Journal of Choice Modelling 23 (2017) 1–8

Contents lists available at ScienceDirect

Journal of Choice Modelling

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

Correlation and scale in mixed logit models

Stephane Hess[a,*], Kenneth Train[b]

[a] *Institute for Transport Studies & Choice Modelling Centre, University of Leeds, UK*
[b] *Department of Economics, University of California at Berkeley, United States*

ARTICLE INFO

Keywords:
Mixed logit
Correlation
Scale heterogeneity

ABSTRACT

This paper examines sources of correlation among utility coefficients in models allowing for random heterogeneity, including correlation that is induced by random scale heterogeneity. We distinguish the capabilities and limitations of various models, including mixed logit, generalized multinomial logit (G-MNL), latent class, and scale-adjusted latent class. We demonstrate that (i) mixed logit allows for all forms of correlation, including scale heterogeneity, (ii) G-MNL is a restricted form of mixed logit that, with an appropriate implementation, can allow for scale heterogeneity but (in its typical form) not other sources of correlation, (iii) none of the models disentangles scale heterogeneity from other sources of correlation, and (iv) models that assume that the only source of correlation is scale heterogeneity necessarily capture, in the estimated scale parameter, whatever other sources of correlation exist.

1. Introduction

Scale heterogeneity has become a widely discussed topic in recent years (see e.g. Swait and Bernardino, 2000; Fiebig et al., 2010). It is defined as variation across individual decision-makers in the impact of factors that are not included in the model, relative to the impact of factors that *are* included. Decision-makers whose choices are greatly affected by factors that are outside of the model have relatively small coefficients, in magnitude, for the variables that are in the model; while people who are little affected by unincluded factors have larger coefficients, in magnitude, for included factors. Scale heterogeneity is a form of correlation among utility coefficients, by which the coefficients of all included variables (including alternative specific constants) are larger in magnitude for some people than others.

Several model specifications have recently been proposed with the goal of estimating scale heterogeneity, and numerous published papers claim to have done so in empirical applications.[1] However, as highlighted by Hess and Rose (2012), scale heterogeneity is not identified separately from other sources of heterogeneity, which means - unfortunately - that these claims are incorrect and the goal itself is misguided. The current paper clarifies the issue of scale within mixed logit models and distinguishes the capabilities and limitations of different specifications. These concepts can be used by researchers to specify and interpret their models within the necessary constraint of identification.

Random coefficients models allow for variation in parameters across individual decision-makers, which raises the possibility of correlation among the individual parameters. Different models handle this correlation in different ways, and we use this distinction to explain the role of scale heterogeneity in each model. We start by discussing the various sources of correlation in mixed logit models, including scale heterogeneity as one of these sources. We differentiate several models that have been introduced to address heterogeneity with respect to how they handle correlations. We point out that mixed logit models with full correlation among utility coefficients allow for all sources of correlation, including scale heterogeneity. However, models that are designed for scale heterogeneity alone, such as most implementations of the "generalized multinomial logit" model, are restricted forms of mixed logit that contain only one correlation

* Corresponding author.
E-mail addresses: s.hess@its.leeds.ac.uk (S. Hess), train@econ.berkeley.edu (K. Train).

[1] We discuss these models in the sections below and give examples of the claims in the Appendix.

http://dx.doi.org/10.1016/j.jocm.2017.03.001

parameter. The estimate of the correlation parameter in these models (also called the scale parameter) captures whatever sources of correlation exist in the data, and cannot be interpreted as representing only scale heterogeneity.

We expand on these concepts below. We first give notation for the mixed logit model. We then discuss the role of correlation in general, and scale heterogeneity as a form of correlation. Several models are compared next, including scaled multinomial logit (S-MNL), generalized multinomial logit (G-MNL), models in willingness-to-pay (WTP) space, latent class, and scale-adjusted latent class (SALC) models. In addition to interpreting these models, we provide practical guidance for model specification in applied work.

2. Mixed logit

Let the utility that person n obtains from alternative j in choice situation t be denoted in the usual way as

$$U_{njt} = \beta_n' x_{njt} + \varepsilon_{njt} \tag{1}$$

where x_{njt} is a vector of observed attributes, β_n is a corresponding vector of utility coefficients that vary randomly over people, and ε_{njt} is a random term that represents the unobserved component of utility. The vector x_{njt} can include 0/1 terms to allow for alternative-specific constants and for individual attribute levels, as well as continuous attributes.

The unobserved term ε_{njt} is assumed to be *iid* extreme value. Under this assumption, the probability that person n chooses alternative i in choice situation t, conditional on β_n, is the logit formula:

$$L_{nit}(\beta_n) = \frac{e^{\beta_n' x_{nit}}}{\sum_j e^{\beta_n' x_{njt}}} \tag{2}$$

The researcher does not observe the utility coefficients of each person and knows that the coefficients vary over people. The cumulative distribution function of β_n in the population is $F(\beta|\theta)$ which depends on parameters θ. The distribution can be continuous or discrete, different elements in β may follow different distributions (including some being fixed), and the elements of β may be correlated with each other.

With continuous F, the choice probability for the person's sequence of choices, given the researcher's information, is:

$$P_{nit} = \int L_{nit}(\beta) f(\beta|\theta) d\beta \tag{3}$$

where f is the density associated with F.

If F is discrete, then the mixed logit formula is

$$P_{nit} = \sum_{r\in S} L_{nit}(\beta_r)\pi_r(\beta_r|\theta) \tag{4}$$

where π is the probability mass function associated with F, and S is its support set with elements indexed by r. The goal of the researcher is to specify F and estimate its parameters θ.

McFadden and Train (2000) have shown that any choice model, with any distribution of preferences, can be approximated to any degree of accuracy by a mixed logit. This result implies that the mixed logit model does not embody any theoretical restrictions on the choice model or distribution of preferences. In any application, the researcher needs to specify F, and the researcher's choice for F might, and usually does, embody restrictions. This paper focuses on the restrictions on correlations that are implied by the researcher's specification of F.

3. Correlation

Correlations among utility coefficients can arise for many reasons, depending on the application. For example:

1. Energy-efficiency programs offer incentives, such as rebates and financing, for purchases of high-efficiency appliances. Consumers who respond greatly to rebates tend also to respond greatly to attractive financing, such that the rebate and financing coefficients are positively correlated (Revelt and Train, 1998).
2. In choice of fishing site, anglers who place a higher-than-average value on the fish stock at the site also tend to place a higher-than-average value on the aesthetic quality of the site, such that the coefficients of these two measures of quality are positively correlated (Train, 1998).
3. In choice among Alpine hiking sites, recreators who value warming huts at the site tend also to prefer sites with easier trails; and people who prefer difficult trails also tend to like having rope assists on the trails (Scarpa et al., 2008).
4. In travellers choices of route by car and public transport, Hess et al. (forthcoming) find complex correlation patterns between the sensitivities to the different time, cost, quality of service and safety attributes. Some correlations are positive while others are negative.

Correlations such as these can be expected in any setting: they simply reflect that a consumers' preferences for one attribute are related to their preferences for another attribute.

Scale heterogeneity constitutes a specific type of correlation among utility coefficients. In empirical analysis, there are some factors that affect people's choices but are not included in the researcher's model, perhaps because the researcher is unable to observe or measure them. The impact of these unincluded factors on people's choices can differ over people: some people might be more affected by the unincluded factors than other people, such that their choices appear more *random* from the perspective of the researcher.

This difference in people's reaction to unincluded factors creates correlation among the coefficients of the included variables. In particular, if a person's choices are determined primarily by unincluded factors, then their choices are *not* affected so much by