Imperial College London

T11

DISCRETE CHOICE MODELS UNDER SOCIAL INTERACTION

Marc de la Barrera i Bardalet

Project supervisor: Dr. Auna Sivakumar

Department of Civil & Environmental Engineering Imperial College London

INTRODUCTION

Discrete Choice Models (DCMs) are a widely used tool to model how individuals make decisions. They are based on the utility of each alternative U_i, with the final choice being the alternative with a greater utility. Most of the models consider decision makers as independent, and this paper studies if this assumption holds in a context of socially connected people.

THE MODEL

When only two alternatives are considered (*smoking* vs. *not smoking*), take $y^* = U_{smoking} - U_{not\ smoking}$. The model including social influence takes the form:

$$y^* = X\beta + XW\theta + \rho W y^* + u$$
$$u = \lambda W u + \varepsilon, \qquad \varepsilon \sim N(0, \sigma^2 I)$$
$$y = \begin{cases} 1 & if\ y^* \geq 0 \\ 0 & if\ y^* < 0 \end{cases}$$

$\boldsymbol{y^*}$: **latent utility.** $\boldsymbol{P_{smoking} = P(y^* > 0) = \Phi(y^*)}$

Overall, does smoking provide a positive utility to an individual?

$\boldsymbol{X\beta}$: **private utility.**

The cost of smoking, the age of the person and their education influence their decision to smoke or not?

$\boldsymbol{XW\theta}$: **contextual effects.**

If young people tend to smoke more, does an old person surrounded by young people have an increasing chance of smoking?

$\boldsymbol{\rho W y^*}$: **endogenous effects.**

Does a person surrounded by smokers have a greater chance of smoking?

$\boldsymbol{\lambda W u}$: **correlated effects.**

If income influences smoking but is not considered in the model, do people with the same income tend to be socially connected?

$\boldsymbol{\beta, \theta, \rho, \lambda, \sigma^2}$: **parameters to estimate.**

THE WEIGHT MATRIX

All three social effects have in common the weight matrix W, which is defined as $w_{ij} > 0$ if i has a relationship with j. This sparse matrix is defined by the analyst who can use different specifications, and it is usually row normalised (rows add to one).

BOUNDARY PROBLEMS

The loss of several observations have a big impact on the resulting network as only people who took the survey can be considered for the model. In order to get a dense network, a high proportion of observations from a closed group must be obtained, making random sampling infeasible.

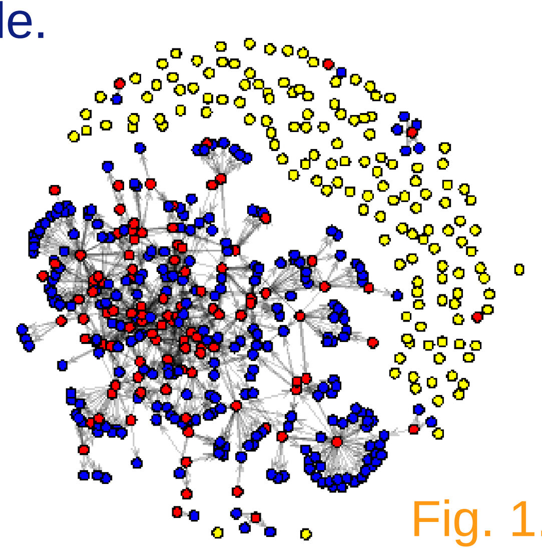

Fig. 1.a

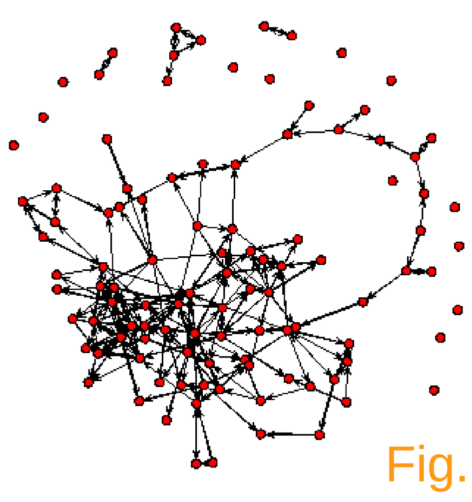

Fig. 1.b

Figure 1.a represents the information available of a network from a workplace of 500 employees. In **red**, people who took the survey and were asked to list their friends, in **blue** people mentioned by them, and in yellow people not mentioned. Figure 1.b shows the resulting useful network for the model, which is significantly less dense.

EMPIRICAL APPLICATION

In a workplace with 500 employees in the UK, 105 did a survey and were asked about their electric vehicle (EV) preferences in 9 scenarios, their personal situation and their peers in the work environment. This dataset is used to test different models, considering each of the effects independently and using different weight matrix specifications.

Table 1 shows one of the models tested, with a symmetric weight matrix $w_{ij} = \max\{w_{ij}, w_{ji}\}$ using only the autoregressive term, estimated with maximum likelihood methods and compared with a linear probit.

TABLE 1

Model formulation	$y^* = \beta X + \varepsilon$	$y^* = \beta X + \rho W y^* + \varepsilon$
Intercept	-0.351 (0.568)	0.043 (0.685)
Range (hundreds of miles)	0.316 (0.224)	0.328 (0.292)
Minimum Range Required>Range	1.461 (0.194) ***	1.468 (0.217) ***
Price (percentage)	-0.539 (0.817)	-0.594 (1.089)
Charge time (hours)	-0.066 (0.020) **	-0.065 (0.026) *
Acceleration (percentage)	3.173 (0.572) ***	3.200 (0.808) ***
Expected Fuel Cost	-0.190 (0.079) *	-0.171 (0.094) .
Expected Convenience to Recharge	0.127 (0.089)	0.129 (0.103)
Expected Overall Satisfaction	0.656 (0.126) ***	0.690 (0.159) ***
Environmental Concern	0.061 (0.020) **	0.057 (0.023) *
Experience with EV (1 if TRUE)	-0.358 (0.219)	-0.292 (0.243)
ρ	-	**0.189 (0.080) ***
LOGLIK	-155.61	-153.12

St. Error in brackets, $\varepsilon \sim N(0,1)$
Significance codes: '***' p-value=0, '**' p-value<0.01, '*' p-value <0.05, '.' p-value<0.1

MODEL DISCUSSION

Although ρ turns out to be significant, it is not able to considerably improve the model fit given its low impact on the loglikelihood function. Moreover, it is highly dependant on the data, as seen in Figure 2.

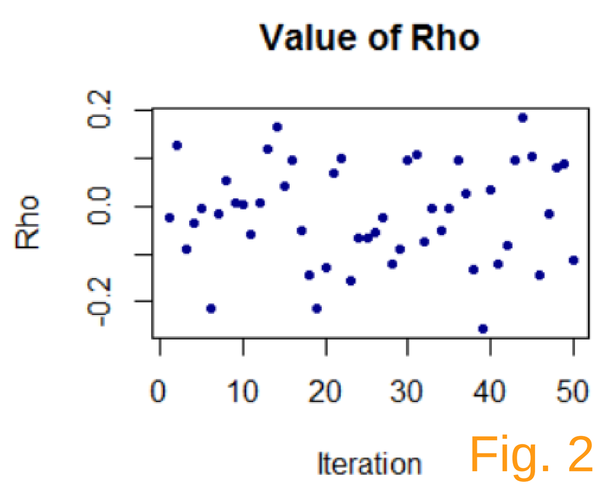

Fig. 2

To study how data dependant ρ is, 50 estimations are done omitting 10% of the observations randomly. ρ takes values from 0.18 to -0.25, with an average value of -0.01, making it impossible to state that ρ has a significant effect on the model.

POSSIBLE EXPLANATIONS

- Decision makers are indeed independent.
- Decision makers are independent when choosing a car
- Boundary problems are too critical to draw any concluding conclusion
- The dataset is too small to show significant social parameters
- The choice distribution is too unevenly distributed (only 8% for EV)

CONCLUSIONS

- A framework to study social interaction in discrete choice models is set and used for the first time.
- In the empirical application, social influence does not seem to improve the model although many reasons could explain it, not only its inexistency.
- Empirical studies must be carried to secure the theory and confirm it is taking the right direction.

ACKNOWLEDGEMENTS

This project would not have been possible without the help and advise of my supervisor Dr. Aruna Sivakumar and the data of the EV provided by John Axsen. Maristes School in Girona (Spain) also collaborated with the project coordinating a survey to its students on academic performance and exam choice.

REFERENCES

Leenders, R., (2002). Modeling social influence through network autocorrelation: constructing the weight matrix. Social Networks, 24, 21-47.

LeSage, J. et al., (2011). New Orleans business recovery in the aftermath of Hurricane Katrina. Journal of the Royal Statistical Society, 174, 1007-1027.

Manski, C., (1993). Identification of Endogeneous Social Effects: The Reflection Problem. Review of Economic Studies, 60, 531-542.