Home bias in P2P Lending Platform

HaoYin Zhi

Shanghai, University of Shanghai for Science and Technology, 200093, China

Email:1981144083@qq.com

Abstract. This article mainly concentrates on the impact of home bias in peer-to-peer (P2P) platforms. Since the absence of geographical barriers in online transfers and visible information flows to investors,, the spatial barriers to trading have been broken down and it is logical that no significant local preference should be shown. However, some studies in economics and finance still suggest that home bias appears in online platforms. Therefore, the purpose of this article is to verify that home bias does influence investment preferences by using the machine learning approach of Random Forest Model [1] to analyze data collected from lending platforms. The results show that home bias does have a significant impact on investment preferences, particularly among local borrowers, where lenders often relax loan terms, making it more likely for them to lend to people from the same area. This study offers a deeper understanding of the impact of home bias, providing constructive suggestions for advancing the push methods of future P2P platforms and applications of constructed models.

Keywords: Home bias; P2P; Random Forest; Lending platform; Transfer

1 Introduction

Home bias refers to an investor's stubborn, obsessive and excessive focus on their home market at the expense of the global market. It also means that agents (companies, funds, etc.) are more likely to deal with people who are geographically closer to them, in the same country or region, rather than strangers. This phenomenon can occur in many fields. In international trade, proximity preference occurs between national borders, i.e., one is more likely to engage in trade behavior with countries that are closer [6]. Local preference characteristics of investor behavior should be considered in market structure analysis, policy formulation, and social welfare provision. Since French and Poterba [2], a growing number of scholars have documented this phenomenon.

Home bias can have an impact not only in terms of investment, but also in other areas. In terms of accounting decisions, since financial reports are the main source of information about a company, home bias will have an impact [3]. Because a company is more well-known locally, investors will pay more attention to the financial statements of this company, so that the local accounting standards will be used in the production of corporate finance to reduce the wariness of investors. Academically, G. Andrew Karolyi has identified a new home bias, which he called academic home bias [8]. That is,

Y. Jiao et al. (eds.), *Proceedings of the 3rd International Conference on Internet Finance and Digital Economy (ICIFDE 2023)*, Atlantis Highlights in Economics, Business and Management 1,
https://doi.org/10.2991/978-94-6463-270-5_15

researchers will be more likely to look at national or local economic phenomena and ignore phenomena in other regions, and these ignored phenomena are often more likely to be cited.

In recent years, researchers have noted that the development of online trading models such as e-commerce and online platforms is expected to reduce investors' home bias when it comes to investing. However, home bias can still have an impact in online transaction models such as online crowdfunding platforms [4] and eBay [5]. For example, people will be more inclined to buy goods from the same region on the eBay platform and to raise funds for people from the same region on online crowdfunding platforms.

There are different explanations for the causes and characteristics of home bias in different contexts. For instance, differences in returns and risks in trading countries have an impact on home preferences and they have a correlation with returns and risks [7]. Moreover, home bias tends to be more pronounced during times of crisis, while receding during periods of relative calm. In addition, the degree of home bias appears to be inversely related to a country's level of corruption control and investment openness [8].

2 Method

This article has collected data in 2007 from two P2P platforms, Prosper and Lending Club. It attempts to demonstrate that home bias affects lenders on online P2P platforms, thereby giving evidence that online platforms do not completely eliminate home bias. Previous studies always used machine learning approach of Random Forest Model, while this article will add Random Forest Model and Linear Model into consideration.

Specifically, the article uses a machine learning approach comprising two stages. The first is the classification of variables. The interest rate can indicate the risk level of a lender, so that it is the explained variable and also a numerical variable. Since there are many factors in the original data set that cannot be specified, such as whether it is a weekend and whether the lender and borrower are from the same region, these variables are incorporated into the analysis as dichotomous variables (0 indicating absence of the attribute and 1 indicating its presence). They are both key explanatory variables. Furthermore, whether one is a homeowner (is_homeowner) is also identified as a key explanatory variable. Other than numerical variables, the study introduces control variables to the model to account for other potentially confounding factors..

In this article, the data set is divided into two parts. One is the training set and the other is the test set. The main function of the training set is to train the model with a large amount of data and tasks. Through continuous iterative training, the model will be fitted to a reasonable state of heterogeneity, and finally the trained model can be put into use. The test set is used to verify whether the model fits or not.

The name of the data set is data_p2p.csv. The data was collected from the two largest P2P lending sites in the US, Lending Club and Prosper, involving all lending data between March 18, 2007 and July 10, 2007. The lending data include various parameters at the time of a debit and various types of information about the borrower and the lender, such as the indicator creation_date that represents the time, the indicator MemberKey that represents a specific lender, and the indicators stated_monthly_income, monthly_debt, etc. that represent the borrower's monthly income and monthly debt. Among these data, the indicators that are important to build the model are BidAmount, which is used primarily to indicate the available funds contributed by the borrower to the lender. BorrowerMaximamRate is used to indicate the interest rate acceptable to the lender. Since the interest rate is used as an indicator of risk in economics, it is also an indicator to represent the borrower's risk. For better modeling, four indicators of is_weekend, homebias, creation_season and creation_time are included additionally. In adding variables, is_weekend, creation_season and creation_time are all time variables. Python is adopted to classify the date and time period. For the indicator of is_weekend, special calculation method is used to determine whether a day is a weekend day by confirming whether the result is greater than 5. For the indicators of creation_season and creation_time, python is used to divide 4 and 3 time periods respectively to determine the current season and time period. For the indicator of homebias, which is an additional indicator to represent the occurrence of the home bias phenomenon, python is used to determine if LenderState and BorrowerState are in the same region. If they are, the output will be 1; if not, the output will be 0.

The sample number of the original data set is 582,980. In order to solve the over-fitting or under-fitting problem occurred in the later modeling process, python is used to divide the training set and test set randomly. As a result, based on the ratio of 7:3, the sample number of the training set is 408086 and the sample number of the test set is 174,894.

2.1 Variable Definition

Interest rate is an indicator that stands for risk. The higher interest rate, the higher risk. Therefore, interest rate is the dependent variable y. The meanings of each variable are shown in Table 1.

Table 1. The Variables

Variable Type	Variable Name
Explained Variables	BorrowerMaximumRate
Key Explanatory Variables	is_weekend
Control Variables	homebias
	is_homeowner
	BidAmount
	ParticipationAmount
	daily_listings_num
	Funded_per
	stated_monthly_income
	monthly_debt

2.2 Regression ModelS

The linear regression model in this article is as follows, the coefficients and significance results of the linear regression model are shown in Table 2.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 is\_homeowner + \beta_8 is\_weekend + \beta_9 homebias$$

Table 2. The Numerical Value of the Variables

	coef	**P>|t|**
BidAmount	-0.0022	0.000
ParticipationAmount	0.0009	0.000
daily_listings_num	-0.0001	0.000
Funded_per	0.4418	0.000
stated_monthly_income	-5.3e-06	0.000
monthly_debt	0.0006	0.000
is_homeowner	-1.7115	0.000
is_weekend	-0.0119	0.534
homebias	0.8388	0.000

Due to the large number of MemberKey and occupation indicators, they are not counted in the table.

From the above table, it can be seen that the explanatory variable is_homeowner has a negative effect on the interest rate and the result is significant. This suggests that when a borrower has a house, the interest rate on the loan will consequently fall because those who have a house have better solvency. The control variable monthly_debt is positively and significantly related to the interest rate. This is because the solvency of those who have debt decreases. The higher the debt, the lower the solvency. The indicator stated_monthly_income is negatively correlated with the interest rate. The higher the income, the better the solvency of a person.

The explanatory variable homebias is positively correlated with the interest rate and the result is significant, indicating that for lenders and borrowers coming from the same region, the loan interest rate will subsequently increase. This represents that borrowers are willing to give loans to people with higher interest rates and greater risk. It also reflects that home bias is present in the P2P lending platforms.

This article then uses the Linear Regression, Ridge, and Lasso models of sklearn to build three regression models and obtains model parameters of linear regression models on the basis of the sclera training data set. It further calculates the MSE and MAE of the training set and test set to judge the effect of models. From the results in the following Table 3, it is found that the models of Linear Regression, Ridge regression, and Lasso regression perform similarly in the training set and test set in terms of MSE and MAE.

Table 3. The Comparison of the Models

Model	**Metrics**	**Training Set**	**Test Set**
Linear	MSE	32.80	32.99

	MAE	4.59	4.59
Ridge	MSE	32.85	32.84
	MAE	4.59	4.59
Lasso	MSE	48.70	48.69
	MAE	5.80	5.80

2.3 Random Forest Model

Similarly, the Random Forest Regression model of sklearn is used to build the random forest model. The training set data are used to obtain the model parameters. Then, mean_squared_errorand mean_absolute_error functions of sklearn are used to calculate the MSEand MAE of the training set and test set, respectively, so as to judge the effect of the model. As can be seen from the Table 4 below, the results of the random forest model are quite different from the linear regression.

The parameters used in the random forest model are all default parameters.

Table 4. Random Forest Model

	Training Set	**Test Set**
MSE	1.61	8.52
MAE	0.59	1.50

2.4 Model Comparison

From the above four models, it can be clearly concluded that the random forest model outperforms other three regression models. This is evident from the considerably smaller values of MSE and MAE obtained by the random forest model as compared to the remaining three models, which indicates that the random forest model exhibits increased stability.

This article presents a comparative analysis between the random forest model and other linear models. From above experimental procedures, it is shown that the results between the training set and test set do not differ much, which proves that the models are both correct. However, the results of fitting the data set using the random forest model are much better than those using the linear regression models. Their MSE and MAE are 32.99, 4.59 and 8.52, 1.5, respectively. Consequently, the random forest model is significantly better than linear regression models in this context.

The partial dependence graph of the Random forest model is further drawn as shown in Figure 1,it can be seen that the explanatory variable is_homeowner has a negative effect on the interest rate. monthly_debt is positively and significantly relat-ed to the interest rate. The in-dicator stated_monthly_income is negatively correlated with the interest rate. The explanatory variable homebias is positively correlated with the interest rate, indicating that for lenders and borrowers coming from the same re-gion, the loan interest rate will subsequently increase.

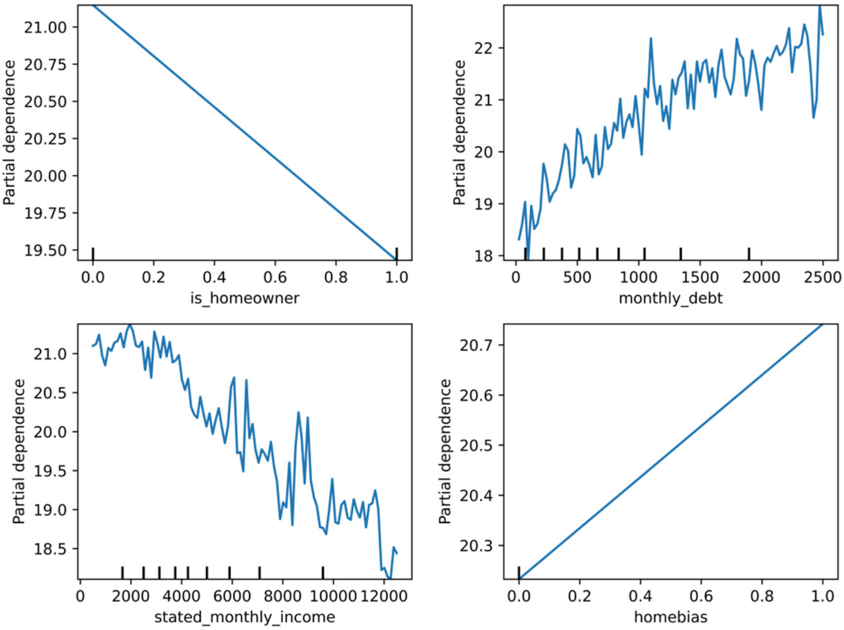

Fig. 1. Partial dependence graph of Random forest

3 Conclusion

In summary, based on the overall results of linear regression models, it is easy to conclude that the key indicators such as homebias, BidAmount, ParticipationAmount, daily_listings_num, funded_per, stated_monthly_income and monthly_debt are all significant, which is sufficient to prove that all these indicators have an impact on the interest rate. The most critical explanatory variable homebias is also positively correlated with the interest rate, demonstrating the presence of home bias in lending platforms.

The findings yielded by the random forest model presented in this article demonstrate its potential for broad extension to most lending platforms, facilitating the prediction of loan interest rates. It can also further improve the pushing methods of lending platforms. For example, when a platform pushes borrowers to lenders, it can give priority to people from the same region, so that the tolerance of lenders to high interest rate borrowers will rise, which can increase the efficiency of the lending platform and optimize the experience of customers.

The proposed model is not limited to the analysis of P2P lending platforms, but can be extended to a wider range of financial fields. In stock trading, the regional characteristics of investors and companies may be leveraged to gauge the degree of investor preference for a particular stock. Another potential application in the Chinese market could be assessing the predilections of local banks towards individual and corporate borrowers. For example, local commercial banks may prefer local individual borrowers or enterprises.

Reference

1. Pal, M. (2005). Random forest classifier for remote sensing classification. International journal of remote sensing, 26(1), 217-222. https://doi.org/10.1080/01431160412331269698.
2. French, K. R., Poterba, J. M. (1991). Investor diversification and international equity markets. The American Economic Review, 81, 222-226. DOI 10.3386/w3609.
3. Bradshaw, M. T., Bushee, B. J., Miller, G. S. (2004). Accounting choice, home bias, and US investment in non-US firms. Journal of accounting research, 42(5), 795-841. https://doi.org/10.1111/j.1475-679X.2004.00157.x.
4. Lin, M., Viswanathan, S. (2016). Home bias in online investments: An empirical study of an online crowdfunding market. Management Science, 62(5), 1393-1414. https://doi.org/10.1287/mnsc.2015.2206.
5. Hortaçsu, A., Martínez-Jerez, F. A., Douglas, J. (2009). The geography of trade in online transactions: Evidence from eBay and mercadolibre. American Economic Journal: Microeconomics, 1(1), 53-74. DOI: 10.1257/mic.1.1.53.
6. Lindblom, T., Mavruk, T., Sjögren, S., Lindblom, T., Mavruk, T., Sjögren, S. (2017). The Relation Between Local Bias, Home Bias, and Financial Sophistication. Proximity Bias in Investors' Portfolio Choice, 241-263. DOI: 10.1007/978-3-319-54762-6_11.
7. Kim, H., Cho, S. H., Kim, Y. (2015). Home bias, risk differential, and cultural spatial spillover effects. Journal of International Money and Finance, 51, 114-136. https://doi.org/10.1016/j.jimonfin.2014.10.007.
8. Andrew Karolyi, G. (2016). Home bias, an academic puzzle. Review of Finance, 20(6), 2049-2078. http://hdl.handle.net/10.1093/rof/rfw007.