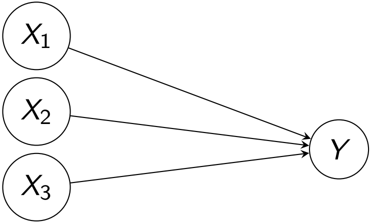
Oral presentation 3 - Econometric Results

Devynck Tom, Druilhe Théo, Fouquet Damien

March 1, 2024

Research question

Are credit allocation model discriminating ? The american experience



Where:

- X_1 is the race of the applicant
- X_2 is he sex of the applicant
- X_3 is the two applicants have the same sex
- Y is the probability of the credit being denied

Main econometric specification

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 $U_i = X_i'\beta$.

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 :

$$U_i = \beta_0 + \beta_1 sex_i + \beta_2 race_i + \beta_3 same\_sex_i + \varepsilon_i$$

Plugging it in ($\mathcal{LM}$), we denote our first model ($\mathcal{LM}1$).

We estimate ($\mathcal{LM}1$) with the Maximum Likelihood (ML) method and find the following estimates:

First model - Results

Coefficient	Estimate	p-value
(Intercept)	-1.68	$<$ 2e-16 ***
Asian	0.09	0.002 **
Black	0.904201	$<$ 2e-16 ***
Native	0.75	$<$ 2e-16 ***
Women	0.028	0.012 *
Same Sex	-0.12	0.004 **

Table: Model 1 - Results

Second model - Motivation

It is obvious that there is endogeneity in this model so that we need to add control variables (since they are contained in the error term and influences $deny_i$ through it).

We now choose:

$$U_i = [\beta_0 + \beta_1 \text{race}_i + \beta_2 \text{sex}_i + \beta_3 \text{same\_sex}_i] + [\beta_4 \text{age}_i + \beta_5 \text{purpose}_i \\ + \beta_6 \text{amount}_i + \beta_7 \text{income}_i + \beta_8 \text{property\_value}_i + \beta_9 \text{loan\_term}_i] + \varepsilon_i$$

and plugging it in $(\mathcal{LM})$, we denote our second model $(\mathcal{LM}2)$.

We estimate $(\mathcal{LM}2)$ with the ML method and find the following estimates.

Second model - Results & Interpretation

Variable	Estimate	p-value
(Intercept)	-2.532	$<$ 2e-16 ***
Asian	0.44	$<$ 2e-16 ***
Black	0.826	$<$ 2e-16 ***
Native	0.604	$<$ 2e-16 ***
Women	-0.05	1.88e-05 ***
Same Sex	2.849e-01	2.00e-11 ***
Age $<$25	-0.068	0.0542
Age $>$74	-0.114	6.95e-05 ***
Age 25-34	-0.132	8.15e-13 ***
Age 45-54	2.506e-05	0.9988
Age 55-64	-0.121	1.42e-11 ***
Age 65-74	-0.20	$<$ 2e-16 ***
Home improvement	1.742	$<$ 2e-16 ***
Other purpose	1.879	$<$ 2e-16 ***
Refinancing	1.090	$<$ 2e-16 ***
Cash-out refinancing	1.094	$<$ 2e-16 ***
Loan amount	-6.591e-09	$<$ 2e-16 ***
Income	-6.493e-06	$<$ 2e-16 ***
Property_value	-8.369e-07	$<$ 2e-16 ***
Loan term	2.279e-03	$<$ 2e-16 ***

Table: Logistic Regression Results

Second model - Results & Interpretation (2)

Coefficients of logistic models cannot be interpreted directly.

We decided to create an "average" candidate, defined as follows:

- **Mean** value for numerical variables
- **Mode**: most common value for categorical variables

Second model - Results & Interpretation (3)

$\bar{X} = (\bar{x_1}, \bar{x_2}, ..., \bar{x_6})$ value of the "average" candidate defined previously

Probability of deny	**Value**
$P(deny \mid race = black, X = \bar{X})$	0.14
$P(deny \mid race = native, X = \bar{X})$	0.115
$P(deny \mid race = asian, X = \bar{X})$	0.10
$P(deny \mid race = white, X = \bar{X})$	0.066

Interpretation:

A black person is more than **2 times more likely** to have his credit rejected, just because he is black.

Second model - Results & Interpretation (4)

Cross-effect of race and sex:

Probability of deny	**Value**
$P(deny \mid race = black, sex = men, X = \bar{X})$	0.14
$P(deny \mid race = black, sex = women, X = \bar{X})$	0.133
$P(deny \mid race = white, sex = men, X = \bar{X})$	0.066
$P(deny \mid race = white, sex = women, X = \bar{X})$	0.063

Interpretation:

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