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 included factors. These people have utility coefficients that are small in magnitude, reflecting the small impact of the included factors relative to unincluded factors. Conversely, if a person's choices are mainly determined by included factors, with unincluded factors having little influence, then their utility coefficients are large in magnitude, reflecting the relatively large impact of the included factors. The coefficients become correlated, with all of the coefficients being larger (in magnitude) for some people and smaller (in magnitude) for other people. This phenomenon is called “scale heterogeneity” because the scale of utility (i.e., the magnitude of all utility coefficients) differs over people, which constitutes a form of correlation among all of the coefficients.

The role of scale in utility can be examined more formally by writing utility as:

$$U_{njt} = \alpha_n' x_{njt} + \frac{1}{\sigma_n}\varepsilon_{njt} \tag{5}$$

where σ_n is inversely proportional to the standard deviation of the error term, and α_n is a random vector. Since utility has no units, Eq. (5) can be written equivalently as:

$$U_{njt} = (\alpha_n \sigma_n)' x_{njt} + \varepsilon_{njt} \tag{6}$$

Suppose now that σ_n varies over people. This variation causes all of the utility coefficients to be correlated with each other, since all the utility coefficients depend on σ_n. Scale heterogeneity creates correlation among all utility coefficients, with the correlation taking a very specific pattern.

A mixed logit model with full covariance among coefficients includes all sources of correlation, including the correlation that is induced by scale heterogeneity. However, the various sources of heterogeneity cannot be distinguished empirically. For example, suppose that two sources of correlation exist for people's choices among appliances: (i) people who are greatly affected by rebates are also greatly affected by attractive financing, creating a positive correlation between the rebate and financing coefficients; and (ii) people who don't want to borrow money tend to want rebates, creating a negative correlation. The researcher estimates the correlation between the two coefficients, and, suppose the estimate is 0.3. This estimate captures the combined effect of the two phenomena. The estimate being positive does not mean that the second source does not exist; it simply means that the combined effect of both phenomena is a positive correlation. Similarly, suppose that scale heterogeneity exists as a third source of correlation. Scale heterogeneity creates positive correlation in the coefficients (assuming, as reasonable, that both have positive means). The correlation of 0.3 then includes the impact of scale heterogeneity in addition to the impact of the other two behavioral phenomena. It is not possible to determine what portion of the 0.3 is due to scale heterogeneity, what portion is due to people liking both financing and rebates, and what portion constitutes the negative correlation induced by the people who do not want to borrow but like rebates.

4. Model comparisons

A mixed logit model that allows all utility coefficients to be randomly distributed and estimates a full covariance matrix among them is the most general form possible. Such a model allows for the type of correlation that would result from scale heterogeneity as well as other behavioral sources that can affect the overall level of correlation between utility coefficients. Such models are computationally feasible in many, if not most, settings; see e.g. the large scale application by Hess et al. (2017). In some situations, however, the researcher might choose to restrict the model, either to avoid the computational burden of a full covariance matrix, or to focus on behavioral factors for which correlations are not necessarily relevant. Such restriction might be not be unreasonable (as we discuss below); however, interpretation of the estimates needs to recognize the implications of the restrictions. In the subsections below, we discuss various types of mixed logit models, including those that were developed to focus on scale heterogeneity.

4.1. S-MNL

A scaled multinomial logit (S-MNL model) is a version of mixed logit where, in Eq. (6), σ_n varies across people while α_n is kept fixed. The utility coefficients are then $\beta_n = \alpha\sigma_n$ where α is a fixed (non-random) vector and σ_n is a random scalar. This model allows for scale heterogeneity, which induces the utility coefficients to vary together through their common dependence on σ_n. The scale parameter is the standard deviation of σ_n: greater variation in σ_n leads to greater variation in utility coefficients and greater covariance among utility coefficients. However, the model restricts the utility coefficients to vary only because of scale heterogeneity, i.e. imposing homogeneity in relative sensitivities, such as WTP.

If utility coefficients vary for reasons other than scale heterogeneity, as one would generally expect in any real world setting, then the estimate of the scale parameter will capture at least some of this variation. As a result, the researcher cannot interpret the estimated scale parameter as measuring the extent of scale heterogeneity. The scale parameter necessarily captures whatever source of variation exists for the utility coefficients. Similarly, if the S-MNL model is found to fit better than a logit with fixed utility coefficients, the improvement does not necessarily indicate that scale heterogeneity exists. It simply means that some form of variation in utility coefficients exists that is captured by the scale parameter. In order to conclude that scale heterogeneity exists, the researcher would need to test the hypothesis that no other forms of variation exist, which requires estimating a more general model. We discuss this testing more directly in the next subsection.

4.2. G-MNL

The most prominent model seeking to capture scale heterogeneity is the “generalized multinomial logit model”, or G-MNL, introduced by Keane (2006), Fiebig et al. (2010), Greene and Hensher (2010). Starting with utility expressed in (6), the authors decompose each element of α_n into its a mean and a person-specific deviation: for the l-th element, $\alpha_{nl} = a_l + \tilde{\alpha}_{nl}$. Then the utility coefficient for the element, β_{nl}, is expressed as:

$$\beta_{nl} = \sigma_n a_l + (\gamma + \sigma_n(1-\gamma))\tilde{\alpha}_{nl} \tag{7}$$

where γ (bounded between 0 and 1) determines the differential influence of scale σ_n upon the person-specific deviations $\tilde{\alpha}_{nl}$. The scale σ_n is assumed to be log-normally distributed with its mean normalized to 1 for identification purposes. Greater variation in σ_n represents greater correlation among utility coefficients.

The estimation of γ is difficult, and many applications set it to 0. The utility coefficients then take the simpler form:

$$\beta_{nl} = \sigma_n a_l + \sigma_n \tilde{\alpha}_{nl} = \sigma_n \ \alpha_{nl} \tag{8}$$

meaning that the impact of σ_n is the same on the means and deviations. The resulting utility thus takes the form of Eq. (6) with both σ_n and α_n being random. The issues regarding correlation and interpretation are the same conceptually whether the G-MNL utilizes Eq. (7) or Eq. (8).

In the theoretical descriptions of the model and in the vast majority of applications, α_n is specified to be a vector of uncorrelated random terms. We will discuss the case of correlation in α_n at the end of this subsection; we assume until then that the elements of α_n are uncorrelated, which is the case in nearly all applications of G-MNL.

When the elements of α_n are uncorrelated, G-MNL is a mixed logit with a highly restricted form of correlation among utility coefficients. For K coefficients, a mixed logit with full correlation contains $K(K-1)/2$ correlation parameters. The G-MNL model reduces the number of correlation parameters to 1, namely, the parameter for scale heterogeneity. While, unlike the S-MNL model in Section 4.1, the G-MNL model does not assume that scale heterogeneity is the only sources of variation in utility coefficients, the model nevertheless assumes that scale heterogeneity is the only source of *correlation* among utility coefficients. The parameter for scale heterogeneity is often called the scale parameter, but it can also be considered the G-MNL model's correlation parameter, since it is the only parameter that represents correlation.

A G-MNL model is appropriate if scale heterogeneity is indeed the only source of correlation. However, estimation of a G-MNL does not identify whether this supposition is true. The correlation parameter is usually found to be statistically significant. However, this result does not mean that scale heterogeneity exists: any source of correlation among coefficients is picked up by this one correlation parameter. In particular, the correlation parameter in the G-MNL captures, to the extent possible, whatever sources of correlation exist in the data. The model does not estimate scale heterogeneity; it estimates the combined impact of all sources of correlation on the model's one correlation parameter.

Researchers who use G-MNL often claim that it disentangles preference heterogeneity from scale heterogeneity. This is not true. As we discussed above, when multiple sources of correlation are present in the real world, the estimated correlation captures the combined effect of all of them. G-MNL assumes that sources other than scale heterogeneity do not exist, and estimates one correlation parameter under this assumption. But the estimate of this one correlation parameter necessarily captures whatever sources of correlation exist in the real-world.

The question arises then: is it possible to conclude from a G-MNL model that scale heterogeneity exists* The answer is, no, at least not stated in that way. The estimate of the model's one correlation parameter can be statistically significant when scale heterogeneity does not exist in a given setting and yet other sources create correlation that is, in part, similar to that created by scale heterogeneity. There *is* a statement about scale that can perhaps be made, but it is a very weak one. The researcher can estimate two models: (1) an unrestricted mixed logit with full covariance and (2) the G-MNL model, which restricts the covariance matrix to the one-parameter form that arises when only scale heterogeneity exists.[2] The researcher then tests the hypothesis that the restrictions implied by the G-MNL model are true. If the hypothesis cannot be rejected, then the researcher can conclude: "the hypothesis cannot be rejected that the correlation pattern among coefficients takes the form that would arise if scale heterogeneity were the only source of correlation."

The work in Keane and Wasi (2013) goes in that direction. However, once a standard mixed logit with full correlations has been estimated, the researcher might not see the need to estimate a G-MNL. Most researchers would probably not think that the only source of correlation in their data is scale heterogeneity. And if the researcher estimates the G-MNL and finds that the restrictions cannot be rejected, the researcher still needs to decide whether the failure means that no other sources of correlation exist, or that the power of the test is low because the data are insufficient to capture the impacts of other sources of correlation.

This brings us to the issue of the name G-MNL. The name states that it is a generalization of a "multinomial logit" model, i.e., a logit model whose coefficients are fixed. However, the name has often been misinterpreted to mean that it is a generalization of mixed logit, which is it not. G-MNL is a highly restricted form of mixed logit, where the full covariance matrix is reduced to one parameter.

Compared to a mixed logit with *uncorrelated* coefficients, G-MNL can be considered a generalization, since G-MNL contains one correlation coefficient. This comparison is perhaps the origin for the misstatements about G-MNL: that some researchers think that a mixed logit always has uncorrelated coefficients. However, the term "mixed logit" has never been defined as only models with uncorrelated coefficients. The McFadden and Train (2000) theorem that mixed logit can approximate any choice model, which is widely cited as justification for using the model, holds only when the definition of mixed logit is not restricted to models with uncorrelated coefficients. And software to estimate mixed logits with no, partial, and full covariance, by both classical and Bayesian

[2] The marginal distribution of utility coefficients would need to be the same in both models, such that the only differences arise in the presence or absence of correlations. A G-MNL model with uncorrelated elements α_n has utility coefficients that, typically, are distributed as the product of a log-normally distributed term and a normally distributed term. The mixed logit with full covariance would need to be specified with the same product of a lognormal times a normal, appropriately normalized for identification, but allow full covariance among the utility coefficients. Otherwise, differences in the log-likelihood function would arise due to the different specifications of marginal distributions rather than only from the restriction on correlations.

methods, have been available for decades.

A further issue arises in that, in many applications of G-MNL, the multiplication by scale σ_n is restricted to attributes that vary over decision-makers and is not applied to the alternative specific constants (ASCs). This practice follows the suggestion by Fiebig et al. (2010) not to scale ASCs because they are "fundamentally different" from other model components. It is important to recognize, however, that this practice is inconsistent with the definition of scale in a random utility model and means that the model, if specified in this way, does not in fact allow for scale heterogeneity. The model allows for a one-parameter form of correlation among the coefficients of some of the variables, which might be useful if these variables are viewed as fundamentally different, but it does not allow for scale heterogeneity.

As stated earlier, it is possible to specify the G-MNL model in such a way as to allow for correlation in α_n; one example is Czajkowski et al. (2014). However, then the G-MNL model simply becomes a mixed logit in the form of Eq. (1). It does not generalize mixed logit, nor does it separately identify the various sources of correlation. The parameters associated with α_n and σ_n are identified empirically by their assumed distributional forms. For example, if σ_n is specified to be log-normal and α_n is joint normal, then the estimated scale parameter captures whatever variation exists that is closer to log-normal than normal.[3]

4.3. Models in WTP-space

In the mixed logit specification of Eq. (1), the willingness to pay (WTP) for an attribute is calculated as $wtp_n = -\beta_n^a/\beta_n^p$, where $\beta_n{}^a$ is the coefficient of the attribute and $\beta_n{}^p$ is the price coefficient. The distribution of WTP is derived from the estimated distribution of $\beta_n{}^a$ and $\beta_n{}^p$.

Models in WTP-space reparameterize utility such that the distribution of WTP is estimated directly (e.g. Train and Weeks, 2005; Scarpa et al., 2008). Utility takes the form:

$$U_{njt} = -\beta_n^p p_{njt} + \beta_n^p wtp_n' x_{njt}^a + \varepsilon_{njt} \tag{9}$$

where p_{njt} is price, $x_{njt}{}^a$ is a vector of non-price attributes, and wtp_n is a corresponding vector of the consumer's WTP for the non-price attributes. The researcher specifies and estimates the distribution of $<\beta_n^p, wtp_n>$. This model is the same as Eq. (1) in the sense that any distribution of utility coefficients in Eq. (1) can be represented by a distribution of $<\beta_n^p, wtp_n>$ in Eq. (9), and vice versa. Models that utilize the parameterization in Eq. (1) are called models in preference-space, and those using Eq. (9) are called models in WTP-space.[4] As stated, the reason for implementing models in WTP-space is to estimate the distribution of WTP directly, rather than estimate the distribution of utility coefficients and then derive the implied distribution of WTP, which may be difficult or impossible with some choices of distributions (Daly et al., 2012). Note that this specification is not useful when the price does not enter utility linearly, since, with nonlinear price effects, WTP is not simply the ratio of the attribute coefficient to the coefficient of the price variable.

The model allows for scale heterogeneity, since each utility coefficient includes $\beta_n{}^p$. If $<\beta_n^p, wtp_n>$ is specified to have full covariance, then the model in WTP-space allows for all sources of correlation. If $<\beta_n^p, wtp_n>$ is restricted to have uncorrelated WTP's, then the model in WTP space does not account for forms of correlation beyond scale heterogeneity; as a result, the estimated variation in $\beta_n{}^p$ can reflect whatever other sources can be captured by this variation.

When a goal of the analysis is to estimate consumers' WTP, or to conduct welfare analysis, it is important that the price coefficient in Eq. (1) be negative for all consumers. That is, the distribution of the price coefficient must have support only for negative values, as occurs with a lognormal distribution on the negative of price. If the distribution overlaps zero, as occurs with a normal distribution, then the mean WTP is undefined (infinite) for all attributes, and the mean welfare gain or loss from any policy is undefined (infinite), as discussed by Daly et al. (2012). The model is therefore unuseable for calculating mean WTPs and for welfare analysis.

This issue brings us back to G-MNL. The price coefficient in a G-MNL is specified as the product of the scale parameter σ_n and the element of α_n that corresponds to price. In all applications to our knowledge, α_n, when random, has been given a normal distribution. As a result, the price coefficient overlaps zero, and mean WTP is undefined. A G-MNL model specified in this standard way cannot be used to calculate mean WTPs or for welfare analysis of any policies. This problem can be avoided by specifying the element of α_n that corresponds to price as having a distribution that does not overlap zero.

Two more notes are required. First, it has come to our attention (by a reviewer of an early revision of this paper) that some researchers think that models in WTP-space are scale-free. This is not true: variation in $\beta_n{}^p$ arises from variation in customers' response to unincluded factors (i.e., scale). Without loss of generality, utility can be normalized to be in dollar units:

$$U_{njt} = -p_{njt} + \beta_n^p wtp_n' x_{njt}^a + (1/\beta_n^p)\varepsilon_{njt} \tag{10}$$

The standard deviation of the unobserved factors is the inverse of the random price coefficient, which represents scale heterogeneity.

Secondly, some analysts seem to believe that models in WTP-space constitute a form of G-MNL where the element of α_n that corresponds to price, labeled $\alpha_n{}^p$, is constrained to equal 1. This is not true: $\beta_n{}^p$ in a model in WTP-space is simply rewritten as $\sigma_n \alpha_n^p$ for a G-MNL. Any variation in σ_n and/or α_n in the G-MNL is represented as variation in $\beta_n{}^p$ in a model in WTP space. The

[3] Several papers published after Hess and Rose (2012) use the G-MNL model to identify demographic variables that relate to scale. Examples include Boerger (2015), Czajkowski et al. (2014b). The scale parameter σ_n is expressed as a function of demographic variables, as in a heteroscedastic mixed logit model. However, as the scale term still includes a random disturbance, and, as with a log-normal distribution, the mean and variance are both a function of the mean of the log of σ, the estimated coefficient of each demographic variable on scale necessarily still captures whatever other sources of variation in utility coefficients is related to that demographic variable.

[4] This model is also called a numeraire model, because the parameters being estimated have been converted to dollar equivalents, with the scale of unobserved terms entering as the price coefficient that is multiplied by all terms. See Eq. (10) below.

model in WTP-space incorporates random scale (i.e., σ_n), without attempting to separate terms that are not separately identified. If α_n and σ_n are specified in the G-MNL to have log-normal distributions (such that the G-MNL has finite mean WTPs), then the price coefficient in the G-MNL, $\sigma_n \alpha_n^p$, is itself log-normal (since the product of two lognormals is lognormal) and the parameters of the two lognormals for α_n and σ_n are not identified: only the parameters of their product are identified. The price coefficient in the G-MNL with this specification is the same as a log-normal price coefficient in a model of WTP space. When the G-MNL is specified to have a normal distribution for ${\alpha_n}^P$ and a lognormal for σ_n (which is the standard form, with undefined mean WTPs), then the equivalent distribution for the price coefficient in a model WTP-space is a lognormal-mixture of normals.

4.4. Latent class models

Latent class models are mixed logits in the form of Eq. (4). Each element r of set S represents a “class.” The utility coefficients β_r are different in each class, and $\pi(\beta_r)$ is the share of the population in class r. The goal of the researcher is to estimated β_r and $\pi(\beta_r)$ for all $r \in S$.

Latent class models, by their definition, allow for full correlation among utility coefficients. The covariance matrix for utility coefficients is $Cov(\beta) = \sum_{r \in S} \pi(\beta_r)[(\beta_r - \overline{\beta})(\beta_r - \overline{\beta})']/R$ where R is the number of classes, and $\overline{\beta}$ is the mean of β_r over classes. As such, any form of correlation is permitted, including the form of correlation that is induced by random scale.

A model called “scale-adjusted latent class,” or SALC, was proposed by Magidson and Vermunt (2005). This model has been said to generalize standard latent class models by allowing for scale heterogeneity. However, standard latent class models already allow for scale heterogeneity, as well as other forms of correlation. SALC allows for scale heterogeneity *within* each class, which standard latent class models do not allow. That is, the SALC allows for the one-parameter form of correlation within each class that is induced by random scale. As such, the SALC model does not disentangle scale heterogeneity from preference heterogeneity: the estimated scale parameter incorporates the impact of all sources of within-class correlation that exist.[5]

5. Conclusions and guidance

In conclusion: Researchers have many options for representing heterogeneity. Allowing for scale heterogeneity is possible using numerous different approaches, and allowing for all sources of correlation is also possible and more general. However, researchers need not feel that representation of any source of correlation is an absolute requirement, or that accommodating scale heterogeneity is somehow more important than other patterns of heterogeneity and correlation.

Given the discussion in this paper, the question arises of how an analyst can, or should, approach the task of specifying his/her model in any given application. We have a few suggestions that we think would help without preventing researchers from pursuing their own objectives in the way that they think is best.

- If you want to allow all forms of correlation among utility coefficients, then estimate a mixed logit model with full covariance. Software to do this is widely available, using classical and Bayesian methods. With classical estimation, it is useful to have good starting values. One practice that we often use is to estimate the model with uncorrelated coefficients first and then enter those estimates as starting values for the model with full covariance. Also, Bayesian estimation procedures are effective for mixed logit models with many parameters, as demonstrated in a recent large scale application by Hess et al. (2017). They are as fast with full covariance as with uncorrelated coefficients, and provide estimates even when the log-likelihood is highly non-quadratic (Huber and Train, 2001). Under fairly benign conditions, the Bayesian estimator is asymptotically equivalent to the maximum likelihood estimator (see e.g. Train, 2009), and so a classicist can treat the Bayesian estimates the same as if they were maximum likelihood. Or the Bayesian estimates can be used as starting values in classical estimation; we have found this procedure to be very effective. Stata contains commands for mixed logit estimation with full or no covariance by both classical and Bayesian procedures. Packages for both types of estimation are also available in R, Matlab, and other programming languages. Models with hundreds of parameters can be readily estimated on these widely-available codes.
- If you want to estimate WTPs and/or do welfare analysis, then be sure to specify a distribution for the price coefficient that does not overlap zero. Also, you might find that using a model in WTP-space is more convenient than models in preference space, because it allows you to estimate the distribution of WTPs directly. Avoid using the standard G-MNL, which has a normal distribution for ${\alpha_n}^P$, because, as we discuss above, its mean WTPs are undefined; if you want to use G-MNL, then respecify it to have, e.g., a lognormal distribution for ${\alpha_n}^P$.
- You may want to restrict the covariance terms in your model, even after considering our first point above. This can be an appropriate specification choice. Keane and Wasi (2013) tested a variety of mixed logit models on ten different datasets and found that restrictions on the full covariance were accepted in many cases, which suggests that full covariance is not needed in all situations. There is a caveat, however: they found that different specifications (full covariance, covariance for scale-heterogeneity only, no covariance) fit best on different datasets, which means that a researcher cannot know, without testing, whether the restrictions they want to place are valid for their own dataset.
- If you restrict your model to allow for scale heterogeneity but not other forms of correlation, it is important that you do not

[5] With S^* classes in the scale layer and S classes in the lower layer, the SALC model will likely offer improvements in fit over a standard latent class model using S classes. This is however simply a result of increasing the distributional flexibility as the new model now uses S^*S classes, and the specific structure imposed by the two layer approach in turn means that a single layer model with S^*S freely estimated classes will offer greater flexibility still.

interpret your results in a way that is inconsistent with your specification. In particular, the estimated scale parameter captures whatever correlation exists in the data that can be accommodated by this one parameter. The estimated parameter being statistically significant does not mean that significant scale heterogeneity exists, because the significance might arise from other sources of correlation that are picked up by the scale parameter. You also cannot state that you have disentangled scale from preference heterogeneity, because the two are not separately identified.

- If you estimate a model that allows for scale heterogeneity but do not scale some of the coefficients (such as the alternative specific constants), then you cannot interpret the model as allowing for scale heterogeneity. The model might be reasonable and appropriate for its purposes (as in point 3 above), but it does not allow for scale heterogeneity.
- If you restrict your model to have uncorrelated utility coefficients, then it does not allow for scale heterogeneity or any other sources of correlation. Importantly, the distribution of ratios of coefficients, which represent WTP and marginal rates of substitution, can be over- or under-estimated because the correlations among coefficients are not captured.

Acknowledgements

The first author acknowledges the financial support by the European Research Council through the consolidator grant 615596-DECISIONS.

Appendix A. Appendix

We list below a few of the incorrect statements about scale that have been made in published papers, in papers submitted to journals and conferences, and by reviewers of papers that were submitted to journals. We simply reproduce the statements and trust that the reader – maybe after having completed our paper, or maybe even before – will recognize the errors. For readers who think that a statement sounds right to them, we refer to our on-line version of this appendix,[6] where we discuss each statement more fully.

"*Despite the importance of the MXL [mixed logit] in accounting for preference heterogeneity, there are other sources of heterogeneity (such as scale heterogeneity) that the model fails to account for*."

"*The data is analyzed using the generalized multinomial logit model, which is able to simultaneously account for both the heterogeneity in taste and scale. This model in essence extends the widely-used random parameter logit (or mixed logit) model by adding the ability to capture un/observed scale heterogeneity*."

"*Later advances in the DCM literature [have] led to the introduction of [the] generalized multinomial logit model (GML) that accounts for both preference and scale heterogeneity*."

"*The structure of these [mixed logit] models can be further enriched [by using G-MNL], allowing for scale heterogeneity and different distributional assumptions for the parameters*."

"*It is well known that for discrete choice models assuming homoscedastic variances (or homogenous scales) would lead to biased and inconsistent preference parameter estimates when the assumption is not true. It is therefore not uncommon for choice modelers to explicitly estimate the scale or variance functions (e.g. heteroscedastic logit/probit, G-MNL)*."

"*There is an emerging literature on the confounding of preference and scale heterogeneity in mixed logit (and other) models*."[7]

"*The generalized mixed logit model accommodates scale heterogeneity*."[8]

"*Given that under some circumstances the MIXL [mixed logit] model can be seen as a special case of the GMNL model, why not use only the GMNL approach**"

"*[P]ublished papers that estimate mixed logit models and claim that preference heterogeneity exists…make the strong assumption of scale homogeneity*."

"*GMNL can model preference heterogeneity over individuals – as RPL is nested within it – and, unlike either RPL or latent class models, can also model scale heterogeneity over individuals or choice tasks*"

"*This shows that, contrary to the assumption in the [mixed logit] model, [scale] should not be normalized across individuals, and that the [G-MNL] specification is the more appropriate model*."

"*the improvement in statistical fit provided by allowing for scale heterogeneity is substantial*."[9]

A paper compares a SALC model to a fixed-coefficients logit model for choices among healthcare innovations, and finds a substantial improvement in fit by the former. The authors state that "*[t]his suggests that there are some people showing different preferences with different error variances (or 'choice uncertainty')*".

[6] http://www.stephanehess.me.uk/papers/Hess_Train_2017_online_appendix.pdf
[7] As if the confounding is not inherent and could, with the right models, be avoided.
[8] As if mixed logit does not.
[9] The improvement in fit came from allowing one parameter for correlation rather than none; the one parameter need not be picking up scale heterogeneity.

"*Unfortunately, in most choice models, including general latent class models, the parameter estimates describing preferences are perfectly confounded with the inverse of the error variance…As such, the use of the [SALC] model in being able to group individuals on the basis of holding similar preferences, whilst accounting for potentially confounding differences in variability, is likely to be attractive to researchers for future research in the field of education research particularly in contexts where identification of distinctive segments is important*."

References

Boerger, T., 2015. Are fast responses more random* Testing the effect of response time on scale in an online choice experiment. Environmental and Resource Economics mar 65 (2), 389–413.⟨http://dx.doi.org/10.1007/s10640-015-9905-1⟩.

Czajkowski, M., Bartczak, A., Giergiczny, M., Navrud, S., Zylicz, T., 2014. Providing preference-based support for forest ecosystem service management. For. Policy Econ. 39, 1–12.

Czajkowski, M., Kadziela, T., Hanley, N., 2014b. We want to sort! assessing households' preferences for sorting waste. Resource and Energy Economics, 36 (1), 290–306, January.⟨http://dx.doi.org/10.1016/j.reseneeco.2013.05.006⟩.

Daly, A., Hess, S., Train, K., 2012. Assuring finite moments for willingness to pay estimates from random coefficients models. Transportation 39 (1), 19–31.

Fiebig, D.G., Keane, M., Louviere, J.J., Wasi, N., 2010. The generalized multinomial logit: accounting for scale and coefficient heterogeneity. Mark. Sci. 29 (3), 393–421.

Greene, W.H., Hensher, D.A., 2010. Does scale heterogeneity across individuals matter* An empirical assessment of alternative logit models. Transportation 37 (3), 413–428.

Hess, S., Rose, J.M., 2012. Can scale and coefficient heterogeneity be separated in random coefficients models*. Transportation 39, 4.

Hess, S., Murphy, P., Le, H., Leong, W., 2017. Estimation of new monetary valuations of travel time, quality of travel and safety for Singapore. Transportation Research Record, forthcoming.

Huber, J., Train, K., 2001. On the similarity of classical and bayesian estimates of individual mean partworths. Mark. Lett. 12 (3), 259–269.

Keane, M., Wasi, N., 2013. Comparing alternative models of heterogeneity in consumer choice behavior. J. Appl. Econ. 28 (6), 1018–1045.

Keane, M., 2006. The generalized logit model: preliminary ideas on a research program. presentation at Motorola-CenSoC Hong Kong meeting, October 22, 2006.

Magidson, J., Vermunt, J., 2005. Technical Guide to Latent Gold Sofrware 4.5. Statistical Innovation.

McFadden, D., Train, K., 2000. Mixed MNL Models for discrete response. J. Appl. Econ. 15 (5), 447–470.

Revelt, D., Train, K., 1998. Mixed Logit with repeated choices: households' choices of appliance efficiency level. Rev. Econ. Stat. 80 (4), 647–657.

Scarpa, R., Thiene, M., Train, K., 2008. Utility in willingness to pay space: a tool to address the confounding random scale effects in destination choice to the alps. Am. J. Agric. Econ. 90 (4), 994–1010.

Swait, J., Bernardino, A., 2000. Distinguishing taste variation from error structure in discrete choice data. Transp. Res. Part B 34 (1), 1–15.

Train, K., 1998. Recreation demand models with taste variation. Land Econ. 74 (1), 230–239.

Train, K., 2009. Discrete Choice Methods with Simulation, second edition. Cambridge University Press, Cambridge, MA.

Train, K., Weeks, M., 2005. Discrete choice models in preference space and willingness-to-pay space. In: Scarpa, R., Alberini, A. (Eds.), Application of simulation methods in environmental and resource economics. Springer, Dordrecht, Ch. 1, pp. 1–16.