

Oral presentation 2 - Data and Main Econometric Specification

DEVYNCK Tom, DRUILHE Théo, FOUQUET Damien

February 1, 2024

Research question (reminder)

Consider a person wishing to buy a house. They often need a mortgage and thus need to apply for a loan in a bank. The bank evaluates their profile and approve or not their application thanks to a *credit allocation model*.

Research question

Do credit allocation models discriminate?

We mainly focus on the impact of the following variables on one's access to credit .

- ▷ Does *ethnicity* have an impact on a credit denial?
- ▷ Does *gender* have an impact on a credit denial?
- ▷ Do *same-sex couple* applications have an impact on a credit denial?

▷ Source :  Consumer Financial
Protection Bureau

- ▷ How they are collected : Each year, financial institutions report mortgage data under Home Mortgage Disclosure Act (HMDA)
- ▷ Sample : Focusing on 234k credit applications in Michigan, in 2022.
- ▷ Individuals data, each row corresponds to one credit application

Target variable

Deny: dummy variable that equals one if the credit is denied, 0 otherwise.

	Application accepted	Application denied
Number	193,584	40,538
Proportion	82.69%	17.31%

Table: Distribution of deny variable

Main discriminatory variable

Race: variable derive by the CFPD (Consumer Financial Protection Bureau) based on the ethnicity of applicants.

	White	Asian	Black	Native
Number	204,768	8,218	19,556	1,580
Proportion	87.46%	3.51%	8.35%	0.66%

Table: Distribution of race variable

Descriptive Statistics

Table: Descriptive Statistics about Loans

Statistic	N	Mean	St. Dev.	Min	Max
loan_to_value_ratio	233,958	73.53	21.64	0.07	149.88
loan_amount	233,958	169,722.40	124,166.10	5,000	995,000

Table: Descriptive Statistics about Income and Property Value

Statistic	N	Mean	St. Dev.	Min	Max
income	233,958	93.32	64.04	0	499
property_value	233,958	289,543.30	160,032.80	5,000	995,000

Econometric Specification

We aim to explain the variable *deny* taking the value 1 if the loan is **rejected** by the bank and 0 otherwise that is :

$$deny_i = \mathbf{1}_{\{U_i \leq 0\}}$$

where U_i is the utility function of the bank for individual i (if it is negative, the bank denies the loan and $y_i = 1$).

We assume:

- ▷ The utility function of the bank depends on financial variables but also on some socio-demographical characteristics, that is:

$$U_i = \beta_0 + \beta_1 * income_i + \beta_2 age_i + \beta_3 debt - income - ratio + \beta_4 race_i + \beta_5 sex_i + \varepsilon_i$$

- ▷ ε_i follows a logistic distribution

Econometric Specification - Logistic Model

We thus obtain a logistic model (\mathcal{LM}) where the relation between $deny_i$ and some individual's characteristics is given by:

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

where $X_i' \beta = [income_i, age_i, debt - income - ratio_i, race_i, sex_i]$

We then construct an estimator of β by the Maximum Likelihood method which is consistent and converges to a normal distribution, under some technical assumptions.

Econometric Specification - Potential issues

- ▷ Regressing separately or together the variables *sex* and *race* might give different results (statistical significance)
- ▷ Endogeneity : omitted variables might affect *deny* through the error term (e.g. savings).
- ▷ Endogeneity : *income* might be a potential endogenous variable and one might need to use instruments to solve this problem.