Are credit allocation models discriminatory ?

DEVYNCK Tom, DRUILHE Théo, FOUQUET Damien

Under the supervision of VANHEMS Anne and WANG Wenxuan

March 7, 2024

Outline

1. **Introduction**
2. **Econometric model and estimation method**
3. **Data**
4. **Results**
5. **Conclusion**

1. Introduction

- Lots of people aspire to become homeowners.
- When savings alone are not sufficient, one must apply for a loan.
- The bank then evaluates the application and makes its decision with an *allocation model*.

1. Introduction - Research question

Are credit allocation models discriminating?

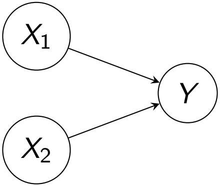

Figure: Causal relation between race, sex and the probability of deny

Where:

- X_1 is the applicant's race
- X_2 is the applicant's sex
- Y is the probability of the applicant's credit denial

1. Introduction - Economic relevance

Demographic and socioeconomic characteristics influence one's access to credit.

- Gender discrimination : Szafarz (2013), microfinance institutions (MFI) and banks in France.
- Gender discrimination : Agier and Szafarz (2011), small-business lending in Brazil.

1. Introduction - Economic relevance

- Racial differences : Bayer and al. (2017), high-cost mortgage lending in the US.
- Racial differences : Myers (1995), racial discrimination in housing markets.

2. Econometric specification - Main approach

We consider the following logistic model:

$$(\mathcal{LM}) \quad : \quad \mathbb{P}(deny_i = 1) = \frac{1}{1 + e^{X_i'\beta}}$$

where:

- $deny_i$ is a dummy variable that equals one if the credit of applicant i is denied, 0 otherwise.
- $X_i'\beta$ is a linear combination of variables of interest

We now want to choose relevant variables to represent the utility function of the bank $U_i = X_i'\beta + \varepsilon_i$

2. First (naive) model

Naively, since we want to evaluate the impact of our discriminatory variables on the probability of having its application denied, we choose :

$$U_i = \beta_0 + \beta_1 sex_i + \beta_2 race_i + \varepsilon_i$$

Plugging it in ($\mathcal{LM}$), we denote our first model ($\mathcal{LM}1$) that we will estimate with the Maximum Likelihood (ML) method.

2. Second model - Motivation

It is obvious that there is endogeneity in this model so that we need to add control variables.

According to the litterature (Hurlin et al., 2021), we now choose:

$$
\begin{aligned}
U_i = {} & [\beta_0 + \beta_1 race_i + \beta_2 sex_i] + [\beta_3 age_i + \beta_4 purpose_i + \beta_5 log(amount_i) \\
& + \beta_6 log(income_i) + \beta_7 log(property\_value_i) + \beta_8 loan\_term_i] + \varepsilon_i
\end{aligned}
$$

and plugging it in $(\mathcal{LM})$, we denote our second model $(\mathcal{LM}2)$ that we will estimate again with the ML method.

3. Data - Initial sample

- Source : Consumer Financial Protection Bureau

- Collecting method : each year, all financial institutions report mortgage data under the Home Mortgage Disclosure Act (HMDA).
- Cross-section, individual-leveled data.
- 470,000 credit applications in Michigan (USA), in 2022.
- One hundred variables containing information about the loan, the applicant and the financial institution.

Our final sample, after cleaning, contains 234,000 credit applications in Michigan (USA), in 2022 and includes the following variables:

initemize

Our target variable : *deny*. A binary variable equal to 1 if the application was denied, 0 otherwise.

Some socio-demographic variables : *sex*, *age*, *race*

Some financial variables: *loan_amount*, *income*, *property_value*, *loan_term*, *loan_to_value_ratio* and *loan_purpose*.

3. Data - Representativeness

We can consider our sample **representative** of the population.

- For the qualitative variables, we verify that the proportions between the population and the final sample are the same (for each category).
- For the quantitative variables, we verify that the median between the population and the final sample is the same.

Example:

- $Me(income) = 75,000\$$ and $Me(sample) = 76,000\$$
- 9% of black people in the population and 8% in our sample.

4. Reminder : Second model specification

Reminder we consider the following logistic model:

$$(\mathcal{LM}2) \quad : \quad \mathbb{P}(deny_i = 1) = \frac{1}{1 + e^{X_i'\beta}}$$

where:

$$U_i = [\beta_0 + \beta_1 race_i + \beta_2 sex_i] + [\beta_3 age_i + \beta_4 purpose_i + \beta_5 log(amount_i) \\ + \beta_6 log(income_i) + \beta_7 log(property\_value_i) + \beta_8 loan\_term_i] + \varepsilon_i$$

4. Second model - Results & Interpretation

	Coefficient	(Std. Error)
Intercept	8.166***	(0.146)
Asian	0.483***	(0.032)
Black	0.765***	(0.017)
Native	0.527***	(0.059)
Applicant Sex (Female)	−0.095***	(0.012)
Applicant Age < 25	−0.194***	(0.036)
Applicant Age 25-34	−0.153***	(0.019)
Applicant Age 45-54	−0.874	(0.017)
Applicant Age 55-64	−0.182***	(0.018)
Applicant Age 65-74	−0.297***	(0.022)
Applicant Age > 74	0.226***	(0.029)
Home Improvement	1.750***	(0.0223)
Other Purpose	1.890***	(0.023)
Refinancing	1.089***	(0.021)
Cash-out refinancing	1.103***	(0.018)
Loan Term	0.0023***	(7.14e-05)
log(Income)	−0.704***	(0.012)
log(Loan Amount)	0.042***	(0.011)
log(Property Value)	−0.324***	(0.014)
Observations	243,334	
Akaike Inf. Crit. (AIC)	199,455	

Note: *p<0.1; **p<0.05; ***p<0.01

4. Second model - Results & Interpretation (2)

Magnitude of coefficients in a logistic model cannot be interpreted directly (only their sign).

Thus, we decided to create an *average* candidate $\bar{x} = (\bar{x_1}, \ldots, \bar{x_7})$, defined as follows:

- If the variable is numerical, $\bar{x_k}$ is the **median** of this variable,
- If the variable is categorical, $\bar{x_k}$ is the **mode**, that is the most common value.

4. Second model - Results & Interpretation (3)

Let $\bar{x} = (\bar{x_1}, \bar{x_2}, ..., \bar{x_7})$ be the value of the *average* candidate defined previously.

Here the *average* candidate is a men aged between 35 and 44 years old that purchases a home worth $255k\$$.

He has an income of $76k\$$ and applies for a loan of $145k\$$, on 30 years.

Effect of the race:

Predicted probability of deny	**Value**
$\hat{\mathbb{P}}(deny = 1 \| race = black, x = \bar{x})$	0.145
$\hat{\mathbb{P}}(deny = 1 \| race = native, x = \bar{x})$	0.115
$\hat{\mathbb{P}}(deny = 1 \| race = asian, x = \bar{x})$	0.114
$\hat{\mathbb{P}}(deny = 1 \| race = white, x = \bar{x})$	0.073

A black person is more than **2 times more likely** to have their credit rejected, just because they are black.

4. Second model - Results & Interpretation (5)

Cross-effect of race and sex. (Case of black and white)

Predicted probability of deny	**Value**
$\hat{\mathbb{P}}(deny = 1 \mid race = black, sex = men, x = \bar{x})$	0.145
$\hat{\mathbb{P}}(deny = 1 \mid race = black, sex = women, x = \bar{x})$	0.134
$\hat{\mathbb{P}}(deny = 1 \mid race = white, sex = men, x = \bar{x})$	0.073
$\hat{\mathbb{P}}(deny = 1 \mid race = white, sex = women, x = \bar{x})$	0.067

There is **not a great effect** of sex, but women always have a slightly lower probability of deny.

4. Extension to the 50 states of the US - Method

We estimate the same model ($\mathcal{LM}2$) using data from all the other states.

This data has the same characteristics than the one from Michigan.

We measure the *racial discrimination* (towards Afro-American people) with the following ratio :

$$ratio_s = \frac{\hat{\mathbb{P}}_s(deny = 1 | race = black, x = \bar{x})}{\hat{\mathbb{P}}_s(deny = 1 | race = white, x = \bar{x})}$$

that we capture for each state s.

4. Extension to the 50 states of the US - Results

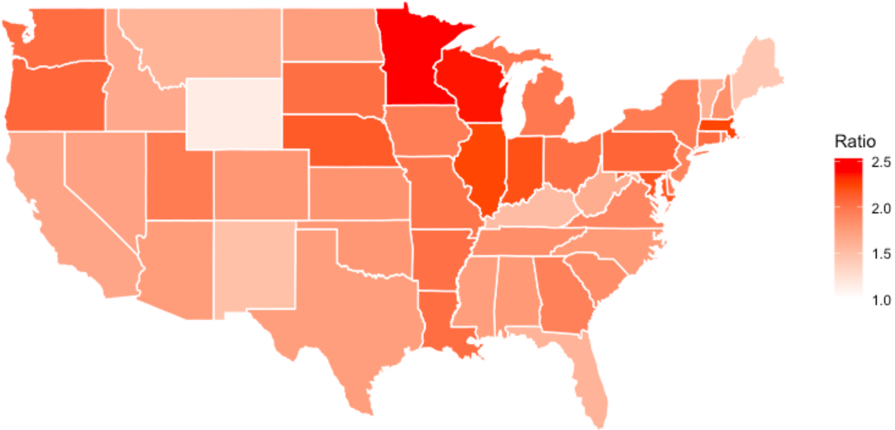

Figure: Discrimination ratio on credit-acceptance between black and white people in the US

4. Extension to the 50 states of the US - Results

According to this criterion,

- The less racist state is Wyoming with a ratio of 1.2.
- The ratio nationwide is 2.1.
- The more racist state is Minnesota with a ratio of 2.5.

5. Conclusion

- We did not show great evidence of gender discrimination, but we revealed racial discrimination in Michigan, in 2022. There, a black or an Asian person is more likely to get their credit application rejected.

- More generally, we showed that credit allocation models are discriminating in the whole United States, in 2022. **A black or Afro-american person is 2 times more likely to get their loan application rejected.**

5. Conclusion (2)

Our analysis could be extended.

- One can consider European countries.
- One can consider past or future years (even we suspect the same results for previous years).

References

1. Bayer, P., Ferreira, F., Stephen, L. (2017), "What Drives Racial and Ethnic Differences in High-Cost Mortgages? The Role of High-Risk Lenders", *The Review of Financial Studies*, Vol.31, Issue 1, p175-205.

2. Szafarz, A., Cozarenco, A. (2014), "Women's Access to Credit in France: How Microfinance Institutions Import Disparate Treatment from Banks", Aix-Marseille School of Economics Working Paper, N° 13-037.

3. Agier, I., Szafarz, A. (2011), "Credit to Women Entrepreneurs: The Curse of the Trustworthier Sex", Universite Libre de Bruxelles Working Papers, CEB: 11-005.

4. Myers, S. (1995), "Racial Discrimination in Housing Markets: Accounting for Credit Risk", *Social Science Quarterly*, Vol.76, no. 3, 543–61.

5. Hurlin, C., Pérignon, C., Saurin, S. (2021), "The Fairness of Credit Scoring Models", HEC Paris Research Paper No. FIN-2021-1411, `https://ssrn.com/abstract=3785882`