Further Examination of the Relationship Between Race and Gender in Mortgage Lending within the State of Oklahoma. *

Opal Fraser†

April 24, 2023

Abstract

The project answers a question about lending in Oklahoma; do mortgage rates differ among race or gender? The difference in difference analysis looks at data before and after the COVID 19 pandemic. Using linear regression the data-set is analyzed to reveal differences. (notfinished).

*

†Department of Economics, University of Oklahoma. E-mail address: opalfraser@ou.edu

1 Introduction

The purpose of this report is to examine the relationship between the amount of a mortgage and the demographic characteristics of the borrower before and after the COVID 19 pandemic. The study uses data collected by the Federal Housing Finance Agency from a random sample of loan-level mortgage acquisitions acquired in 2019 and 2021. The analysis employs multiple linear regression models to estimate the impact of several independent variables, including race, gender, age, income, credit score, and loan-to-value ratio, on the dependent variable, the amount of the mortgage note/percentage rate.

2 Literature Review

Previous work by [1] shows that educational decisions are an important determinant of later-life earnings. This point is driven further in follow-up work by [1] and [1].

3 Data

The primary data source for this research is the FHL Bank Public Use Database. Table 1 contains summary statistics.

4 Empirical Methods

While my approach explores a number of different approaches, the primary empirical model can be depicted in the following equation:(I will code math properly for final report!) [1] LTV log-

noteamt + noteratepercent + LTV + bo1race*year + bo1gender*year + bo1age + debtexpenseratio [2] lognoteamt noteratepercent + LTV + bo1race*year + bo1gender*year + bo1age + debtexpenseratio + hsexpenseratio [3] noteratepercent debtexpenseratio + lognoteamt*year + LTV + bo1race*year + bo1gender*year + bo1age

$$Y_{it} = \alpha_0 + \alpha_1 Z_{it} + \alpha_2 X_{it} + \varepsilon, \tag{1}$$

where Y_{it} is a continuous outcome variable for unit i in year t, and Z_{it} are characteristics about the firm at which i is working, while X_{it} are characteristics about i. The parameter of interest is α_1.

5 Research Findings

The main results are reported Figure 1, 2, 3

6 Conclusion

Here I will add my conclusion on the regression analysis. I did not get this far since I've been doing the data search, wrangling and now the difference in difference analysis.

References

[1] Arpit Gupta, Christopher Hansman, and Pierre Mabille. Financial constraints and the racial housing gap. 2022.

	Unique (#)	Missing (%)	Mean	SD	Min	Median	Max	
totmonthlyincome	1090	0	9249.2	8286.5	1138.0	7260.0	110741.0	
LTV	82	0	79.3	16.2	7.0	80.0	104.0	
bo1age	68	0	46.6	14.7	20.0	45.0	90.0	
noteratepercent	26	0	4.0	0.5	2.9	3.9	5.8	
noteamt	783	0	194408.9	102657.3	26800.0	172000.0	484350.0	
hsexpenseratio	950	0	18.9	7.9	0.0	18.2	78.0	
debtexpenseratio	972	0	31.0	9.0	1.1	31.4	78.0	
lognoteamt	783	0	12.0	0.5	10.2	12.1	13.1	
logtotmonthincome	1090	0	8.9	0.6	7.0	8.9	11.6	

Figure 1: 2019 data

Figures and Tables

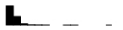
	Unique (#)	Missing (%)	Mean	SD	Min	Median	Max	
totmonthlyincome	1002	0	9838.6	8693.7	1342.0	8016.5	141878.0	
LTV	80	0	75.6	15.0	16.0	80.0	104.0	
bo1age	65	0	47.1	13.4	20.0	45.0	86.0	
noteratepercent	15	0	2.8	0.3	2.0	2.9	3.6	
noteamt	647	0	212522.4	115728.0	29000.0	182760.5	548250.0	
hsexpenseratio	855	0	18.0	7.9	0.0	16.9	50.2	
debtexpenseratio	882	0	30.6	9.2	5.6	31.2	64.5	
lognoteamt	647	0	12.1	0.6	10.3	12.1	13.2	
logtotmonthincome	1002	0	9.0	0.6	7.2	9.0	11.9	

Figure 2: 2021 data

	(1)	(2)	(3)
(Intercept)	−10.623	11.414***	4.907***
	(7.541)	(0.140)	(0.267)
lognoteamt	6.636***		−0.113***
	(0.554)		(0.022)
noteratepercent	4.855***	−0.078**	
	(0.740)	(0.028)	
bo1race2	−7.545*	0.314**	−0.124
	(3.058)	(0.114)	(0.087)
bo1race3	−1.252	0.048	−0.049
	(3.206)	(0.119)	(0.091)
bo1race4	10.529	0.037	0.682+

Figure 3: regression results

```
Residuals:
    Min      1Q  Median      3Q     Max
-66.639  -7.178   0.954   9.303  37.741

Coefficients: (1 not defined because of singularities)
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)       -10.62301    7.54128  -1.409  0.15908
lognoteamt          6.63556    0.55365  11.985  < 2e-16 ***
noteratepercent     4.85497    0.73952   6.565 6.47e-11 ***
bo1race2           -7.54544    3.05848  -2.467  0.01370 *
bo1race3           -1.25218    3.20649  -0.391  0.69620
bo1race4           10.52857   13.97287   0.754  0.45123
bo1race5           -4.76355    1.43934  -3.310  0.00095 ***
year2021           -0.06996    2.25285  -0.031  0.97523
bo1gender2          1.77288    0.88610   2.001  0.04554 *
bo1age             -0.31515    0.02127 -14.817  < 2e-16 ***
debtexpenseratio    0.29758    0.03314   8.979  < 2e-16 ***
bo1race2:year2021   9.80301    4.65489   2.106  0.03532 *
bo1race3:year2021   0.88545    4.48463   0.197  0.84350
bo1race4:year2021        NA         NA      NA       NA
bo1race5:year2021   1.98127    2.10128   0.943  0.34584
year2021:bo1gender2 -0.69886   1.28266  -0.545  0.58591
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.89 on 2194 degrees of freedom
Multiple R-squared:  0.226,    Adjusted R-squared:  0.221
F-statistic: 45.75 on 14 and 2194 DF,  p-value: < 2.2e-16
```

Figure 4: LTV regression results

```
Call:
lm(formula = lognoteamt ~ noteratepercent + LTV + bo1race * year +
    bo1gender * year + bo1age + debtexpenseratio + hsexpenseratio,
    data = df)

Residuals:
     Min       1Q   Median       3Q      Max
-1.65680 -0.35408 -0.00591  0.37239  1.65926

Coefficients: (1 not defined because of singularities)
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)        11.4138188  0.1399465  81.558  < 2e-16 ***
noteratepercent    -0.0784339  0.0277452  -2.827  0.00474 **
LTV                 0.0093461  0.0007699  12.139  < 2e-16 ***
bo1race2            0.3137623  0.1138377   2.756  0.00590 **
bo1race3            0.0484186  0.1193184   0.406  0.68493
bo1race4            0.0370061  0.5200609   0.071  0.94328
bo1race5            0.1467065  0.0536095   2.737  0.00626 **
year2021            0.0312175  0.0838709   0.372  0.70977
bo1gender2         -0.2154022  0.0327097  -6.585 5.67e-11 ***
bo1age             -0.0008992  0.0008304  -1.083  0.27896
debtexpenseratio    0.0010746  0.0014860   0.723  0.46966
hsexpenseratio      0.0071392  0.0016733   4.267 2.07e-05 ***
bo1race2:year2021  -0.2196751  0.1733272  -1.267  0.20515
bo1race3:year2021  -0.0122286  0.1669414  -0.073  0.94161
bo1race4:year2021          NA         NA      NA       NA
bo1race5:year2021  -0.0161812  0.0782071  -0.207  0.83611
year2021:bo1gender2  0.0717689  0.0477083   1.504  0.13264
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5167 on 2193 degrees of freedom
Multiple R-squared:  0.1249,  Adjusted R-squared:  0.1189
F-statistic: 20.87 on 15 and 2193 DF,  p-value: < 2.2e-16
```

Figure 5: lognoteamount regression results

```
Call:
lm(formula = noteratepercent ~ debtexpenseratio + lognoteamt *
    year + LTV + bo1race * year + bo1gender * year + bo1age,
    data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-1.10116 -0.26905 -0.02616  0.25431  1.84979

Coefficients: (1 not defined because of singularities)
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)              4.9073518  0.2671989  18.366  < 2e-16 ***
debtexpenseratio         0.0035498  0.0009580   3.705 0.000216 ***
lognoteamt              -0.1127740  0.0222035  -5.079 4.11e-07 ***
year2021                -2.7785186  0.3797667  -7.316 3.56e-13 ***
LTV                      0.0040004  0.0006021   6.644 3.85e-11 ***
bo1race2                -0.1236231  0.0872952  -1.416 0.156874
bo1race3                -0.0494412  0.0913110  -0.541 0.588246
bo1race4                 0.6816855  0.3977135   1.714 0.086668 .
bo1race5                 0.0287699  0.0411125   0.700 0.484136
bo1gender2              -0.0466421  0.0254411  -1.833 0.066888 .
bo1age                  -0.0004380  0.0006352  -0.690 0.490559
lognoteamt:year2021      0.1332211  0.0311411   4.278 1.97e-05 ***
year2021:bo1race2        0.0471251  0.1328974   0.355 0.722925
year2021:bo1race3        0.1120776  0.1276958   0.878 0.380207
year2021:bo1race4              NA         NA      NA       NA
year2021:bo1race5        0.0045826  0.0599208   0.076 0.939047
year2021:bo1gender2      0.1047463  0.0368786   2.840 0.004549 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3954 on 2193 degrees of freedom
Multiple R-squared:  0.6846,  Adjusted R-squared:  0.6825
F-statistic: 317.4 on 15 and 2193 DF,  p-value: < 2.2e-16
```

Figure 6: note rate percentage regression results