Paper:
Hess, S and Rose, JM *Can scale and coefficient heterogeneity be separated in random coefficients models?* Transportation. (Submitted)

Can scale and coefficient heterogeneity be separated in random coefficients models?

Stephane Hess* John M. Rose†

Abstract

There is growing interest in the notion that a significant component of the heterogeneity retrieved in random coefficients models may actually relate to variations in absolute sensitivities, a phenomenon referred to as scale heterogeneity. As a result, a number of authors have tried to explicitly model such scale heterogeneity, which is shared across coefficients, and separate it from heterogeneity in individual coefficients. This direction of work has in part motivated the development of specialised modelling tools such as the G-MNL model. While not disagreeing with the notion that scale heterogeneity across respondents exists, this paper argues that attempts in the literature to disentangle scale heterogeneity from heterogeneity in individual coefficients in discrete choice models are misguided. In particular, we show how the various model specifications can in fact simply be seen as different parameterisations, and that any gains in fit obtained in *random scale* models are the result of using more flexible distributions, rather than an ability to capture scale heterogeneity. We illustrate our arguments through an empirical example and show how the conclusions from past work are based on misinterpretations of model results.

Keywords: random scale; discrete choice; taste heterogeneity; scale heterogeneity; mixed logit

1 Introduction

After several decades of empirical research, it has been conclusively shown that even when faced with similar choice situations, two different decision makers will often exhibit different preferences. Given that the vast majority of discrete

*Institute for Transport Studies, University of Leeds, s.hess@its.leeds.ac.uk, Tel: +44 (0)113 34 36611

†Institute of Transport and Logistics Studies, The University of Sydney, john.rose@sydney.edu.au

choice studies rely on pooled data collected from multiple individuals, constructing representative models accounting for such heterogeneous response behaviour is important for several reasons. Most critically, parameter estimates from models that fail to account for heterogeneous sensitivities may be biased which in turn may impact upon other important outputs such as willingness to pay or elasticity estimates.

In recent years, particular attention has been paid to how to adequately represent random variations across respondents, i.e. heterogeneity that cannot be linked to measurable aspects of the decision makers (see for example the discussions in Train, 2009; Hensher and Greene, 2003; Swait, 2006). In this context, a number of researchers have openly questioned whether what is being captured in random parameter models is actually not heterogeneity in individual sensitivities[1], but at least in part differences in scale across choice tasks or respondents (see e.g. Louviere et al., 1999, 2002; Louviere and Eagle, 2006; Louviere and Meyer, 2008; Louviere et al., 2008).

As is well known (see e.g. Ben-Akiva and Lerman, 1985), scale is both confounded with the deterministic component of utility as well as being inversely related to the error variance within the choice data. As such, the larger (smaller) the error variance, the smaller (larger) the parameters of the deterministic component of utility will be. Any observed differences in estimated parameters could be the result of different marginal utilities, different error variances, or both. Taking this argument further, this relationship therefore also has the potential to impact upon how one might view heterogeneity of the sort obtained from using Mixed Multinomial Logit (MMNL) models. In particular, it is possible that some of what is being modelled is not heterogeneity in individual sensitivities, but rather scale heterogeneity that would impact on all parameters in the same way. This point is precisely the argument put forward by Louviere et al. (1999, 2002, 2008), Louviere and Eagle (2006), and Louviere and Meyer (2008). Some authors have gone as far as suggesting that homogeneity in relative sensitivities (which would arise if all heterogeneity was caused by scale differences) may be more common than previously thought and that differences in estimated utility parameters are primarily the result of scale differences (see e.g. Swait and Bernardino, 2000; Fiebig et al., 2010).

While acknowledging that scale heterogeneity may indeed play a role, the present paper puts forward the argument that recent efforts to separately identify random scale heterogeneity have been misguided. In particular, we base this reasoning on the fact that, econometrically, a linear in parameters specification

[1]We focus on random scale heterogeneity alone in this paper, thus not looking at efforts to link scale heterogeneity to characteristics of the respondent, the alternative, or the choice set.

of the logit model perfectly confounds scale with taste sensitivity. A stochastic treatment of scale thus implies a (perfectly correlated) stochastic treatment of taste intensities. Similarly, a stochastic treatment of heterogeneity in individual sensitivities means that a model also allows for scale heterogeneity.

The above reasoning implies that recent work aimed at providing separate and uncorrelated stochastic treatments of 'scale' and 'taste sensitivities' (e.g. Fiebig et al., 2010; Greene and Hensher, 2010; Hess and Rose, 2010) ignores the existence of the the scale/taste sensitivity confound. We argue that, as a result, the existing interpretation of results from these models is incorrect. Models estimated in this manner simply allow for more flexible distributions, thus uncovering from the data particular correlation structures within the heterogeneity that is being modelled whilst maintaining the scale/taste sensitivity confound.

The remainder of this paper is organised as follows. The next section outlines the theory behind our claims, which is followed in Section 3 by a discussion of the implications for past and future work. Section 4 provides empirical support to our claims, and Section 5 summarises the findings of the paper and presents the conclusions of the research.

2 Theory

2.1 Background

Random utility models decompose utility into a deterministic component and a random component, or error term. The *scale* of the model is inversely proportional to the variance of this error term; if the variance of the error term goes up, the *scale* of the model goes down and vice versa. The confounding issue arises as such scale differences can also be accommodated by increasing or decreasing the parameters in the deterministic utility.

To illustrate this point, denote the deterministic component of utility for alternative i of person n in choice situation t as $V_{int} = \beta' x_{int}$, where x_{int} is a vector of attributes describing alternative i as well as decision maker n in choice situation t, and β is a vector of parameters to be estimated. We define α to be the scale parameter of the extreme value distribution that is assumed for the error term in the Multinomial Logit (MNL), giving us the following choice probabilities:

$$P_{int}(\alpha, \beta) = \frac{e^{\alpha V_{int}}}{\sum_{j=1}^{J_{nt}} e^{\alpha V_{jnt}}}. \tag{1}$$

It is immediately clear that changes in model scale can be accommodated both through changes in β and changes in α, meaning that the model is overspecified;

the two components α and β are not separately identified, and what is in fact estimated is $\theta \equiv \alpha\beta$. It is sometimes stated that the normalisation $\alpha = 1$ is placed on the model, but this statement is just an alternative way of saying that only the product $\alpha\beta$ is actually identified and estimated.

2.2 Random coefficients models and the role of correlation

Recent interest in scale heterogeneity has focussed on random variations across respondents. In a random parameters model, such as the MMNL model, we allow $\theta \equiv \alpha\beta$ to vary across respondents. Working on the basis of intra-respondent homogeneity, the probability of the observed sequence of T choices for respondent n is given by:

$$P_n(\Omega) = \int_\theta \prod_{t=1}^{T} P_{i_c nt}(\theta) f(\theta \mid \Omega) \, \mathrm{d}\theta, \tag{2}$$

where $i_c nt$ is the alternative chosen by respondent n in choice situation t, and where the choice probabilities are obtained through integration of MNL probabilities (as in Equation 1) over the assumed distribution of the vector of θ, $f(\theta \mid \Omega)$, where Ω represents the parameters of this distribution.

This random coefficients framework is, of course, equivalent to one specified in terms of separate random α and β components, as long as a correlated distribution is used for the vector θ (since, when separately estimated, the scalar α multiplies all elements of the vector β). As we will see, it is this requirement to use correlated distributions which avoids the need to explicitly multiply the distribution of individual coefficients by a distribution of a random scale parameter.

Three possible scenarios arise. If all random heterogeneity across respondents is in individual sensitivities, then the model will be able to capture such heterogeneity alongside any correlation in the heterogeneity for different coefficients. If on the other hand, there is only scale heterogeneity, then this will be captured through perfect correlation amongst individual coefficients. In practice, a mixture of the two is likely to arise, with some heterogeneity being in individual coefficients, and some heterogeneity being shared across coefficients, i.e. scale heterogeneity, with the latter being captured through increasing the correlation between individual coefficients. This observation also explains apparently counterintuitive results showing for example positive correlation between time and cost coefficients.

To demonstrate the issue of correlation further, consider the case where some heterogeneity is coefficient-specific, while a remaining part is shared across coefficients. The multiplication of a common scale, α, across all elements in β must by definition induce a particular correlation structure in the marginal utilities, which are given by $\theta = \alpha\beta$. Now, let us contrast two separate approaches. The first approach makes use of the specification from Equation 2 and estimates the distribution of θ. The second approach separately estimates the two components, α^* and β^*, where * denotes the separate estimation. If the first model makes use of a correlated distribution for θ, then it is structurally equivalent to a model that separately estimates α^* and β^*, *provided* that the directly estimated θ follows the same distributional form as $\alpha^*\beta^*$. Any MMNL model that allows for correlated random parameters thus uncovers not just correlated taste sensitivities, but also simultaneously allows for random scale heterogeneity, conditional on all parameters being included in this multivariate distribution. Counter to this, a model that assumes uncorrelated random parameters in θ also assumes that scale is homogenous within the sampled distribution (i.e., the variance of α is zero.) In the presence of a non-trivial amount of scale heterogeneity, such a model is likely to overstate the degree of heterogeneity in individual (and hence relative) sensitivities, as it can only capture scale heterogeneity through increased variance without being able to accommodate the fact that such variation is perfectly correlated across individual coefficients.

The above discussion has highlighted the key role played by an analyst's assumptions relating to correlation in random coefficients models. It also implies that all coefficients of the model should be treated as random, including any alternative specific constants (ASC). Indeed, if any coefficient or ASC is treated as being fixed, then that coefficient cannot by definition be correlated with any of the remaining coefficients, and this is equivalent to imposing the assumption of homogenous scale within the sampled population. It should be acknowledged that such a specification can be difficult to estimate in practice, given limited information content in the data. The same however also applies to any specification with $\theta = \alpha^*\beta^*$, as even with some fixed or independent elements in β^*, all elements in θ will be random.

2.3 Disentangling sources of heterogeneity

The aim of recent work in this area has been to disentangle heterogeneity in individual coefficients from scale heterogeneity. At least in part, these efforts were motivated by concerns that any scale differences between respondents that were not properly accounted for may lead to bias not only in the heterogeneity levels for individual coefficients, but also in the estimated correlation between marginal utility coefficients.

A simple model explicitly aiming to incorporate a random scale component can be specified as follows. Making use once more of the * notation, let us assume that the scale parameter α^* follows a random distribution $h\left(\alpha^* \mid \Omega_{\alpha^*}\right)$ across respondents, where Ω_{α^*} is a vector of parameters. Similarly, denote the distribution of β^* by $f\left(\beta^* \mid \Omega_{\beta^*}\right)$, with parameter vector Ω_{β^*}. We then have:

$$P_n\left(\Omega_{\alpha^*}, \Omega_{\beta^*}\right) = \int_{\alpha^*} \int_{\beta^*} \prod_{t=1}^{T} P_{i_c n t}\left(\alpha^*, \beta^*\right) h\left(\alpha^* \mid \Omega_{\alpha^*}\right) f\left(\beta^* \mid \Omega_{\beta^*}\right) \mathrm{d}\alpha^* \mathrm{d}\beta^*, \quad (3)$$

with $P_{i_c n t}\left(\alpha^*, \beta^*\right)$ being defined once again as in Equation 1, and where it is important to note that the scale parameter α^* and the vector β^* are distributed independently of one another. The scale parameter α^* is positive by definition, leading to the requirement of a constraint on its distribution (or an appropriate transform). Additionally, since only the distribution (i.e., moments) of $\theta = \alpha\beta$ are identified, some normalisation is required when the distributions of α^* and β^* are estimated separately. A convenient normalisation for the means is to set $E(\alpha^*) = 1$ or $E(\beta_k^*) = 1$ for some element k. Other normalisations are specification-specific. For example, if α^* is lognormal and β^* is jointly lognormal, then the variance of α^* or one element of β^* must be normalised to a fixed value, since the product of lognormals is itself lognormal.

The estimation of the model in Equation 3 can lead to four possible outcomes, as follows:

1. the model results do not show significant variance in either α^* or β^*;
2. the model results show significant variance only in α^*;
3. the model results show significant variance only in β^*; and
4. the model results show significant variance in both α^* and β^*.

If the first outcome arises, the model collapses back to a simple MNL specification. The analyst may reach the conclusion that any heterogeneity has already been explained in the deterministic component of the utility. However, in practice, this outcome will be extremely rare, and may be the result of overly restrictive distributional assumptions.

If the second outcome arises, an analyst may reach the conclusion that any heterogeneity across respondents is solely due to differences in scale. This conclusion may again be misguided, as the variation captured in α^* may to a certain (or even large) extent be caused by heterogeneity in β^* that the distributions imposed by the analyst fail to capture.

If the third outcome arises, the analyst may take this as an indication of an absence of scale heterogeneity. However, this could once again be a misguided conclusion as any scale heterogeneity present in the data could have been captured in the correlation structure for the multivariate distribution of β^*, hence making the additional α^* component redundant. We will return to this issue below, in the context of the empirical example.

The fourth outcome is in many ways the one sought by an analyst interested in disentangling the two components of heterogeneity. However, the risk of misguided conclusions remains. Indeed, by using $V = \alpha^*\beta^*x$, we obtain a specification with $V = \theta x$, where it is possible (or even likely) that the correlated multivariate distribution of θ, which is the product of two separate distributions, may be more flexible than those distributions typically used in "simple" MMNL specifications, i.e. θ in Equation 2. As a result, any improvements in model fit obtained by using the specification in Equation 3 instead of the specification in Equation 2 may simply be due to the fact that this more flexible distribution better explains the behaviour in the data.

Conclusions as to the presence of scale heterogeneity can thus not be drawn when the distribution of $\alpha^*\beta^*$ in Equation 3 is different from the distribution of θ in Equation 2. Conversely, as highlighted in the empirical application, when the distribution of θ is equivalent to the distribution of $\alpha^*\beta^*$, the specification in Equation 3 is identical to that in Equation 2, and it is once again not possible to disentangle the two components of heterogeneity.

3 Implications for past and future work

In this section, we highlight the implications of the earlier discussions for past and future work. We first note the equivalence between a number of commonly used specifications before focussing on past work aiming to disentangle the various components of random heterogeneity.

3.1 Equivalence of common specifications

Three common departures from a "standard" specification of the MMNL model have been discussed in the context of random scale heterogeneity. Alongside the simple model discussed in Section 2.3, special attention needs to be given to the generalised multinomial logit (G-MNL) model and models estimated in willingness to pay (WTP) space.

The G-MNL model, first proposed by Keane (2006) and operationalised by Fiebig et al. (2010) and Greene and Hensher (2010), allows for separate impacts of α upon the mean and standard deviation parameters of the MMNL model. As an example, when working with independent normally distributed elements in β, each coefficient can be represented as its mean plus its standard deviation times a standard normal term: i.e., $\beta_k = \mu_k + \sigma_k \eta_{n,k}$ where $\eta_{n,k}$ is iid standard normal.

In the simpler version of the G-MNL, the marginal utility for attribute k is given as

$$\theta_{n,k} = \alpha_n \mu_k + \alpha_n \sigma_k \eta_{n,k} \tag{4}$$

with the normalization $E(\alpha) = 1$. The role of α_n is the same as described above, representing scale. The more general form of the model allows the person-specific part of β_k to enter in two ways: with one part multiplied by α_n and one part not multiplied by α_n, with weighting for the two parts. In particular:

$$\theta_{n,k} = \alpha_n \mu_k + \gamma \sigma_k \eta_{n,k} + (1 - \gamma)\, \alpha_n \sigma_k \eta_{n,k} \tag{5}$$

The weighting parameter γ is bounded between 0 and 1 and reflects the extent to which α_n operates on the person-specific component of β_k, i.e. the variance across respondents in β_k, given by $\sigma_k \eta_{n,k}$. When $\gamma > 0$, α_n no longer represents scale in its traditional form, since not all elements in $\theta_{n,k}$ are multiplied by α_n. In applications, α has been assumed to be lognormally distributed with its expected value normalised to 1. It can be seen that this is not in fact a different model from the one in Equation 2, but rather a different parameterisation, and that with appropriate distributional assumptions, they become equivalent.

In a WTP space specification, denote the element that represents cost as $k = c$ and specify its distribution to have support only over strictly negative numbers, such that the cost coefficient is negative for all n as required for WTP calculations. In most applications, this assumption is implemented by entering cost as the negative of cost and placing a lognormal distribution on its coefficient. In accordance with this practice, let $\theta_{n,c}$ be the coefficient of the negative of cost, such that $\theta_{n,c} > 0$. For all $k \neq c$, the WTP for marginal changes in the attribute is $\lambda_{n,k} = \theta_{n,k}/\theta_{n,c}$. The marginal utility for attribute $k \neq c$ is then, by definition:

$$\theta_{n,k} = \theta_{n,c} \lambda_{n,k} \tag{6}$$

The model is completed by specifying the distribution of the vector of WTP's, i.e., the $\lambda_{n,k}$'s. This approach allows the analyst to specify and estimate the distribution of WTP's directly, rather than deriving the distribution of WTP's from the estimated distribution of θ_n.

For this parameterisation in WTP space, it is sometimes stated that $-\beta_{n,c}$ is normalised to 1, such that $\theta_{n,c} = \alpha_n$ and each $\lambda_{n,k}$ is multiplied by the scale α_n. However, the more accurate statement is that $\theta_{n,c}$ is the product of α_n and $-\beta_{n,c}$, such that each $\lambda_{n,k}$, $k \neq c$ is multiplied by $-\alpha_n \beta_{n,c}$. The only restriction of the specification is that the cost coefficient is assumed to be negative for all n, which is a requirement for economically meaningful WTP's.

It should also be noted that it was in the context of working in WTP space that some of the early discussions on scale heterogeneity took place, with Scarpa et al. (2008, page 996) noting confounding between scale and preference heterogeneity, by observing that:

> *If the scale parameter varies and [the relative sensitivities] are fixed, then the utility coefficients vary with perfect correlation. If the utility coefficients have correlation less than unity, then [the relative sensitivities] are necessarily varying in addition to, or instead of, the scale parameter. Finally, even if [the scale parameter] does not vary over [respondents] ..., utility coefficients can be correlated simply due to correlations among tastes for various attributes. (Scarpa et al., 2008, page 996)*

This discussion has shown that all three parameterisations are equivalent; a given specification for one parameterisation can be replicated with another one by making appropriate distributional assumptions. The same also applies for the "base" model in Equation 2. This observation shows us that none of these specifications is more flexible, or more able to disentangle scale heterogeneity and heterogeneity in relative sensitivities then the others. With an appropriate correlated distribution for the coefficients, each specification captures scale heterogeneity as well as heterogeneity in relative sensitivities, but the two cannot be separately identified. Any *gains* in fit by one specification over the other are simply the result of more flexible distributional assumptions; as such, the differences between models arise in the ease in which given distributional shapes can be accommodated, with advantages for different specifications in different settings.

3.2 Re-interpretation of past results

The discussions in the early parts of this paper have no bearing on past work making use of a deterministic treatment of scale heterogeneity, i.e. where the scale parameter is parameterised on the basis of measurable information relating to the respondent or the choice environment, using a Heteroscedastic Multinomial Logit (HMNL) model (see e.g. Caussade et al., 2005; Dellaert et al., 1999; Hensher et al., 1998; Louviere et al., 2000; Swait and Adamowicz, 2001; Swait and Louviere, 1993). Nevertheless, care is required to avoid a situation in which this specification confounds such heterogeneity with unaccounted for heterogeneity in the β parameters. Separate identification of random heterogeneity in scale and individual coefficients should also be possible in work where the estimation of the distribution of α draws on additional model components, such as the work of Hess and Stathopoulos (2011) where α is used jointly in the choice model component and a separate measurement equations component.

Early work by Breffle and Morey (2000) proposed a model allowing for random scale heterogeneity while maintaining homogeneity in the relative sensitivities. The shortcoming of this approach, as shown in Section 4, is that it is impossible to establish whether the heterogeneity retrieved for α is in fact scale heterogeneity, as any heterogeneity that exists in β may well be captured in α, given the homogeneity assumption imposed on the former.

Recently, there has been growing interest in the use of the the G-MNL model. In line with our own discussions, Fiebig et al. (2010) concede that an alternative interpretation of the G-MNL model, at least in so far as derived conditional individual level parameter distributions are concerned, is that the model provides for more flexible distributions. We argue that this is in fact the only correct interpretation, and is contrary to e.g. Greene and Hensher (2010) who interpret the outputs as separately identified scale and taste intensity heterogeneity. This claim is underlined by noting that both Fiebig et al. (2010) and Greene and Hensher (2010) use a G-MNL specification where a lognormal α is multiplied by a normal β, and contrast this with a MMNL specification using a normal β.

Other work concerned with scale includes the multiplicative error model of Fosgerau and Bierlaire (2009), but this is based on a non-random scale parameter. Error components or models with random alternative specific constants have also been used to accommodate scale differences (see e.g. Walker, 2001; Brownstone et al., 2000), but the main interest has generally been on heteroscedasticity across alternatives, and even if the scale heterogeneity treatment could be limited to be across respondents, such an approach works in a linearly additive manner, thus not incorporating an interaction with the explanatory variables used in the model.

4 Empirical example

We now present a brief empirical example, making use of stated choice data collected for the DATIV study carried out in Denmark in 2004 (cf. Burge and Rohr, 2004). A binary unlabelled route choice experiment was used, with two attributes, namely travel time (TT) and travel cost (TC), describing the alternatives. The final sample used in our analysis contains $17,020$ observations collected from $2,197$ respondents, with up to 8 choice situations per respondent. The various models were coded in Ox 6.2 (Doornik, 2001), using 500 Halton draws (cf. Halton, 1960; Bhat, 2001).

Table 1 first shows a simple MNL model alongside a scale heterogeneity model with homogeneity in relative sensitivities, where this latter model is referred to

Table 1: Estimation results for empirical example: part I

	MNL		S-MNL	
Final LL	-11,378.70		-10,257.30	
par.	2		3	
adj. ρ^2	0.0353		0.1303	
	est.	t-rat.	est.	t-rat.
θ_{TT}	-0.0365	-14.67	n/a[a]	
θ_{TC}	-0.0578	-22.77	n/a[a]	
β^*_{TT}	n/a[b]		-0.0078	-7.80
β^*_{TC}	n/a[b]		-0.0544	-7.96
$\sigma_{\ln(\alpha^*)}$	n/a[b]		-3.7590	-21.57
$cv_{\theta_{\mathrm{TT}}}$	0		1,170.33	
$cv_{\theta_{\mathrm{TC}}}$	0		1,170.33	
$cv_{\beta^*_{\mathrm{TT}}}$	n/a[b]		0	
$cv_{\beta^*_{\mathrm{TC}}}$	n/a[b]		0	

[a] separate components estimated

[b] no separate components estimated

as S-MNL. In the MNL model, we estimate fixed time (θ_{TT}) and cost (θ_{TC}) coefficients. In the S-MNL model, we estimate separate β^* and α^* components, where the random scale parameter α^* follows a lognormal distribution, i.e. using a normal distribution for $\ln(\alpha^*)$. This specification ensures positive signs only, where the mean for the underlying normal distribution was set to zero for identification (giving a median of 1 for α^*), with $\sigma_{\ln(\alpha^*)}$ giving the standard deviation for the underlying normal distribution. No heterogeneity is estimated for θ in the MNL model, and while we make a homogeneity assumption for β^* in the S-MNL model, the heterogeneity in α^* leads to heterogeneity in θ in this model.

The results show very significant gains in model fit for S-MNL over MNL, with an increase in log-likelihood (LL) by $1,121.4$ units for one additional parameter. With the S-MNL model not allowing for additional heterogeneity in β^*, it is conceivable that part of the heterogeneity captured by α^* is in fact caused by heterogeneity in individual sensitivities. No reliable conclusions in relation to scale heterogeneity can thus be drawn from this model. Indeed, the use of a randomly distributed scalar α^* with a fixed vector β^* imposes (rather than reveals) perfectly correlated random elements in θ. It is also conceivable that the

Table 2: Estimation results for empirical example: part II

	MMNL_{UC}		$\text{S-MMNL}_{\text{UC}}$		MMNL_{C}		S-MMNL_{C}	
Final LL	-9663.57		-9367.48		-9653.97		-9347.30	
par.	4		5		5		6	
adj. ρ^2	0.1805		0.2055		0.1813		0.2072	
	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.
$\mu_{\theta_{\text{TT}}}$	-0.1094	-12.48	n/a[a]		-0.1059	-11.06	n/a[a]	
$\mu_{\theta_{\text{TC}}}$	-0.3183	-23.92	n/a[a]		-0.3064	-22.48	n/a[a]	
$s_{11,\theta}$	0.2855	28.73	n/a[a]		0.3204	24.18	n/a[a]	
$s_{21,\theta}$	0	-	n/a[a]		0.0557	4.53	n/a[a]	
$s_{22,\theta}$	-0.2418	-22.82	n/a[a]		-0.2316	-21.35	n/a[a]	
$\mu_{\beta^*_{\text{TT}}}$	n/a[b]		-0.1610	-15.52	n/a[b]		-0.1558	-10.54
$\mu_{\beta^*_{\text{TC}}}$	n/a[b]		-0.4037	-16.81	n/a[b]		-0.4205	-11.25
s_{11,β^*}	n/a[b]		0.2743	17.01	n/a[b]		0.1991	11.42
s_{21,β^*}	n/a[b]		0	-	n/a[b]		-0.1499	-9.50
s_{22,β^*}	n/a[b]		-0.2349	-16.20	n/a[b]		-0.1412	-11.75
$\sigma_{\ln(\alpha^*)}$	n/a[b]		1.9102	20.23	n/a[b]		-2.0843	-20.60
$cv_{\theta_{\text{TT}}}$	2.61		406.87		3.02		461.47	
$cv_{\theta_{\text{TC}}}$	0.76		19.93		0.78		20.25	
$cv_{\beta^*_{\text{TT}}}$	n/a[b]		1.70		n/a[b]		1.28	
$cv_{\beta^*_{\text{TC}}}$	n/a[b]		0.58		n/a[b]		0.49	

[a] separate components estimated

[b] no separate components estimated

excessively long tails obtained with this S-MNL model (see the coefficient of variation for θ) are the result of this assumption; when imposing perfect correlation between coefficients, the true degree of heterogeneity is overstated. Finally, it is interesting to note a disproportionally large impact on β^*_{TT} in the S-MNL model, where the value of time drops from 37.91DKK/hr to just 8.65DKK/hr.

As a next step, we make use of normally distributed coefficients for the time and cost sensitivities. This specification is commonly used in the literature despite extensive discussions about its inappropriateness from a micro-economic perspective (see e.g. Hess et al., 2005) while also causing issues in the computation of willingness to pay measures (cf. Daly et al., 2011). Here, we include it as an illustration of the issues highlighted in Section 2.3. Specifically, we use both uncorrelated (UC) as well as correlated (C) specifications for the distribution of θ, with the resulting models referred to as MMNL_{UC} and MMNL_{C} respectively. We also estimate models which separate out α^* and β^*, once again using a log-normal distribution for α^*, with models referred to as $\text{S-MMNL}_{\text{UC}}$ (uncorrelated) and S-MMNL_{C} (correlated) respectively.

The results for these models are summarised in Table 2. Looking at the notation for the MMNL models, $\mu_{\theta_{\text{TT}}}$ and $\mu_{\theta_{\text{TC}}}$ give the mean parameters, with $s_{11,\theta}$ and $s_{22,\theta}$ giving the diagonal elements in the Cholesky matrix, and $s_{21,\theta}$ giving the off-diagonal element[2]. A corresponding notation is used in the other models.

The two MMNL models obtain highly significant gains in model fit over their MNL counterpart, with the same applying when comparing the two S-MMNL models to their S-MNL counterpart. Looking first at the models with uncorrelated distributions, we see substantial gains in model fit for the $\text{S-MMNL}_{\text{UC}}$ model over the MMNL_{UC} model (296 units for one additional parameter) as a result of separating out α^*. While the degree of heterogeneity in β^* is lower in the $\text{S-MMNL}_{\text{UC}}$ model than was the case for the heterogeneity in θ in the MMNL_{UC} model, the overall heterogeneity in θ is increased substantially in the $\text{S-MMNL}_{\text{UC}}$ model. This result is a reflection of the fact that the distribution of $\theta = \alpha^*\beta^*$ in the $\text{S-MMNL}_{\text{UC}}$ model is different from that in the MMNL_{UC} model. This difference in flexibility also accounts for at least part of the gains in fit (and the split in heterogeneity between α^* and β^*), making it impossible to make inferences about the retrieval of scale heterogeneity.

Using a correlated distribution for θ_{TT} and θ_{TC} in the MMNL_{C} model leads to a small but significant gain in fit (MMNL_{C} *vs.* MMNL_{UC}), while also suggesting positive correlation between the two marginal utility coefficients (note that the product between $s_{11,\theta}$ and $s_{21,\theta}$ is positive). This observation could be seen as a direct result of this model capturing some of the scale heterogeneity in the data. Turning to the S-MMNLC model, we see that separate estimation of α^* and β^* once again leads to substantial gains in model fit (306.7 units for one additional parameter). Just as in the uncorrelated models, the degree of heterogeneity in β^* is lower in the S-MMNL_{C} model than was the case for the heterogeneity in θ in the MMNL_{C} model, and we also note a reversal of the sign of the correlation in that distribution. However, the heterogeneity in the overall distribution θ is once again increased and the correlation between the elements in θ is again positive (noting that $\sigma_{\ln(\alpha)}$ dominates in the covariance between θ_{TT} and θ_{TC}).

[2]With ξ_1 and ξ_2 being independent standard normal variates, draws from the distribution of θ_{TT} are obtained as $\mu_{\theta_{\text{TT}}} + s_{11}(\theta)\,\xi_1$, while draws from the distribution of θ_{TC} are obtained as $\mu_{\theta_{\text{TC}}} + s_{21}(\theta)\,\xi_1 + s_{22}(\theta)\,\xi_2$. Correlation is allowed for as ξ_1 is used for both coefficients, with the covariance between θ_{TT} and θ_{TC} being given by $s_{21}(\theta)\,s_{11,\theta}$.

In past work, the results from this example would have been used as evidence that a) there exists significant scale heterogeneity in the data, and b) the model at hand is able to disentangle the two components of heterogeneity. However, two issues arise. Firstly, the specification used for the distribution of the two marginal utility coefficients is inappropriate in the present context, as it would imply a substantial share of respondents with incorrectly signed time and cost coefficients. Secondly, the distribution used for $\theta = \alpha^*\beta^*$ in the S-MMNL$_{\text{UC}}$ and S-MMNL$_{\text{C}}$ models is now a product between a lognormal distribution and a normal distribution. The resulting distribution, known as the normal lognormal mixture (NLNM) distribution, is commonly used in the financial time series literature, and has well defined mathematical properties (see e.g. Clark, 1973; Yang, 2008). The NLNM distribution is more flexible than the typically assumed normal or lognormal distributions used when estimating MMNL models, in that the distribution (depending on the moments of the two underlying distributions) is non-symmetrical, being leptokurtic with negative skew, and not bounded at zero. From this perspective, the gains in fit (and the different patterns of heterogeneity) obtained by incorporating a random α are arguably at least in part due to this gain in flexibility. In the MMNL$_{\text{UC}}$ *vs.* S-MMNL$_{\text{UC}}$ comparison, the additional issue arises that while the distribution of θ in the former is uncorrelated, the multiplication of uncorrelated β^* distributions by a common α^* distribution leads to correlation in $\theta = \alpha^*\beta^*$.

Informed by these discussions, we now make use of distributional assumptions that will a) ensure meaningful results from a micro-economic theory perspective and b) allow us to avoid issues with differences in flexibility between θ in the MMNL model and $\theta = \alpha^*\beta^*$ in the S-MMNL model. This double aim is achieved by making use of a lognormal distribution for θ in the MMNL model, and lognormal distributions for both α^* and β^* in the S-MMNL model, ensuring that the resulting $\theta = \alpha^*\beta^*$ distribution will similarly be lognormal.

The results are summarised in Table 3, using much the same notation as before, where all estimates now relate to the normal distributions of the logarithms of coefficients. We once again see significant improvement of the MMNL and S-MMNL models over their MNL and S-MNL counterparts in Table 1, while the fit (in terms of adjusted ρ^2) is for each model also superior to that of the corresponding model from Table 2.

In the discussion of these four models, we first focus on the two MMNL models. We see an improvement in model fit by 136 units for one additional parameter (the off-diagonal Cholesky term) when comparing MMNL$_{\text{C}}$ to MMNL$_{\text{UC}}$. This observation highlights the importance of allowing for the correlation between the individual elements in θ, where in the present case, we observe positive correlation (the product between $s_{11,\ln(\theta)}$ and $s_{21,\ln(\theta)}$ is positive), along with substantial

Table 3: Estimation results for empirical example: part III

	MMNL_{UC}	$\text{S-MMNL}_{\text{UC}}$	MMNL_{C}	S-MMNL_{C}
Final LL	-9,462.84	-9,326.45	-9,326.70	-9,323.86
par.	4	5	5	6
adj. ρ^2	0.1975	0.2090	0.2090	0.2092

	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.
$\mu_{\ln(\theta_{\text{TT}})}$	-2.2203	-35.58	n/a[a]		-1.7688	-26.81	n/a[a]	
$\mu_{\ln(\theta_{\text{TC}})}$	-1.1830	-23.03	n/a[a]		-0.9430	-14.98	n/a[a]	
$s_{11,\ln(\theta)}$	1.1842	28.89	n/a[a]		-1.8876	-18.99	n/a[a]	
$s_{21,\ln(\theta)}$	0	-	n/a[a]		-1.7371	-17.99	n/a[a]	
$s_{22,\ln(\theta)}$	1.6605	32.34	n/a[a]		1.5415	60.53	n/a[a]	
$\mu_{\ln\left(\beta^*_{\text{TT}}\right)}$	n/a[b]		-1.7498	-27.54	n/a[b]		-1.7353	-26.23
$\mu_{\ln\left(\beta^*_{\text{TC}}\right)}$	n/a[b]		-0.9205	-14.90	n/a[b]		-0.9238	-13.95
$s_{11,\ln(\beta^*)}$	n/a[b]		0.4127	19.49	n/a[b]		1.9004	18.91
$s_{21,\ln(\beta^*)}$	n/a[b]		0	-	n/a[b]		1.8105	18.44
$s_{22,\ln(\beta^*)}$	n/a[b]		1.4721	63.32	n/a[b]		1.5296	49.52
$\sigma_{\ln(\alpha^*)}$	n/a[b]		1.8556	18.40	n/a[b]		-0.1169	-0.37
$cv_{\theta_{\text{TT}}}$	1.75		6.01		5.85		6.04	
$cv_{\theta_{\text{TC}}}$	3.84		16.50		14.80		16.67	
$cv_{\beta^*_{\text{TT}}}$	1.75		0.43		5.85		6.00	
$cv_{\beta^*_{\text{TC}}}$	3.84		2.78		14.80		16.56	

[a] separate components estimated

[b] no separate components estimated

increases in the degree of heterogeneity. This result is in line with the comparison between MMNL_{C} to MMNL_{UC} in the normal case (cf. Table 2).

We now proceed to the discussion of the two S-MMNL models. In both models, the coefficients used to multiply the attributes in the utility functions are given by $\theta = \alpha^*\beta^*$. The multiplication of two lognormals produces another lognormal distribution, where, independently of whether β^* is uncorrelated or correlated, the resulting distribution for θ will be correlated. With the exception of MMNL_{UC}, the various specifications in Table 3 are thus formally equivalent, as reflected in the results, up to differences caused by simulation error. Furthermore, it can be noted that the S-MMNL_{C} model is in fact over-specified; there are multiple solutions for $s_{11,\ln(\beta^*)}$, $s_{21,\ln(\beta^*)}$, $s_{22,\ln(\beta^*)}$ and $\sigma_{\ln(\alpha^*)}$ that give the same covariance for θ, a conclusion that can be reached by noting that only three values are needed to specify the covariance matrix between two coefficients[3]. This issue would always arise when the distribution of $\theta = \alpha^*\beta^*$ in a model of the type in Equation 3 is consistent with the distribution of a directly estimated θ in a model of the type in Equation 2. This point illustrates that when satisfying the condition that the distribution used for θ in the base model is of the same degree of flexibility as that used for $\theta = \alpha^*\beta^*$ in the S-MMNL model, efforts to disentangle scale heterogeneity from heterogeneity in relative sensitivities are in vain. Conversely, if the distribution for θ in the simple MMNL model differs in flexibility from that of $\theta = \alpha^*\beta^*$ in the S-MMNL model, it is impossible to say whether any gains in fit are the result of more flexible distributional assumptions or a sign of an ability to retrieve scale heterogeneity.

5 Summary and conclusions

There has been growing interest of late in the possibility that a large share of the heterogeneity retrieved in random coefficients models relates to variations in absolute sensitivities, i.e. scale heterogeneity, rather than variations in relative sensitivities.

This paper has not set out to discredit the possibility that scale heterogeneity across respondents exists. Our focus has rather been on attempts in the literature to disentangle the two components of heterogeneity, i.e. scale heterogeneity and heterogeneity in individual sensitivities. While the ability to separately identify the two components might be regarded as interesting from a behavioural analysis perspective, we argue that this is not in fact possible in a random heterogeneity context.

Our reasoning is based on two key principles. Firstly, an appropriately specified "standard" Mixed Logit model, in particular one making use of correlated

[3] In the $\mathrm{MMNL_C}$ model, draws from θ_{TT} are obtained as $e^{\mu_{\ln(\theta_{\mathrm{TT}})}+s_{11,\ln(\theta)}\xi_1}$, with draws from β_{TC} being obtained as $e^{\mu_{\ln(\theta_{\mathrm{TC}})}+s_{21,\ln(\theta)}\xi_1+s_{22,\ln(\theta)}\xi_2}$, with ξ_1 and ξ_2 once again giving independent standard normal variates. In the S-$\mathrm{MMNL_C}$ model, draws from α^* are obtained as $e^{\sigma_{\ln(\alpha^*)}\xi_3}$, where ξ_3 is an additional standard normal variate. The draws for $\theta_{\mathrm{TT}} = \alpha^*\beta^*_{\mathrm{TT}}$ are thus given by $e^{\mu_{\ln\left(\beta^*_{\mathrm{TT}}\right)}+s_{11,\ln(\beta^*)}\xi_1+\sigma_{\ln(\alpha^*)}\xi_3}$, while the draws for $\theta_{\mathrm{TC}} = \alpha^*\beta^*_{\mathrm{TC}}$ are given by $e^{\mu_{\ln\left(\beta^*_{\mathrm{TC}}\right)}+s_{21,\ln(\beta^*)}\xi_1+s_{22,\ln(\beta^*)}\xi_2+\sigma_{\ln(\alpha^*)}\xi_3}$. Working on the basis of the underlying normal distribution, we can see that the variance of $\ln(\theta_{\mathrm{TT}})$ is equal to ${s_{11,\ln(\beta^*)}}^2 + {\sigma_{\ln(\alpha^*)}}^2$, while, for $\ln(\theta_{\mathrm{TC}})$, the variance is given by ${s_{21,\ln(\beta^*)}}^2 + {s_{22,\ln(\beta^*)}}^2 + {\sigma_{\ln(\alpha^*)}}^2$. Finally, the covariance between $\ln(\theta_{\mathrm{TT}})$ and $\ln(\theta_{\mathrm{TC}})$ is given by $s_{21,\ln(\beta^*)}s_{11,\ln(\beta^*)} + {\sigma_{\ln(\alpha^*)}}^2$. However, the exact same covariance matrix can be obtained on the basis of a correlated θ alone, as in the $\mathrm{MMNL_C}$ model, with one less parameter, meaning that the S-$\mathrm{MMNL_C}$ model is indeed over-specified.

distributions, is by definition capable of capturing scale heterogeneity alongside heterogeneity in individual coefficients. Secondly, attempts to disentangle the two have been based on making use of models where the marginal utility is given by a product of two random terms, one of which is shared across attributes. This specification leads to a more flexible distributional form, and any gains in fit may be the result of that flexibility in shape rather than an ability to capture scale heterogeneity. In summary, this means that while some of the heterogeneity captured in the "scale" parameter in such models (α in our notation) may indeed relate to scale heterogeneity, there is no way for the analyst to determine whether that is indeed the case, or what share of the heterogeneity that may be.

The observations in this paper apply not just to work on random scale heterogeneity, but also work on WTP space estimation. Such models differ from preference space models only in terms of the ease by which standard distributions can be used within each specification and how the parameters are interpreted. If a model estimated in WTP space fits better on a given data set, this is simply a reflection that the resulting distributional assumptions better match the data being modelled. The same argument applies to comparisons between say the G-MNL model and a model not attempting to include a multiplicative random term shared across coefficients. The G-MNL model (or any model multiplying two random parameters) simply has a greater candidate set of distributions for the marginal utility coefficients. That is to say that all such models are strictly nested and that what is being examined in such comparisons are simply alternative distributional assumptions and the ease by which each parameterisation can accommodate the assumptions. This observation also relates to discussions in McFadden and Train (2000).

As a direction for future research, we encourage analysts interested in scale heterogeneity to attempt to explain such heterogeneity through exogenous means (e.g. Caussade et al., 2005; Dellaert et al., 1999; Hensher et al., 1998; Louviere et al., 2000; Swait and Adamowicz, 2001; Swait and Louviere, 1993) or by making use of additional model components to quantify the role scale parameter (e.g. Hess and Stathopoulos, 2011). Furthermore, it is possible to break the linear correlation and hence unconfound β and α by using a model with utility being non-linear in β; this is however done extremely rarely. With any of these approaches, special care is still required, and, as with any treatment of heterogeneity, there is always a risk that what is captured relates in part to phenomena other than those targeted by the analyst.

6 Acknowledgements

The first author acknowledges the financial support by the Leverhulme Trust, in the form of a *Leverhulme Early Career Fellowship*. The majority of this work was carried out during a visit by the first author to the Institute of Transport and Logistics Studies at the University of Sydney, which was made possible by a Faculty of Economics and Business Visiting Scholar Grant. The authors would like to thank Kenneth Train and Thijs Dekker for valuable feedback on an earlier version of this paper.

References

Ben-Akiva, M., Lerman, S. R., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. MIT Press, Cambridge, MA.

Bhat, C. R., 2001. Quasi-random maximum simulated likelihood estimation of the mixed multinomial Logit model. Transportation Research Part B 35 (7), 677–693.

Breffle, W. S., Morey, E. R., 2000. Investigating preference heterogeneity in a repeated discrete-choice recreation demand model of atlantic salmon fishing. Marine Resource Economics 15, 1–20.

Brownstone, D., Bunch, D. S., Train, K., 2000. Joint Mixed Logit models of stated and revealed preferences for alternative-fuel vehicles. Transportation Research Part B 34 (5), 315–338.

Burge, P., Rohr, C., 2004. DATIV: SP Design: Proposed approach for pilot survey. Tetra-Plan in cooperation with RAND Europe and Gallup A/S.

Caussade, S., Ortúzar, J. de D., Rizzi, L. I., Hensher, D. A., 2005. Assessing the influence of design dimensions on stated choice experiment estimates. Transportation Research Part B 39 (7), 621–640.

Clark, P. K., 1973. A subordinate stochastic process model with finite variance for speculative prices. Econometrica 41 (1), 135–155.

Daly, A., Hess, S., Train, K., 2011. Assuring finite moments for willingness to pay estimates from random coefficients models. Transportation forthcoming.

Dellaert, B., Brazell, J. D., Louviere, J. J., 1999. The effect of attribute variation on consumer choice consistency. Marketing Letters 10 (2), 139–147.

Doornik, J. A., 2001. Ox: An Object-Oriented Matrix Language. Timberlake Consultants Press, London.

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

Fosgerau, M., Bierlaire, M., 2009. Discrete choice models with multiplicative error terms. Transportation Research Part B 43 (5), 494–505.

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.

Halton, J., 1960. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. Numerische Mathematik 2 (1), 84–90.

Hensher, D. A., Greene, W. H., 2003. The Mixed Logit Model: The State of Practice. Transportation 30 (2), 133–176.

Hensher, D. A., Louviere, J. J., Swait, J., 1998. Combining sources of preference data. Journal of Econometrics 89 (1-2), 197–221.

Hess, S., Bierlaire, M., Polak, J. W., 2005. Estimation of value of travel-time savings using mixed logit models. Transportation Research Part A 39 (2-3), 221–236.

Hess, S., Rose, J. M., 2010. Random scale heterogeneity in discrete choice models. paper presented at the 89^{th} Annual Meeting of the Transportation Research Board, Washington, D.C.

Hess, S., Stathopoulos, A., 2011. Linking response quality to survey engagement: a combined random scale and latent variable approach. ITS working paper. Institute for Transport Studies, University of Leeds.

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

Louviere, J. J., Eagle, T., 2006. Confound it! That pesky little scale constant messes up our convenient assumptions. Proceedings of the Sawtooth Software Conference 2006.

Louviere, J. J., Hensher, D. A., Meyer, R. J., Irwin, J., Bunch, D. S., Carson, R., Dellaert, B., Hanemann, W. M., 1999. Combining sources of preference data for modeling complex decision processes. Marketing Letters 10 (3), 205–217.

Louviere, J. J., Hensher, D. A., Swait, J., 2000. Stated Choice Models: Analysis and Application. Cambridge University Press, Cambridge.

Louviere, J. J., Islam, T., Wasi, N., Street, D., Burgess, L., 2008. Designing discrete choice experiments: Do optimal designs come at a price? Journal of Consumer Research 35, 360–375.

Louviere, J. J., Meyer, R. J., 2008. Formal choice models of informal choices: What choice modeling research can (and can't) learn from behavioral theory. Review of Marketing Research 4, 3–32.

Louviere, J. J., Street, D., Carson, R., Ainslie, A., DeShazo, J. R., Cameron, T., Hensher, D. A., Kohn, R., Marley, T., 2002. Dissecting the random component of utility. Marketing Letters 13 (3), 177–193.

McFadden, D., Train, K., 2000. Mixed MNL Models for discrete response. Journal of Applied Econometrics 15 (5), 447–470.

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. American Journal of Agricultural Economics 90 (4), 994–1010.

Swait, J., 2006. Advanced choice models. In: Kanninen, B. (Ed.), Valuing Environmental Amenities Using Stated Choice Studies: A Common Sense Approach to Theory and Practice. Springer, Dordrecht.

Swait, J., Adamowicz, W., 2001. Choice environment, market complexity, and consumer behavior: A theoretical and empirical approach for incorporating decision complexity into models of consumer choice. Organizational Behavior and Human Decision Processes 86 (2), 141–167.

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

Swait, J., Louviere, J. J., 1993. The role of the scale parameter in the estimation and comparison of multinomial logit models. Journal of Marketing Research 30 (3), 305–314.

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

Walker, J., 2001. Extended discrete choice models: Integrated framework, flexible error structures, and latent variables. Ph.D. thesis, MIT, Cambridge, MA.

Yang, M., 2008. Normal log-normal mixture: Leptokurtosis, skewness and applications. Applied Economics Letters 15 (9), 737–742.