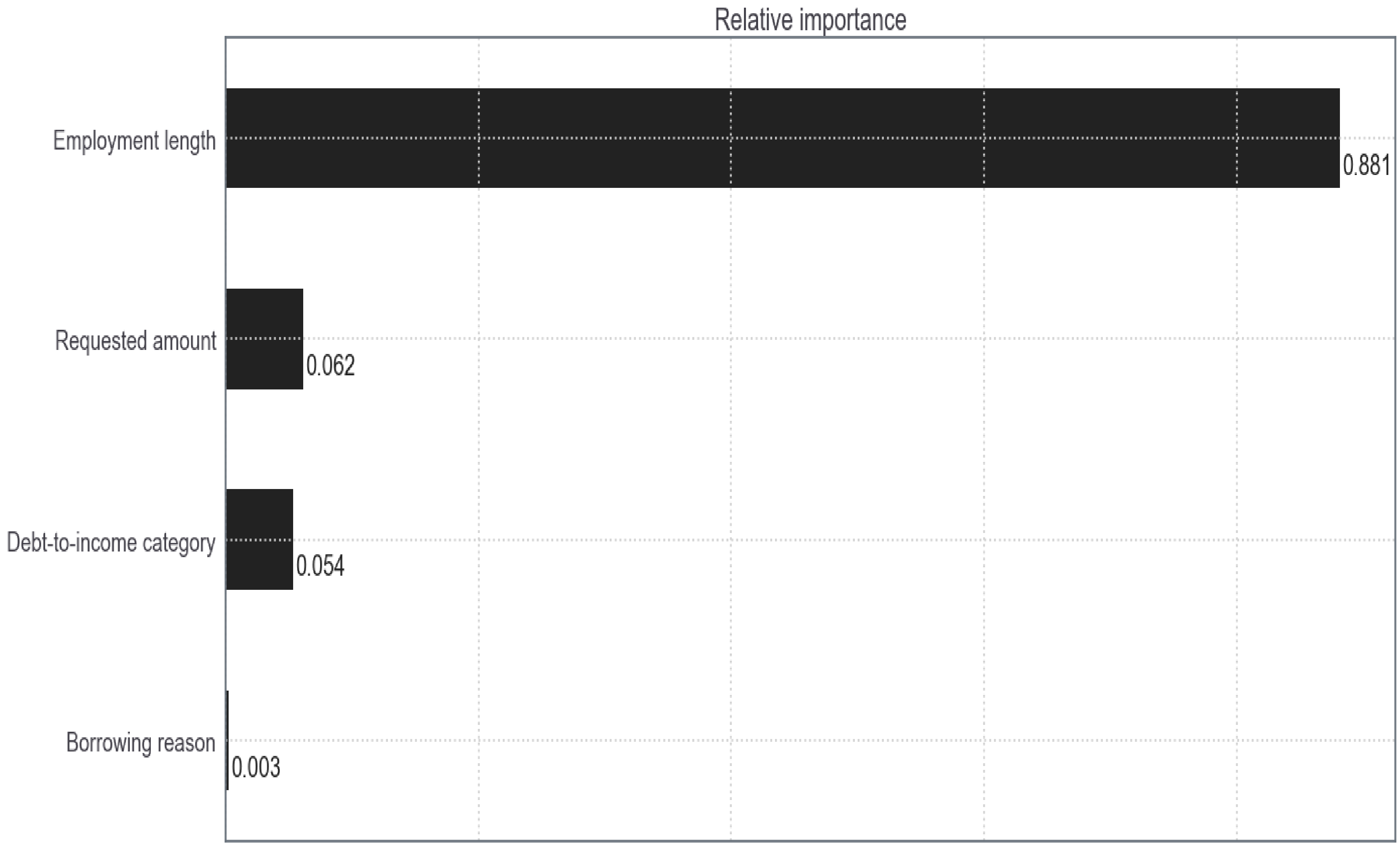

Figure 6: Feature importance for funding outcome

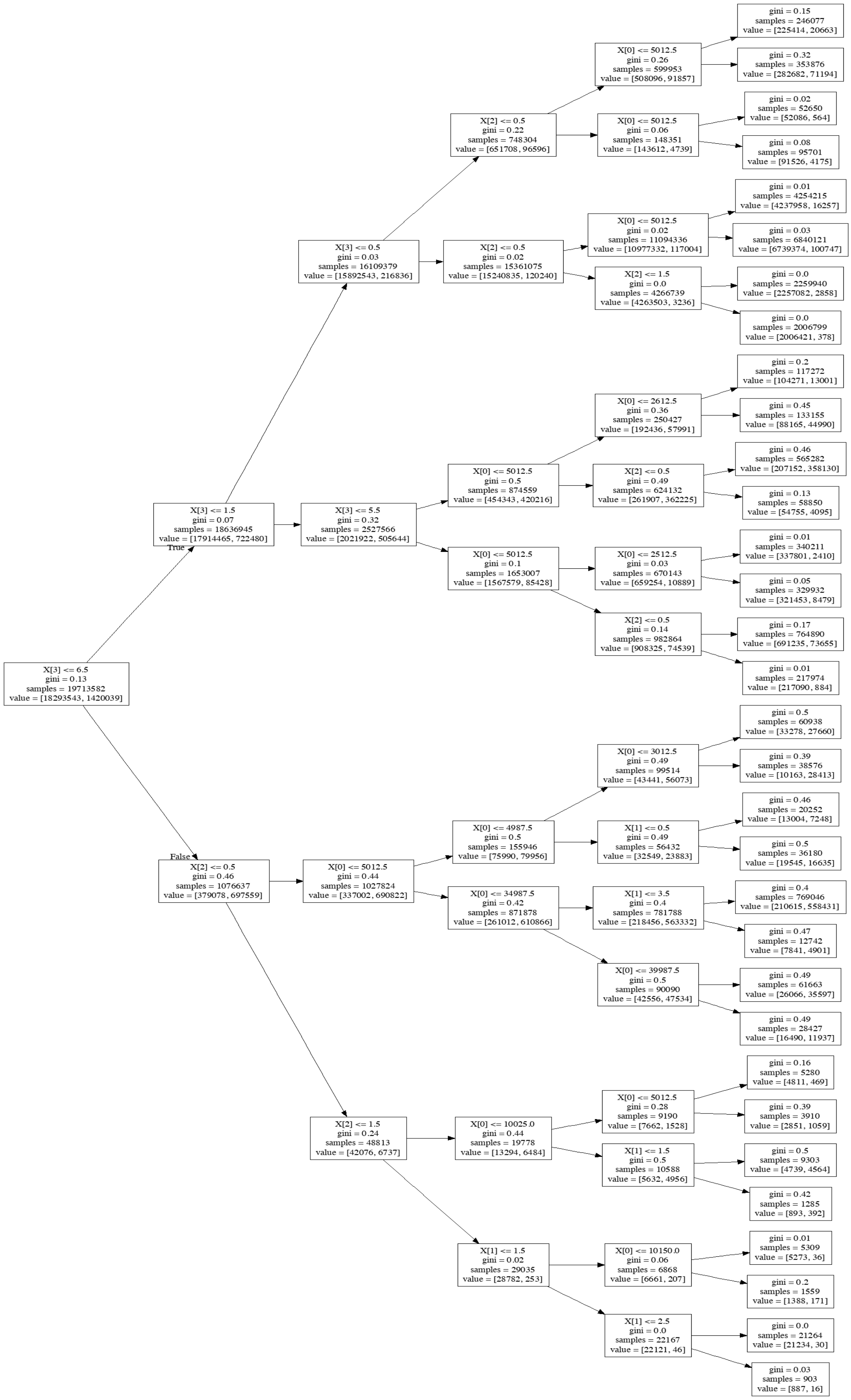

Figure 7: Funding decision process: X[0,1,2,3,] = [Requested amount, Borrowing reason, Debt-to-income category, Employment length]

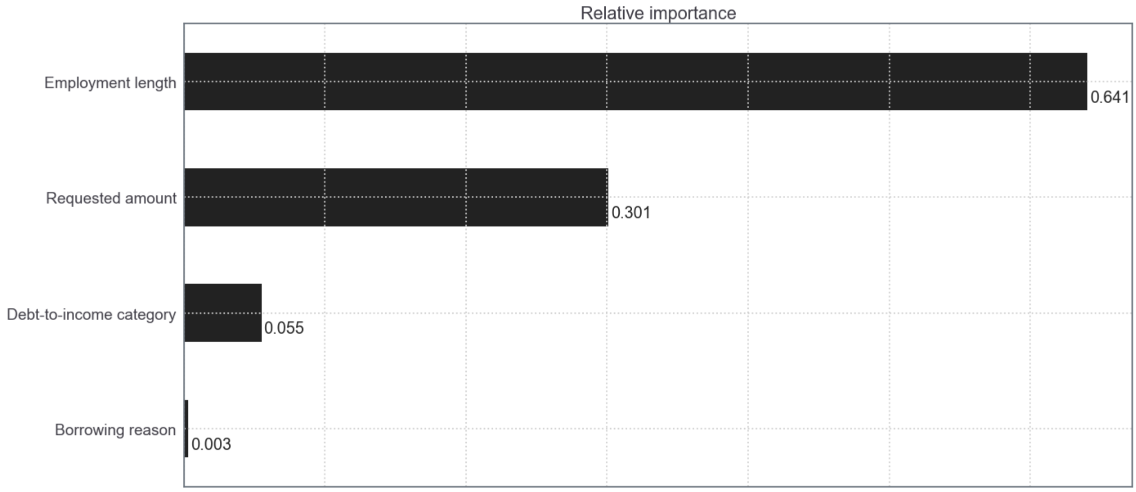

Figure 8: Feature importance for funding amount

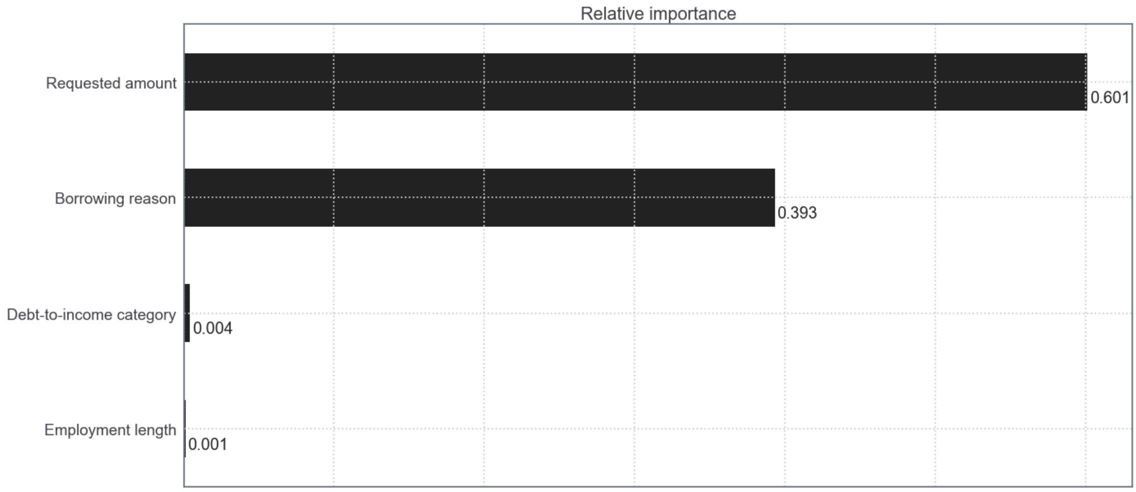

Figure 9: Feature importance for credit pricing (without loan grade)

Figure 10: Feature importance for credit pricing (with loan grade)

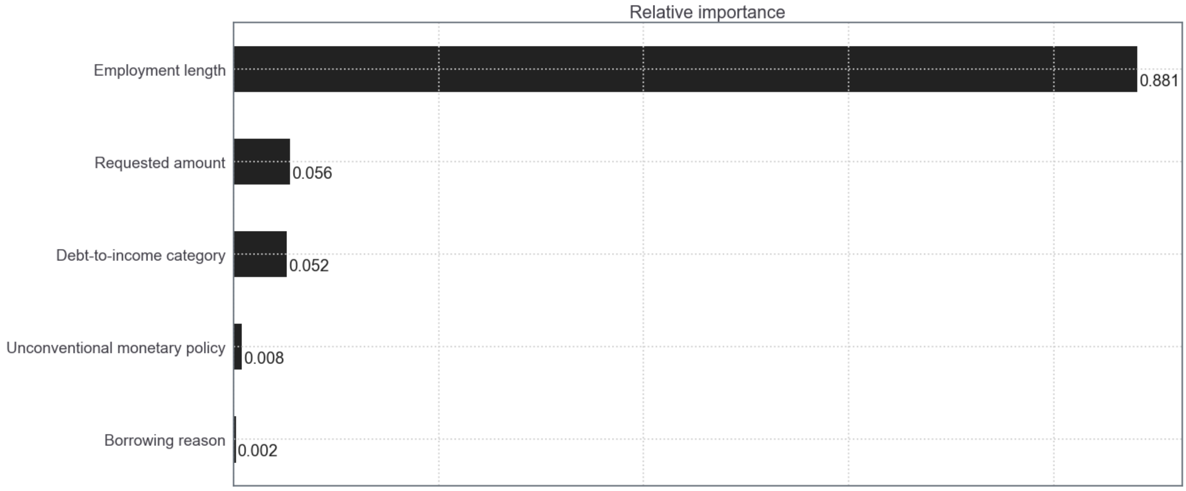

Figure 11: Feature importance for funding outcome including monetary policy regime

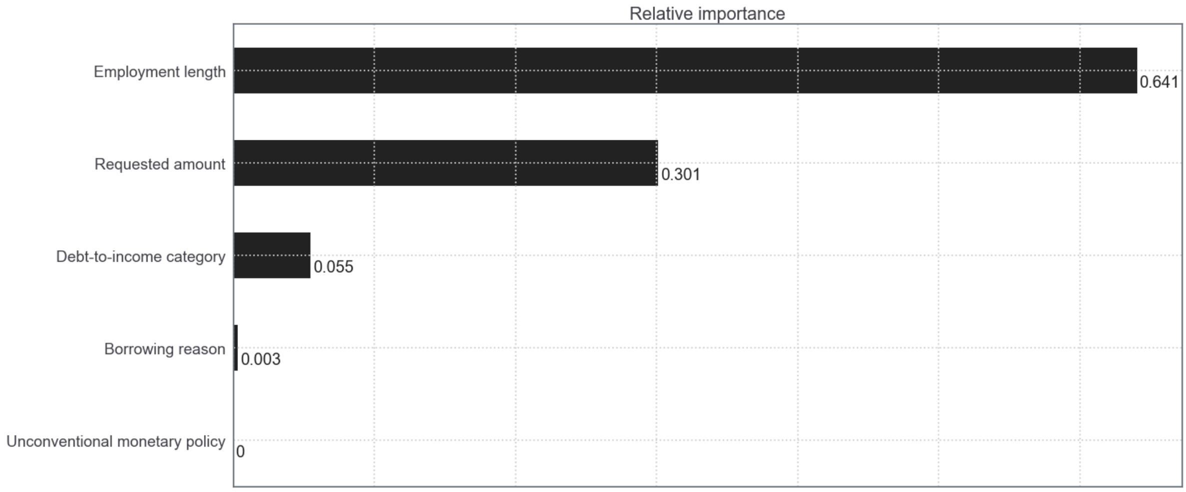

Figure 12: Feature importance for funding amount including monetary policy regime

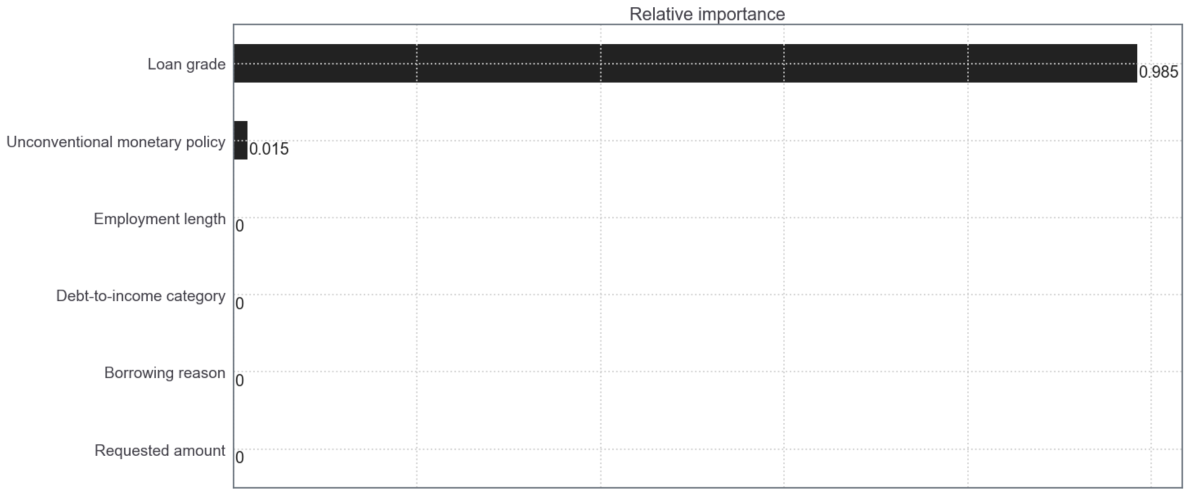

Figure 13: Feature importance for credit pricing including monetary policy regime