Journal of Choice Modelling 28 (2018) 41–55

Contents lists available at ScienceDirect

Journal of Choice Modelling

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

# Estimation of factor structured covariance mixed logit models

Jonathan James

*Department of Economics, California Polytechnic State University, USA*

ARTICLE INFO

*JEL classification codes:*
C02
C13
C25
C35
C38

*Keywords:*
Discrete choice
Mixed logit
EM algorithm
Factor models

ABSTRACT

Mixed logit models with normally distributed random coefficients are typically estimated under the extreme assumptions that either the random coefficients are completely independent or fully correlated. A factor structured covariance offers a range of alternatives between these two assumptions. However, because these models are more difficult to estimate they are not frequently used to model preference heterogeneity. This paper develops a simple expectation-maximization algorithm for estimating mixed logit models when preferences are generated from a factor structured covariance. The algorithm is easy to implement for both exploratory and confirmatory factor models. The estimator is applied to stated-preference survey data from residential energy customers (Train, 2007). Comparing the fit across five different models, which differed in their assumptions on the covariance of preferences, the results show that all three factor specifications produced a better fit of the data than the fully correlated model measured by BIC and two out of three performed better in terms of AIC.

## 1. Introduction

The mixed logit with normally distributed random coefficients is one of the most widely used specifications of random utility models. Researchers typically estimate these models under the extreme assumptions that either the random coefficients are completely independent or fully correlated. While the fully correlated model nests the independence model, the number of parameters in the fully correlated model increases exponentially with the number of random coefficients, causing researchers in some situations to prefer the independence model because of computation time or because the model with fewer parameters performs better on model selection criteria like AIC or BIC. For example, Keane and Wasi (2013) estimate mixed logit models with normally distributed random coefficients on ten different stated-preference data sets and found that the uncorrelated specification performed better than the correlated specification using BIC in the majority of the data sets.

An alternative approach for modeling correlation in preferences that is less demanding of the data than the full covariance model is a factor structured covariance. Factor models are used extensively in the social sciences to study correlated random variables and have been applied to modeling preference heterogeneity in discrete choice models in a number of previous studies. In a factor structured covariance model, individual preferences over product attributes are a function of a low-dimensional number of latent factors, which generates correlation in individual preferences with fewer parameters. In random utility models, the factor-analytic logit model, for example, Elrod (1988), Chintagunta (1994), Goettler and Shachar (2001), and Wen et al. (2014), uses a factor structure to model correlation in preferences over *unobserved* product attributes. Factor-analytic logits have been further discussed in Ben-Akiva and Bolduc (2001), Ben-Akiva and Bierlaire (1999) and Dube et al. (2002). In addition, factor structured covariances have been used to model preference heterogeneity over *observed* product attributes in Elrod and Keane (1995), Keane and Wasi (2012) and Fiebig et al. (2010).

*E-mail address:* jjames04@calpoly.edu.
*URL:* https://www.calpoly.edu/~jjames04.

https://doi.org/10.1016/j.jocm.2018.05.006
Received 19 December 2017; Received in revised form 24 April 2018; Accepted 25 May 2018
1755-5345/ 

The factor approach offers three main benefits over the full covariance model. First, factor covariance models have substantially fewer parameters than models with a full covariance matrix. For example, with 10 product characteristics, a full covariance matrix contains 55 parameters while the covariance with a single factor is represented with only 20 parameters. In many cases, this more parsimonious representation of the preference distribution can yield more precise estimates of the parameters and superior performance in terms of AIC and BIC, providing higher confidence that the estimates are closer to the true model. In sum, because it has fewer parameters the factor approach is less demanding of the data which may lead to a faster and overall better fitting model. The second benefit of the factor structured covariance is that it provides a more parsimonious decomposition of preferences, where a large number of preferences over product attributes can be described by a handful of factors, which helps reveal the main features of the data. For example, as suggested by Elrod (1988), the loadings from the factor-analytic logit model can be used to place unobserved brand attributes for a very large number of brands on a much lower, typically two to four, dimensional choice map, which helps the researcher better understand how different brands are positioned in the product space.[1] The final benefit of the factor structured covariance is that it is extremely flexible. Not only can researchers choose the number of factors to consider with exploratory factor models, but they can also test hypothesis or consider model driven or theory based restrictions on the covariance through confirmatory factor models.

Despite the benefits, factor structured covariances are not frequently used in mixed logit models in part due to their computational complexity. Like all mixed logit models, estimating a model with a factor structured covariance requires the maximization of an integrated likelihood function. However, maximization of the log-likelihood with a factor structured covariance is extremely slow with quasi-Newton methods because the gradient of the log-likelihood is difficult to take analytically, and optimization must rely on much slower numerical gradients.[2] The main issue is that in the fully correlated model there exists a mapping between the model parameters and the Cholesky components of the preference covariance matrix, facilitating a reformulation of the likelihood in terms of the Cholesky components where analytical gradients can be taken over these parameters. In the case of the factor structured covariance there is no mapping between the model parameters and the Cholesky components of the preference covariance matrix, so numerical gradients are typically the only option.

The contribution of this paper is to develop a simple estimator for factor structured covariance mixed logit models that can be used to estimate both exploratory and confirmatory factor models. The estimator is an extension of the expectation-maximization (EM) algorithms developed in (Train, 2007) and Train (2008). These papers show that mixed logit models can be easily estimated by taking draws from the mixing distribution and then updating the moments of the mixing distribution by computing weighted averages of the draws. Repeating this simple process yields maximum likelihood estimates of the model parameters. Given that the EM algorithm is widely used to estimate factor models (Rubin and Thayer, 1982), the algorithm in Train (2007) can be easily modified to allow for factor structured covariances. The algorithm easily generalizes to any factor structure without the need to derive or code problem specific gradients.

The algorithm is applied to the stated-preference survey on residential energy customers in (Train, 2007). In this data, customers are asked to make choices when faced with different energy pricing options and providers. While the fully correlated model performs better than the independence model, both a 1-factor and 2-factor model perform better in terms of BIC than the fully correlated model, providing evidence that correlation in preferences may be a function of a lower dimensional number of factors. To gain further insight into the structure of preferences, a confirmatory factor analysis is performed to test the hypothesis that consumer preference over the six attributes in the data set are a function of only two latent variables: a preference over the price components and a preference over the supplier components. The results show that this theory based specification outperforms all of the other models in part because it explains nearly the same amount of variation in the data with fewer parameters. In terms of computational savings, the proposed EM algorithm for factor structured covariances was between 13 and 25 times faster than quasi-Newton methods with numerical gradients.

The remainder of this paper is organized as follows. Section 2 describes the mixed logit model with a factor structured covariance matrix. Section 3 discusses an example of a factor structured covariance using the model in Train (2007) and discusses the shortcomings of current estimation methods. Section 4 outlines the EM algorithm to estimate the model. Section 5 studies the performance of the algorithm with stated-preference data. Finally, Section 6 concludes.

## 2. Factor structured covariance mixed logit model

In a mixed logit model, if consumer $i$ chooses product $j$ in choice situation $t$ they obtain utility $U_{ijt} = x'_{ijt}\beta_i + \varepsilon_{ijt}$. Where $x_{ijt}$ is the observed characteristics of product $j$, $\beta_i$ is individual $i$'s preferences over the product characteristics, and $\varepsilon$ is an i.i.d. random utility shock. Individual preferences are unobserved but are drawn from some known distribution $f(\cdot)$ that is parameterized by unknown $\Psi$, i.e., $\beta_i \sim f(\beta|\Psi)$. The goal of estimation is to recover maximum likelihood estimates of the parameters of the mixing distribution $\hat{\Psi}$. Since preferences are not observed, the likelihood of the observed data is written conditional on a given value of preferences and then integrated over the full distribution using $f(\cdot)$. Assuming there are $J$ products to choose from and assuming $\varepsilon$ is distributed type-I extreme value, the probability that $i$ chooses $j$ in choice situation $t$ given preferences $\beta$ has the familiar logit expression,

[1] While this type of data reduction could be performed on the full covariance matrix after estimation, it is more efficient to estimate the factor structure directly from the discrete choice data.

[2] In their influential paper, Fiebig et al. (2010) kindly provide MATLAB code for a number of mixed logit specifications with alternative covariance structures, which includes a factor analytic structure that relies on numerical gradients.