ARTICLE IN PRESS

Journal of Choice Modelling xxx (2017) 1–17

Contents lists available at ScienceDirect

Journal of Choice Modelling

journal homepage: www.elsevier.com/locate/jocm

Comparison of parametric and semiparametric representations of unobserved preference heterogeneity in logit models

Prateek Bansal [a,*], Ricardo A. Daziano [b], Martin Achtnicht [c]

[a] School of Civil and Environmental Engineering, Cornell University, 301 Hollister Hall, Ithaca, NY 14853, USA
[b] School of Civil and Environmental Engineering, Cornell University, 305 Hollister Hall, Ithaca, NY 14853, USA
[c] Leibniz Institute of Ecological Urban and Regional Development (IOER), Dresden and Centre for European Economic Research (ZEW), Mannheim, Germany

ARTICLE INFO

Keywords:
Mixed logit
Semiparametric
Unobserved taste heterogeneity
Flexible mixing

ABSTRACT

The logit-mixed logit (LML) model is a very recent advancement in semiparametric discrete choice models. LML represents the mixing distribution of a logit kernel as a sieve function (polynomials, step functions, and splines, among many other variants). In the first part of this paper, we conduct Monte-Carlo studies to analyze the number of required parameters (e.g., polynomial order) in three LML variants to recover the true population distributions, and also compare the performance (in terms of accuracy, precision, estimation time, and model fit) of LML and a mixed multinomial logit with normal heterogeneity (MMNL-N). Our results indicate that adding too many parameters in LML may not be the best strategy to retrieve underlying taste heterogeneity; in fact, over-specified models generally perform worst in terms of BIC. We recommend to use neither minimum-BIC nor the most flexible specification, but we rather suggest to start with the same number of parameters as a parametric model (such as MMNL-N) while checking changes in the derived histogram of the mixing distribution. As expected, LML was able to recover bimodal-normal, lognormal, and uniform distributions much better than the misspecified MMNL-N. Computational efficiency makes LML advantageous in the process of searching for the final specification. In the second part of the paper, we estimate the willingness-to-pay (WTP) estimates of German consumers for different vehicle attributes when making alternative-fuel-car purchase choices. LML was able to capture the bimodal nature of WTP for vehicle attributes, which was not possible to retrieve using standard parametric specifications.

1. Introduction: random preference heterogeneity in choice modeling

In random utility maximization-based discrete choice modeling, the multinomial or conditional logit (MNL) model (McFadden, 1973) has been widely used, but cannot handle unobserved differences in preferences across decision makers. In the past two decades, researchers have realized the importance of incorporating random taste heterogeneity in many practical situations, including the valuation of travel time savings that vary across commuters. MNL has been subsequently extended to random parameter logit models, such as the mixed multinomial logit (MMNL) model (McFadden and Train, 2000) that assumes continuous parametric heterogeneity distributions. In addition to MMNL, the literature offers several parametric and semiparametric logit-type models to specify random taste heterogeneity of the consumers. However, there is no agreement among researchers in terms of choosing any specific model (or

* Corresponding author.
E-mail addresses: pb422@cornell.edu (P. Bansal), daziano@cornell.edu (R.A. Daziano), m.achtnicht@ioer.de (M. Achtnicht).

https://doi.org/10.1016/j.jocm.2017.10.002
Received 14 May 2017; Received in revised form 27 September 2017; Accepted 24 October 2017
Available online xxxx
1755-5345/© 2017 Elsevier Ltd. All rights reserved.

Please cite this article in press as: Bansal, P., et al., Comparison of parametric and semiparametric representations of unobserved preference heterogeneity in logit models, Journal of Choice Modelling (2017), https://doi.org/10.1016/j.jocm.2017.10.002

mixing distribution).

Keane and Wasi (2013) and Fosgerau and Hess (2007) are two seminal papes that have compared different random parameter logit models. Whereas Keane and Wasi (2013) used data obtained from 10 stated-preference (SP) discrete choice experiments with a focus on identifying the best parametric random parameter logit model in terms of both data fit –basically using the Bayes information criterion (BIC)– and capturing specific behavioral patterns, Fosgerau and Hess (2007) compared semiparametric models with parametric models using SP data and Monte-Carlo studies to explore the best strategy to retrieve the true random heterogeneity in the population. The findings of these studies are discussed below, in addition to other studies.

In terms of parametric heterogeneity distributions, MMNL with normally distributed random parameters (MMNL-N) is the most commonly used specification in research.[1] However, the normal distribution may be too restrictive for some practical situations and may create misspecification issues.[2] Additionally, when Keane and Wasi (2013) compared MMNL-N with other parametric models[3] across 10 SP datasets, the authors never found MMNL-N to be preferable in terms of BIC. MMNL-N also did worse relative to other parametric models in capturing extreme consumer behavior (such as lexicographic behavior, when consumer choices are mainly determined by a single attribute of the alternatives).

Whereas MMNL-N does not appear to be a universally appropriate choice of mixing distribution and there is no way to know the mixing distribution before estimation, a few studies (Bajari et al., 2007; Fosgerau and Bierlaire, 2007; Train, 2008; Fox et al., 2011; Bastin et al., 2010; Fosgerau and Mabit, 2013) have specified semiparametric logit models, which consider more flexible and non-parametric heterogeneity distributions. These models are generally computationally efficient and easier to implement when compared to parametric models. A quick review of these models can be found in a companion paper (Bansal et al., 2017). Fosgerau and Hess (2007) compared a Legendre polynomial-based semiparametric logit model (Fosgerau and Bierlaire, 2007) with other parametric logit models and found it the best in terms of retrieving the true distribution of the random parameters across case studies (with true distributions ranging from uniform to multimodal). Such finding is expected because of the flexibility of specifying a higher number of parameters in the semiparametric approaches. Fosgerau and Hess (2007) concludes with a different perspective on the advantage of semiparametric approaches: by allowing a higher number of parameters, semiparametric models can be used as initial diagnostic tool to identify the underlying heterogeneity distribution, and parametric logit models with fewer parameters could be subsequently estimated for inference and prediction.

Train (2016) recently proposed a semiparametric Logit-Mixed Logit (LML) model (see Section 2.2), which generalizes many previous parametric and semiparametric logit models (see Sections 2.2.1 and 2.2.2 for details). As the name suggests, this model contains two logit formulations: one for the decision maker's probability to choose an alternative and another for the probability of selecting a given parameter value from a finite parameter space. The actual shape of the logarithm of the mixing distribution can be defined by different type of functions such as polynomials (see Section 2.2.1), step functions (see Section 2.2.2), and splines (see Section 2.2.3), among many others.

Since LML provides a generalized framework for semiparametric logit models, the required number of parameters (i.e., 'order' of polynomial, 'levels' in step function, and 'knots' in spline) to retrieve specific heterogeneity distributions is worth exploring. Thus, in the first part of this paper, we conduct Monte-Carlo studies to analyze the required number of LML parameters to recover different shapes of random taste heterogeneity (see Section 3). In addition, model fit – BIC, estimation time, and random heterogeneity retrieval (in terms of finite sample bias and probability distribution functions of the random parameters) of the LML model are also compared with MMNL-N. In the context of LML, we also investigate the observation of Fosgerau and Hess (2007), which suggests that a higher number of parameters yields a better approximation of the true distribution.[4] In the second part of this paper, we analyze purchase preferences of German consumers for alternative-fuel vehicles using MMNL-N and differing LML specifications. The objective of this empirical application is to explore the implications of alternative LML specifications (with varying number of parameters) on the estimates of willingness to pay (WTP) for various vehicle attributes. Since the true WTP distribution is unknown, we compare the WTP estimates (and probability density functions) of LML with MMNL-N and explicitly state the benefits of using LML over MMNL-N.

The remaining paper is organized as follows: Section 2 discusses mathematical details of MMNL-N and LML; Section 3 lays out the Monte-Carlo study design, and draws insights about performance of different logit models under different types of random taste heterogeneity; Section 4 focuses on the implications of using LML specifications over MMNL-N in estimating stated purchase preferences for alternative-fuel vehicles; and Section 5 concludes with practical recommendations.

2. Mixed multinomial logit (MMNL) and logit-mixed logit (LML) models

As stated in the introduction, MMNL dominates research in random parameter logit models. In MMNL, the indirect utility derived by decision-maker i from choosing alternative j in choice situation t is:

[1] A few studies have also used other distributions such as lognormal, Johnson's S_b, gamma, and triangular.

[2] For example, the marginal utility of price components has to be negative, by microeconomic principles. However, a normal distribution for price parameters may misleadingly yield positive estimates.

[3] These other parametric models include: generalized MNL (see Fiebig et al., 2010), theory-constraint MMNL (e.g., lognormal distribution of a price parameter), and Mixed-Mixed MNL (MM-MNL, Burda et al., 2008; Rossi et al., 2012).

[4] The loglikelihood of the model with a higher number of parameters may be higher, but it is worth exploring whether model specifications with more flexible distributions are preferable in terms of BIC, for instance.

Download English Version:

https://daneshyari.com/en/article/7356824

Download Persian Version:

https://daneshyari.com/article/7356824

Daneshyari.com