Empirical Analysis of Factors Affecting Mortgage Permissibility

Pengyu Mu

University of International Business and Economics, Beijing, 100029

Keywords: Mortgage permit rate, empirical analysis, econometrics, Gauss-Markov assumption

Abstract: This paper uses the sample data of 1756 mortgage applicants from the US Econ_loan to study the influencing factors of the mortgage loan permit rate. Starting from the Gauss-Markov hypothesis, this paper discusses the measurement model hypothesis testing, estimation error, function form, virtual variables and heteroscedasticity. The natural logarithm of the estimated value of assets has a statistically significant positive impact on the mortgage permit rate. Personal credit characteristics have a joint significance on the mortgage permit rate. There is a race for the mortgage permit rate discrimination.

1. Introduction

Mortgage business is one of the main asset business of commercial banks [1]. The loan applicant submits a mortgage application to the bank, submits information about income, debt, assets, credit and funds available for down payment and settlement costs, including personal characteristics of the loan applicant's economic and non-economic aspects [2]. The bank evaluates the credit status of the loan applicant based on these interests, and finally determines whether the loan applicant can obtain the loan and the amount of the loan [3]. Then, what factors affect the mortgage permit rate, what is the significance of these factors? Is there any racial discrimination in the mortgage permit rate? These issues should be further studied.

2. Hypothesis

Hypothesis 1: The natural logarithm of the estimated value of an asset $\ln apr$ is a statistically significant positive impact on the mortgage loan permit rate. According to the principle of the return on mortgage loans, the higher the fixed asset value of the mortgage asset, the greater the probability that a mortgage applicant can apply for a loan, and the higher the maximum mortgage loan cap available high.

Hypothesis 2: Personal credit characteristic factors have a joint significance to the mortgage permit rate. Personal credit characteristic factors are an important factor for banks to evaluate mortgage applications, and it is expected to have a high joint significance.

Hypothesis 3: There is racial discrimination in the mortgage permit rate. Even in developed countries that are recognized as equal and free, racial discrimination is still widespread. This article predicts that racial discrimination still exists in the United States and that racial discrimination is highly statistically significant.

3. Research methods

3.1 Research Methods

(1) Descriptive statistics

In the definition of relevant variables, the descriptive statistical method is used to analyze the mortgage loan approval rate of mortgage applicants: economic factors, loan characteristics, personal social characteristics, and personal credit characteristics.

(2) Econometric analysis

This study uses STATA 16 software to perform regression analysis on the cross-section data. Based on the Gauss-Markov assumption, the mortgage permit rate is used to regressionly analyze the loan applicant's economic factors, loan characteristics, personal social characteristics and personal credit characteristics, put forward hypotheses and test t-test and F-test at significance levels of 10%, 5%, and 1%; then, discuss possible estimation bias, and bring about whether the explained variable is a binary variable. The heteroscedasticity problem is discussed, and the robustness test is used to evaluate the strength of the model and index interpretation ability, and the use of dummy variables is discussed. Finally, the coefficient of the explanatory variable is explained economically.

3.2 Data Source

According to Econ_loan data, 1756 sample data are selected from the situation of U.S. resident loans, including economic factors of mortgage applicants, loan characteristics, personal social characteristics, and personal credit characteristics.

(1) Processing of missing data

In the econometric model, some samples have missing data and cannot be compensated and replaced. For this part of the sample (217), the relevant data is directly removed, which is the percentage of minors in the local population ($\min 30$, 175), monthly income ($atotinc$, 16), missing gender data ($male$, 15), credit rating data is missing (rep, 6), missing marriage data and relative data ($married$ with dep, 3) and missing housing numbers ($unit$, 2).

The above data loss is random, rather than systemic data loss.If it is directly removed, it will reduce the sample size.

(2) Handling of outliers

In addition, there are still outliers in the sample that cannot be compensated and replaced. For this part of the sample (16), the relevant data is directly removed, which are current assets (liq, 9), mortgage term ($term$, 4), credit history ($gdlin$, 2) and report credits ($lines$, 1).

(3) Discussion of the function form-logarithm of some explanatory variables

Taking the logarithm of the explanatory variable that involves a large amount of assets can reduce the impact of outliers such as maximum values on the estimation results. The effect of measuring the percentage change of the independent variable on the explanatory variable can be discussed without considering the unit of measurement. It is beneficial to make the sample close to the normal distribution, and taking the logarithm is helpful to obtain the corresponding elasticity, which has economic meaning.

Generate new variables $\ln appinc$ natural logarithm of the applicant's annual income, $\ln atotinc$ natural logarithm of the applicant's monthly income, $\ln h\exp$ natural logarithm of the applicant's monthly mortgage payments, $\ln price$ natural logarithm of the applicant's daily expenses, $\ln liq$ natural logarithm of fixed assets, $\ln apr$ natural logarithm of the estimated value of the asset, $\ln netw$ natural log of net worth.

In the process of generating new variables, there are data with negative explanatory variables (9) after processing. In order not to discard the relevant data of the 9 sample mortgage applicants, the processing of this part of the data is to add the explanatory variables take the natural logarithm after 1. $\ln netw$ has 85 samples that are negative, so the natural log of the net. $\ln netw$ value is divided into two steps: the first step is to take the absolute value of the net value plus the natural logarithm of 1, and the second step is to take the natural log symbol according to the sign of the net value. Finally, the sample capacity in the measurement model is 1756.

4. Analysis process

4.1 Multiple Regression Analysis Assumptions

Assume that MLR.1 is linear to the parameters, assume that MLR.2 is randomly sampled, assume that MLR.3 is not completely collinear, assume that the conditional value of MLR.4 is zero, assume that MLR.5 is homoscedastic, and assume that MLR.6 is normal.

Based on the assumptions MLR.1 ~ MLR.4, the unbiased estimator of the measurement model OLS can be obtained; based on the assumptions MLR.1 ~ MLR.5, the Gauss-Markov theorem of the measurement model in this paper can be fully satisfied, and the effectiveness of the OLS is satisfied. The required parameter is the optimal linear unbiased estimator; based on the assumptions MLR.1 ~ MLR.6, t-test and F-test of the parameters to be estimated can be performed. In particular, relax the assumption MLR.6, under large sample conditions, You can perform the parameter estimation χ^2 test.

4.2 Factors Affecting the Mortgage Permit Ratio

In summary, for the above econometric model, the method of gradually increasing the control variable is used to perform OLS regression analysis. The specific results are shown in Table 1:

Table 1. Estimated results of influencing factors on mortgage permit rate—OLS

variable	Model 1	Model 2	Model 3	Model 4
lnappinc	0.0037(0.0143)	-0.0053(0.0154)	-0.0061(0.0152)	0.0002(0.0149)
lnapr	0.0829***(0.0189)	0.0617***(0.0207)	0.0541**(0.0213)	0.0447**(0.0213)
multi	-0.1640***(0.0232)	-0.1337***(0.0270)	-0.1061***(0.0271)	-0.1040***(0.0265)
self	-0.0606***(0.2072)	-0.0626***(0.0224)	-0.0611***(0.0221)	-0.0658***(0.0217)
hrat		-0.0032***(0.0012)	-0.0033***(0.0012)	-0.0034***(0.0011)
loanprc		-0.1639***(0.0436)	-0.1537***(0.0438)	-0.1475***(0.0429)
fixadj		0.0147(0.0166)	0.0172(0.0163)	0.0189(0.0160)
cosign		0.0039(0.0458)	0.0014(0.0454)	0.0089(0.0444)
term		0.0000(0.0001)	0.0000(0.0001)	0.0001(0.0001)
vr		-0.0349**(0.0153)	-0.0251*(0.0151)	-0.0250*(0.0148)
pubrec		-0.3157***(0.0290)	-0.2843***(0.0289)	-0.2108***(0.2941)
male			-0.0186(0.0202)	-0.0209(0.0198)
old			-0.0424***(0.0156)	-0.0490***(0.0153)
sch			-0.0224(00184)	-0.0261(0.0180)
married			0.0395**(0.0177)	0.0384**(0.0174)
dep			-0.0009(0.0248)	-0.0019(0.0071)
white			0.1606***(0.1606)	0.1284***(0.0246)
hispan			0.0536(0.0536)	0.0278(0.0375)
chist				-0.0890(0.0441)
mortlat1				-0.0461(0.0522)
mortno				0.0123(0.0167)
cons				-0.0578***(0.0100)
Sample size	1756	1756	1756	1756
R^2	0.0357	0.1182	0.1490	0.1885
$\overline{R}^2$	0.0335	0.1127	0.1403	0.1783

Note: *, **, and *** represent that the coefficients are significant at the levels of 10%, 5%, and 1%, respectively.

A. Testing hypotheses

(1) Hypothesis testing of a single population parameter

For hypothesis 1: The natural logarithm of the estimated value of an asset $\ln apr$ is a statistically significant positive effect on the mortgage permit rate. In Model 4, for $\ln apr$ using t test:

Null hypothesis $H_0 : \beta_{\ln apr} = 0$

Two-sided alternative hypothesis: $H_1 : \beta_{\ln apr} \neq 0$

Construct t statistics: $t_{\ln apr} = \frac{0.0447}{0.0213} \approx 2.10$

And at a significance level of 5%, $t_{1756-22-1} = t_{1733} = 1.96$

Therefore, 2.10> 1.96

Therefore, at a significance level of 5%, rejecting the null hypothesis, accepting the alternative hypothesis, the natural logarithm of the estimated value of the asset has a statistically significant positive impact on the mortgage loan permit rate, and the hypothesis is true.

This shows that for every 1 percentage point increase in the estimated value of the asset, the mortgage application permit rate will increase by 0.0447 percentage points. The t-test operation process for the overall parameters of other explanatory variables is the same, and the symbol "*" in Table 3 indicates significance level.

(2) Test a linear combination hypothesis about parameters

One of the four factors that affect the mortgage loan approval rate is the personal credit characteristic factor. Among the personal credit characteristic factors of Model 1 to Model 4, the significance level of different explanatory variables is different, and most of them are insignificant; but as a measure of mortgage. As an important indicator of loan applicants' credit, personal credit characteristics still need to be placed in the model. Therefore, hypothesis 2: There is a joint significance of economic factors on the mortgage loan permit rate. In Model 3 and Model 4, f is used for hypothesis 2. test:

Null hypothesis: H_0: $\beta_{chist} = \beta_{mortlat1} = \beta_{mortno} = \beta_{cons} = 0$

Alternative hypothesis: H_1: β_{chist}, $\beta_{mortlat1}$, β_{mortno}, β_{cons} at least one is not 0

Construct f statistics: $F = \frac{(R_4^2 - R_3^2)/q}{(1-R_4^2)/(n-k-1)} = \frac{(0.1885-0.1490)/4}{(1-0.1885)/(1756-22-1)} = 21.09$

And at a significance level of 1%, $F_{(4,1733)} = 3.32$

Therefore, 21.09> 3.32

Therefore, at the significance level of 1%, rejecting the null hypothesis and accepting the alternative hypothesis, the personal credit characteristic factors have joint significance to the mortgage permit rate, and hypothesis 2 holds.

This shows that the personal credit characteristics as an important indicator of the mortgage applicant's credit still need to be placed in the model.

4.3 Estimation Bias

In the regression process of Model 1 to Model 4, the estimation error in Model 1 may be caused by the lack of explanatory variables. The lack of ethnicity of explanatory variables, etc., will cause the problem of missing variables, which will cause the OLS estimator of Model 1. Because of racial discrimination, white people often own higher assets than non-white people, so the effect of the natural logarithm of the fixed value of assets on the mortgage permit rate is exaggerated.

Specifically, Model 1 is compared with the other three models, $\tilde{\beta}_{\ln apr} = \beta_{\ln apr} + \beta_{white}\delta_1$,among them $\beta_{white} > 0$. In regression $white = \delta_0 + \delta_1 \ln apr + v$ in, $\delta_1 > 0$ and therefore makes $\tilde{\beta}_{\ln apr}$ an upward bias. Table 2 is the missing race variable *white* explanatory variables $\ln apr$ error.

Table 2. Missing Race Variables *white* Explanatory variables $\ln apr$ Bias

variable	Model 1	Model 5
lnappinc	0.0037(0.0143)	0.0289(0.0140)
lnapr	0.0829***(0.0189)	0.0626***(0.0187)
multi	-0.1640***(0.0232)	-0.1241***(0.0272)
self	-0.0606***(0.2072)	-0.0640***(0.0228)
white		-0.1832***(0.0210)
Sample size	1756	1756
R^2	0.0357	0.0756
$\overline{R}^2$	0.0335	0.0730

In addition, other explanatory variables may cause the explanatory variables to be generally biased forward. Therefore, in Model 2, Model 3, and Model 4, $\beta_{\ln apr}$ has been some decline.

4.4 Dummy Variables

Of the 22 explanatory variables in this paper, there are a total of 15 dummy variables. The dummy variables have the functions of 0-1 classification, multiple classification, and cross-terms. The following uses human race as an example to analyze the role of dummy variables and the interpretation of dummy coefficients.

First, let's look at the multi-classification effect of dummy variables. In the original data, there are three types of races: white, black, and Hispanic, but if you set 3 dummy variables, you will enter the dummy variable trap-the problem of complete collinearity. Therefore, When setting up dummy variables, two dummy variables are used to measure the difference in mortgage permit rates between different races.

Second, look at the interpretation of the coefficients of the dummy variables. In Model 4, the explanatory variables *white* coefficient β_{white}. Expressed that, with other conditions unchanged, whites increased mortgage permit rates by 0.1284 percentage points over blacks; explanatory variables *hispan* coefficient β_{hispan} says Hispanic raises mortgage approval rate by 0.0278 percentage points over blacks; therefore, whites increase $\beta_{white} - \beta_{hispan} = 0.1284 - 0.0278 = 0.1006$ percentage points.

Because the original model in this article is a linear probability model and the explanatory variable is a binary variable, there must be a heteroscedasticity problem. The fourth section analyzes the heteroscedasticity problem.

4.5 Heteroscedasticity

(1) White test

White test for heteroscedasticity for Model 4, the results are shown in Table 3.

Table 3. White test results

Source	chi2	df	p
Heteroskedasticity	458.45	258	0.0000
Skewness	362.23	22	0.0000
Kurtosis	222.47	1	0.0000
Total	1043.14	281	0.0000

It can be found that the test results are significant at the 1% level, and there is a serious heteroscedasticity problem in Model 4, which violates the assumption of MLR.5 homoscedasticity. Although the OLS estimation results are still unbiased, the validity may not even be necessarily satisfied. There is also a certain bias in the test results. Therefore, the robustness test is performed next.

(2) Robustness test

Assuming that the form of heteroscedasticity is known to be a fixed function or related to an explanatory variable, you can construct the form of homoskedasticity and apply generalized least squares methods such as weighted least squares to solve the heteroscedasticity problem of Model 4, but due to the large sample size There are many explanatory variables, so the form of heteroskedasticity cannot be determined, and the robustness test is directly used. The results are shown in Table 4.

From Table 4, there is basically no difference in the coefficients of the robust test. The robust standard error is basically similar to the standard error under heteroscedastic conditions, with only minor differences, and most of the robust standard errors are slightly larger than the standard error of model 4.

Table 4. Comparison of heteroscedastic structure and modified heteroscedastic results

variable	Model 4	Model 6
lnappinc	0.0002(0.0149)	0.0002(0.0257)
lnapr	0.0447**(0.0213)	0.0446(0.0207)
multi	-0.1040***(0.0265)	-0.1040***(0.0270)
self	-0.0658***(0.0217)	-0.0658***(0.0224)
hrat	-0.0034***(0.0011)	-0.0034**(0.0012)
loanprc	-0.1475***(0.0429)	-0.1475***(0.0436)
fixadj	0.0189(0.0160)	0.0189(0.0166)
cosign	0.0089(0.0444)	0.0089(0.0458)
term	0.0001(0.0001)	0.0001(0.0001)
vr	-0.0250*(0.0148)	-0.0250(0.0153)
pubrec	-0.2108***(0.2941)	-0.2108***(0.0290)
male	-0.0209(0.0198)	-0.0209(0.0210)
old	-0.0490***(0.0153)	-0.0490***(0.0159)
sch	-0.0261(0.0180)	-0.0261(0.0186)
married	0.0384**(0.0174)	0.0384**(0.0186)
dep	-0.0019(0.0071)	-0.0019(0.0073)
white	0.1284***(0.0246)	0.1284***(0.0327)
hispan	0.0278(0.0375)	0.0278(0.0500)
chist	-0.0890(0.0441)	-0.0890*(0.0534)
mortlat1	-0.0461(0.0522)	-0.0461(0.0679)
mortno	0.0123(0.0167)	0.0123(0.0155)
cons	-0.0578***(0.0100)	-0.0578***(0.0114)
Sample size	1756	1756
R^2	0.1885	0.1885
$\overline{R}^2$	0.1783	

5. Conclusion

Using a sample of 1,756 U.S. mortgage applicants in the data Econ_loan, this article conducted a related study on the factors that affect the mortgage permit rate, and reached the following conclusions: First, the natural logarithm of the estimated value of assets $\ln apr$. There is a statistically significant positive impact on the mortgage permit rate; second, the personal credit characteristic factors have a joint significance on the mortgage permit rate; third, there is racial discrimination in the mortgage permit rate; and fourth, the estimated asset value, the natural logarithm, marital status, and race have a significant positive impact on the mortgage permit rate, and the real estate situation, self-employment, mortgage rate, mortgage rate, bankruptcy filing, and age have a significant negative impact on the mortgage permit rate .

In the final model 4, R^2 although it is only 0.1885, which means that only 18.85% of the explained variables are explained by the explanatory variables, it is not only necessary to consider in the social disciplines R^2 . The size of the formula also needs to consider the economic meaning of the formula, so it has certain validity.

References

[1] Jeffrey M. Woodridge. Introduction to Econometrics: A Modern Perspective (Fifth Edition) [M]. Beijing: Renmin University of China Press, 2018.

[2] Li Shi, Wei Zhong, Ding Sai. An Empirical Analysis of the Unequal Distribution of the Property of Chinese Residents and Its Causes [J]. Economic Research, 2005, 34(6): 34-41.

[3] Liang Yunwen, Huo Zhen, Liu Kai. An Empirical Study of the Property of Chinese Urban and Rural Residents [J]. Economic Research, 2010, 54(10): 14-21.