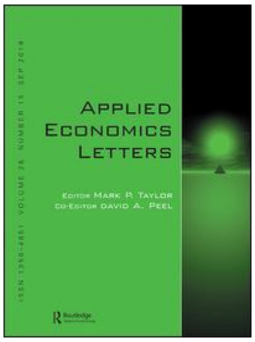

Routledge
Taylor & Francis Group

Applied Economics Letters

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rael20

A Bayesian Mixed Multinomial Logit Model for choice-sets and decision-makers' heterogeneity

Cinzia Carota & Consuelo Rubina Nava

APPLIED ECONOMICS LETTERS
https://doi.org/10.1080/13504851.2020.1854430

ARTICLE

A Bayesian Mixed Multinomial Logit Model for choice-sets and decision-makers' heterogeneity

Cinzia Carota[a] and Consuelo Rubina Nava[b]

[a]Department of Economics and Statistics "Cognetti de Martiis", Università Degli Studi di Torino, Torino, Italy; [b]Department of Economic and Political Sciences, Università Della Valle d'Aosta, Aosta, Italy; Department of Economic Policy, Università Cattolica Del Sacro Cuore, Milano, Italy

ABSTRACT
We propose a Bayesian Mixed Multinomial Logit Model (MMLM) to deal with the critical issue of choice-set heterogeneity often present in policy evaluation studies enriched with microsimulated data. We also exploit the comparison of three clustering methods to capture decision-makers' heterogeneity through a specific random effect. A case study, which aims to describe the determinants of labour choices of females in couples with microsimulated fiscal variables, is the test-bed for our methodological proposal. By virtue of this very flexible specification of the random components, the Bayesian MMLM proves to be more accurate, parsimonious and consistent in terms of point estimates with the research field than other models.

KEYWORDS
Bayesian Mixed Multinomial Logit Model labour supply heterogeneous choice-set heterogeneous decision-makers microsimulation

JEL CLASSIFICATION
C40; C11; H31; J10

I. Introduction

The description of individual choices, in the presence of a set of mutually exclusive discrete alternatives in a choice set C, and the evaluation of public policies are currently jointly treated taking advantage of random utility models (RUMs) (Train 2009). Moreover, recent contributions study public policy impacts by means of (partially) microsimulated data to anticipate and estimate the effects of socio-economic interventions as, for instance, tax and benefit reforms (see, e.g. 2). Microsimulation models combine results from observational studies, real data and expert opinions to simulate unobserved choice features under actual or hypothetical policies. Hence, microsimulation, matched with discrete choice models, has a high potential (Labeaga, Oliver, and Spadaro 2008), especially in the evaluation of public programmes.

However, as a result of the fiscal variable microsimulation and/or of features' discretization (via sampling) of available alternatives (Aaberge, Colombino, and Strøm 1999), the choice set C often does not exhibit the required homogeneity across decision-makers. This violates an important assumption underlying RUMs; nevertheless, some solutions have been proposed only for voter decisions with different sets of parties (Gallego et al. 2014) and for Logit models (Guevara and Ben-Akiva 2013).

Our contribution is, therefore, threefold in this framework. First, we relax the independence of irrelevant alternatives (IIA) assumption (Train 2009) and capture the decision-makers' heterogeneity by specifying a Mixed Multinomial Logit Model (MMLM) (McFadden and Train 2000) within a Bayesian setting to achieve high levels of estimation accuracy. Second, we compare three clustering methods to gather decision-makers' heterogeneity specifying a further random component. Third, heterogeneity of choice sets across decision-makers, due to the use of partially microsimulated data, is explicitly modelled. We test our proposal in the case study described below.

II. Data and methods

We consider a sub-sample of females in couples (2,955) observed in the Italian Survey on Household Income and Wealth (SHIW) (1998) and enriched with some microsimulated variables (gross and net wages per hour, taxes and benefits) via EUROMOD (Immervoll, O'donoghue, and Sutherland 1999), according to the 1998 Italian fiscal system (Colombino, 2015). Each female –

CONTACT Consuelo Consuelo Rubina Nava c.nava@univda.it Department of Economic and Political Sciences, Università Della Valle d'Aosta, Aosta, Italy

aged between 20 and 55, neither retired nor student – is offered the same number I of job-types (ten plus the non-working option, labelled 0). In SHIW, jobs are defined by 1 of 10 8-h intervals which map from 1 to 80 weekly working hours (wwh). To microsimulate fiscal variables, a unique integer number of wwh within the specific interval of each job-type must be selected. Here eight different choice sets are engendered including $10+1$ job-types with increasing amounts of wwh as shown in Figure 1. Hereafter, each female $j \in \{1, \dots, J\}$ and her partner in the couple are randomly assigned to a choice set h, with $h = 1, .., 8$, i.e. $C_j = C_h$[1]

To model female job preferences, available variables are partitioned in two groups: variables directly introduced in the analysis since typically employed as covariates in this literature (wwh, gross wages, age, number of children, alternatives in the choice set, taxes and benefits) (Colombino, 2015), and 'auxiliary' variables (listed in columns of Table 2) exploited to group homogeneous females by means of cluster analyses.

While in recent labour studies (Colombino 2013; Colombino, 2015) the presence of heterogeneous choice sets is ignored and the decision-makers' heterogeneity is not always explored in depth, here we explicitly address these issues by introducing a more flexible Bayesian MMLM, relying on clustering methods.

In our RUM, each female $j \in \{1, \dots, J\}$ is endowed with a random utility $U_{ij}|C_h : c_i^h \to \mathrm{IR}$

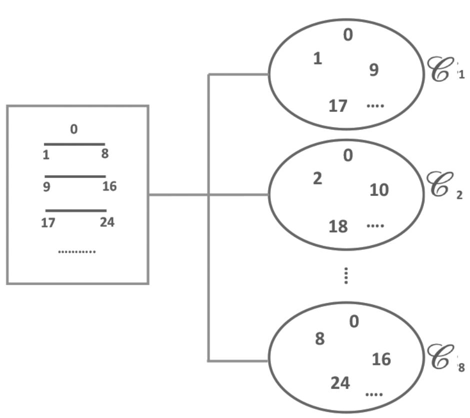

Figure 1. Construction of the eight choice sets (balls) from the discretization of the intervals of wwh (square).

for each alternative c_i^h in her choice set $C_h = \{c_1^h, c_2^h, \dots, c_I^h\}$ to reflect her job preferences. Every U_{ij} is the sum of a score, V_{ij}, assigned to each $c_i^h \in C_h$, and a zero mean stochastic component, ε_{ij}, that accommodates sources of uncertainty (Train 2009). Hence,

$$\mathbb{E}[U_{ij}|C_h] = V_{ij}|C_h = \mathbf{x}'_{ij}\beta + \mathbf{z}'_{hj}\gamma_{hj} \tag{1}$$

where $\mathbf{x}_{ij}$ is a $r \times 1$ vector of covariates, β denotes the corresponding vector of fixed effects, while $\mathbf{z}_{hj}$ represents a $s \times 1$ design vector and γ_{hj} is a vector of individual-specific and choice set–specific random effects. Errors ε_{ij} are standard Gumbel which lead to the well-known MMLM; this heavier tail assumption accommodates 'slightly more aberrant behaviour than the normal' [1, p. 39]. The resulting MMLM relaxes the IIA assumption (Train 2009) and accounts for heterogeneity (crucial for evaluating public policy impacts (Aaberge, Colombino, and Strøm 1999)) of both females and choice-sets.

Going into detail, the probability π_{ij} that job $i = 1, \dots, I$ is selected by female $j = 1, \dots, J$ can be written as

$$\pi_{ij}|h = \Pr(Y_j = i|C_j = C_h) = \frac{\exp\left\{X'_{ij}\beta + \alpha_k K_j + \delta_{hj} P_j + \eta_h C_h\right\}}{\sum_{i=1}^{I} \exp\left\{X'_{ij}\beta + \alpha_k K_j + \delta_{hj} P_j + \eta_h C_h\right\}} \tag{2}$$

where Y_j is a random variable taking values between 1 and $11(=I)$, $\mathbf{z}_{hj} = \{k_j, p_j, C_h\}$ and $\gamma_{hj} = \{\alpha_k, \delta_{hj}, \eta_h\}$, with entries of the last two vectors specified as follows.

First, we exploit individual-specific auxiliary variables to assign each female j to a cluster id $k_j \in \{1, \dots, K\}$ with an associated random effect α_k. Clusters are originated from (a) the k-means method on the Euclidean distance with a posteriori selection of the optimal number of clusters; (b) method (a) with a self-organizing map (SOM) (Kohonen 2013) grid as a preliminary step; and (c) the Chinese restaurant process (CRP) (Teh 2011) clustering method. All these methods avoid an *ex-ante* specification of the final number of clusters and have increasing ability in clustering large, complex data-sets handling different types of variables, as in our case.

[1]Number of females in $C_1, \dots, C_8$: 388, 355, 356, 359, 353, 390, 382, 372..

Table 1. Bayesian estimates for females in couples under MLM and MMLM.

	MLM	MMLM (a) *k*-means	MMLM (b) SOM	MMLM (c) CRP
c_2	−0.076	0.113	0.273**	0.09
c_3	−0.092	0.159*	0.106	0.067
c_4	−0.125	0.267***	0.217*	0.128*
c_5	−0.044	0.171**	0.134	0.096*
c_6	−0.107	0.188***	0.207**	0.149**
c_7	−0.087	0.159**	0.064	0.183***
c_8	−0.106	0.176***	0.117	0.151**
c_9	−0.055	0.203***	0.165*	0.136*
c_{10}	−0.142	0.140	0.127*	0.145***
c_{11}	−0.122	0.221**	0.164*	0.103*
Hours	0.010 ***	0.006***	0.005***	0.003***
Wage	0.008 ***	−0.004***	−0.002***	0.002***
Tax	−0.00021 ***	0.00007***	0.00005	−0.00003
Age	−0.007 ***	−0.006***	−0.006***	−0.005***
Children	−0.029 ***	−0.012***	−0.070***	−0.047***
DIC	14,090.55	13,848.14	13,759.20	13,752.12
Obs.	2,955	2,955	2,955	2,955

***, **, * and '.' indicate respectively that the 99.9%, 99%, 95% and 90% HPD intervals are bounded away from zero.

Second, we account for p_j – the job chosen by the partner of female j in C_h – and the corresponding random effect δ_{hj} in order to model couple heterogeneity.

Third, we cope with choice-set heterogeneity by including the random effect η_h related to the C_h assigned to female j and her partner.

Moreover, we adopt a Bayesian approach to inference, embedding Equation 2 in the following hierarchy:

$$\begin{aligned} &Y_j|h \sim \text{Multinom}(1, \pi_j) \quad j = 1, \ldots, n \\ &\text{logit}(\pi_{ij}) = \mathbf{x}_{ij}'\beta + \mathbf{z}_{hj}'\gamma_{hj} \\ &\beta \sim \text{N}(\mu_\beta, V_\beta), \quad \gamma_{hj} \sim \text{N}(0, V_\gamma), \quad V_\gamma \sim \text{IW}(\Psi, \nu), \end{aligned} \tag{3}$$

where $\pi_j = \{\pi_{1j}, \ldots, \pi_{Ij}\}$ is the vector of 'success' probabilities for each alternative in C_j, β and γ_{hj} are assigned vague Gaussian priors, and, finally, the random effect (co)variance matrix V_γ is taken to be an inverse-Wishart (IW).

III. Results

The analysis under Model (3) has been conducted in R. A preliminary Chi-square test confirms differences in the I job-type frequencies, justifying the use of a Multinomial model.

Cluster id k_js are obtained, given 'auxiliary' variables, as follows. Method (a) identifies, with the analysis of 30 indices (package NbClust), two distinct clusters as well as method (b), which is based on the construction of a5 $\times$ 5 SOM grid (package kohonen), while, method (c) recognizes a more satisfactory partition of the sample composed of four groups (package CRPClustering). Cluster peculiarities are shown in Table 2.

A Markov Chain Monte Carlo method with a block Gibbs sampling algorithm is implemented to estimate Bayesian model coefficients: we run 44,000 iterations, with a burn-in of 4,000 and a thinning interval of 10 (package MCMCglmm). Hyperparameters are: $\mu_\beta = 0$, $V_\beta = \mathbf{I}_4 \cdot 10^{10}$, where $\mathbf{I}_r$ is a $r \times r$ identity matrix. The residual covariance matrix was $\frac{1}{I \cdot 10^2} \cdot (\mathbf{I}_{I-1} + \mathbf{U})$, where $\mathbf{U}$ is a $(I-1) \times (I-1)$ unit matrix, while Ψ is $\mathbf{I}_{I-1}$.

Point estimates under Model (3), compared to the ones under a Multinomial Logit model (MLM) (McFadden 1974), are shown in Table 1, juxtaposing different clustering methods to absorb female heterogeneity.

MMLMs always over-perform the MLM (according to DIC) and show a larger number of highest posterior density (HPD) intervals bounded away from zero. In particular, the MMLM with CRP captures the best female partition. The CRP, indeed, simultaneously clusters females and defines the number K of groups (treated as an unbounded random variable) non-parametrically. The CRP is a distribution over cluster assignments that parsimoniously identifies groups, i.e. heterogeneity, *a posteriori* and in a more satisfactory way than the other methods.

Therefore, for the majority of selected covariates, the MMLM with CRP leads to point

Table 2. Mean values of the (scaled) auxiliary variables, given the three clustering methods, for each emerging group.

			Auxiliary variables								
Method	Cluster	Obs.	Years of education	Taxes not job related	Years of employment	Other non-working earnings (female)	Other non-working earnings (male)	Children aged 0–5	Children aged 6–10	Children aged 11–17	Social insurance
(a) *k*-means	1	90	1,076	3,587	1,694	3,320	1,848	0,050	0,012	0,053	3,160
	2	2,865	−0,034	−0,113	−0,053	−0,104	−0,058	−0,002	0,000	−0,002	−0,099
(b) SOM	1	207	0,434	2,569	0,645	2,574	−0,023	−0,241	−0,293	0,159	2,538
	2	2,748	−0,033	−0,194	−0,049	−0,194	0,002	0,018	0,022	−0,012	−0,191
(c) CRP	1	550	−0,069	−0,284	−0,219	−0,341	−0,134	0,101	1,211	0,079	−0,304
	2	1,482	−0,137	−0,235	−0,208	−0,277	−0,085	−0,008	−0,588	0,039	−0,241
	3	238	0,719	1,596	1,183	1,628	1,204	−0,010	−0,172	−0,039	1,287
	4	685	0,101	0,181	0,215	0,307	−0,127	−0,059	0,360	−0,134	0,317

estimates consistent with the economic theory (e.g. wage and tax), thus improving the interpretability of the results. In Table 1, a negative tax effect explains female reactions to fiscal reforms, that is by reducing wwh for increments in the fiscal pressure. Analogously, an increasing number of children requires more free time. This model also describes preferences among job-types through positive, very often statistically significant, marginal effects compared to the non-working alternative c_1 (see, e.g. c_7 and c_{10}).

IV. Conclusions

In the test-bed for our methodological proposal, we consider only female in couples to avoid both identification and quasi-complete separation issues, and we limit multicollinearity (differently from Colombino, 2015) by suitably selecting covariates. Our Bayesian MMLM specifications are more accurate and parsimonious than the ones proposed in the literature (Colombino, 2015), given their ability to simultaneously capture decision-makers' and choice-set heterogeneity. To the best of our knowledge, this is the first time that, (not only) in this framework, heterogeneity has been indirectly captured via clustering to obtain more parsimonious and accurate models, crucial for policy evaluation and simulations. Here, the CRP clustering technique leads to the best model which guarantees higher precision when comparing different tax reform scenarios and when computing labour supply elasticity. Finally, our methodological proposal can be directly adapted to different economic applications involving multiple sources of heterogeneity.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Consuelo Rubina Nava http://orcid.org/0000-0002-8046-8185

References

Aaberge, R., U. Colombino, and S. Strøm. 1999. “Labour Supply in Italy: An Empirical Analysis of Joint Household Decisions, with Taxes and Quantity Constraints.” *Journal of Applied Econometrics* 14 (4): 403–422. doi:10.1002/(SICI)1099-1255(199907/08)14:4<403::AID-JAE520>3.0.CO;2-4.

Colombino, U. 2013. “A New Equilibrium Simulation Procedure with Discrete Choice Models.” *International Journal of Microsimulation* 6 (3): 25–49. doi:10.34196/ijm.00087.

Colombino, U. 2015. “Five Crossroads on the Way to Basic Income: An Italian Tour.” *Italian Economic Journal* 1 (3): 353–389. doi:10.1007/s40797-015-0018-3.

Gallego, M., N. Schofield, K. McAlister, and J. S. Jeon. 2014. “The Variable Choice Set Logit Model Applied to the 2004 Canadian Election.” *Public Choice* 158 (3–4): 427–463. doi:10.1007/s11127-013-0109-3.

Guevara, C. A., and M. E. Ben-Akiva. 2013. “Sampling of Alternatives in Logit Mixture Models.” *Transportation Research Part B: Methodological* 58: 185–198. doi:10.1016/j.trb.2013.08.011.

Immervoll, H., C. O'donoghue, and H. Sutherland. 1999. “An introduction to EUROMOD.” *EUROMOD Working Paper* No. EM0/99, Microsimulation Unit, Department of Applied Economics, University of Cambridge.

Kohonen, T. 2013. “Essentials of the Self-organizing Map.” *Neural Networks* 37: 52–65. doi:10.1016/j.neunet.2012.09.018.

Labeaga, J. M., X. Oliver, and A. Spadaro. 2008. “Discrete Choice Models of Labour Supply, Behavioural Microsimulation and the Spanish Tax Reforms.” *The Journal of Economic Inequality* 6 (3): 247–273. doi:10.1007/s10888-007-9057-9.

McFadden, D., and P. Zarembka. 1974. “;Frontiers in econometrics.” *Conditional logit analysis of qualitative choice behavior*, 105-142.

McFadden, D., and K. Train. 2000. “Mixed Mnl Models for Discrete Response.” *Journal of Applied Econometrics* 15 (5): 447–470. doi:10.1002/1099-1255(200009/10)15:5<447::AID-JAE570>3.0.CO;2-1.

Teh, Y. W. 2011. “Dirichlet process”. In *Encyclopedia of machine learning*. Springer. 280–287.

Train, K. E. 2009. *Discrete Choice Methods with Simulation*. Cambridge University Press.