Correlation and Scale in Mixed Logit Models
Stephane Hess & Kenneth Train

Online appendix

Following the work of Hess & Rose (2012), the authors of the present paper are not aware of papers in transport research (where that paper was published) affected by the types of misunderstanding and misrepresentation discussed earlier in the present paper. The same is however not true in other disciplines. We conducted a SCOPUS search of all papers published from 2013 onwards (i.e., after Hess and Rose, 2012) that estimated discrete choice models and mention scale heterogeneity, SALC and/or the G-MNL model. A list of the 25 papers we found is shown Table 1, where 14 had a health focus, 9 were in environment, resource & ecology, one in education and one in food studies. Three of the papers (4, 9 and 25) used SALC while the rested used G-MNL.

Table 1: recent papers mentioning scale heterogeneity, SALC and/or G-MNL

ID	Citation	Topic area	Application
1	Baba et al., 2016	food	preferences for beef attributes
2	Börger, 2015	environment, resource & ecology	offshore windfarms
3	Brown et al, 2015	health	support of public vs private hospital care
4	Burke et al., 2015	education	early career teachers' preferences for support
5	Czajkowski et al. 2014a	environment, resource & ecology	forest management strategies
6	Czajkowski et al. 2014b	environment, resource & ecology	waste management
7	Czajkowski et al., 2016	environment, resource & ecology	biodiversity conservation
8	Doiron et al., 2014	health	graduate preferences for working in nursing
9	Erdem and Thompson, 2014	health	health investment
10	Fiebig et al., 2015	health	contraception choices
11	Hall et al., 2013	health	care in terminal illness
12	Ida, 2015	health	smoking behaviour
13	Islam & Meade, 2013	environment, resource & ecology	solar cell adoption
14	Kaambwa et al., 2015	health	consumer directed care
15	Knox et al., 2013	health	contraception choices
16	Kragt, 2013	environment, resource & ecology	catchment areas
17	Lagarde et al., 2013	health	job incentives in Thailand
18	Li et al., 2014a	health	rural location preferences
19	Li et al., 2014b	environment, resource & ecology	refrigerator choice
20	Milte et al., 2014	health	cognitive ability
21	Rezende et al., 2015	environment, resource & ecology	mangrove restoration
22	Rockers et al., 2013	health	preferences for rural postings by nurses
23	Scott et al., 2013	health	GP preferences for rural locations
24	Scott et al., 2015	health	nurses' and midwifes' sensitivities to job characteristics
25	Thiene et al., 2015	environment, resource & ecology	water quality preferences

We will first look at the G-MNL applications.

Ten papers out of 25 (numbers 8, 10, 11, 13, 15, 16, 17, 19, 21, 23) conclude that scale heterogeneity exists in their data, even though the models are incapable of disentangling scale heterogeneity from variation in relative sensitivities. These papers assume that there is no correlation among relative sensitivities, which means that, whatever correlation they estimated through variation in the scale term, the authors attribute, by assumption, to scale heterogeneity. Most of them (10, 11, 15, 17, 23, 13, 21) also do not allow all of the non-scale parameters to vary randomly across respondents.

Five papers (numbers 2, 3, 10, 11, 15) claim to allow for random scale and yet estimated a G-MNL with unscaled alternative specific constants (ASCs). That is, the ASCs were not multiplied by the random scale term. This practice follows the suggestion by Fiebig et al. (2010) not to scale ASCs because they are *"fundamentally different"* from other model components. However, as we noted above, this restriction means that the models do not in fact allow for random scale. They can of course be fully accurate and reliable; they just do not allow for random scale.

Three papers (3, 11, and 15) did not scale some of the non-ASC coefficients, and one did not scale the cost coefficient (paper 3). Again, these models do not allow for random scale, even though the papers claim to do so. A large share of the other papers did not report which coefficients were multiplied by the random scale parameter. The issue of not allowing for random scale while claiming to do so may therefore affect more than just the papers listed above.

In some cases, a misunderstanding of their model structure has led the authors to draw inappropriate conclusions. For example, one paper (15) did not scale its ASCs (which means that the model does not allow for scale heterogeneity) and imposed homogeneity on relative preferences, but concluded that *"explicitly specifying both sources of heterogeneity (taste and scale) in the final model ... gives the best fit for each of the three information conditions and hence supports its use in our analysis".*

The terminology in most of the papers indicates that the authors believe that G-MNL is different from, and more general than, a mixed logit model, and that mixed logit models do not allow for random scale. As discussed above, G-MNL is actually a restricted type of mixed logit, and any mixed logit with correlated coefficients allows for random scale. This confusion about terminology is related to the misinterpretation of results that we identify above: the recognition that scale heterogeneity is identified only by assuming no correlation in relative preferences (i.e., that G-MNL is a restricted form of mixed logits with full covariance among coefficients) prevents the researcher from making these kinds of unsupported conclusions. Examples include quotes such as (in 16) *"This shows that, contrary to the assumption in the [MMNL] model, [μ] should not be normalised across individuals, and that the [G-MNL] specification is the more appropriate model."* Perhaps the worst offender in this context is Li et al. 2014, who state that *"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 (… Hess and Rose 2012)"* where the paper by Hess & Rose (2012) which they cite actually states the opposite.

The emphasis on scale heterogeneity also seems to have affected the specification choices of researchers. Four papers (3, 10, 11, 15) specified models that do not allow for *any* type of heterogeneity in coefficients aside from that induced by random scale. This practice is not necessarily inappropriate if the authors actually believe that scale heterogeneity is the only relevant source of heterogeneity in their data. However, it seems doubtful (to us at least) that researchers thinking about behavioural differences in consumers' choices would reach this conclusion (and without any empirical testing.) It seems more likely that the exclusive interest in scale heterogeneity is driven by the misconception that models like G-MNL that contain a random scale term are an improvement over "mixed logit," without realizing that these models are actually restricted versions of mixed logits that allow general forms of heterogeneity.

Similarly, almost half of all papers (3, 10, 11, 13, 14, 15, 17, 20, 22, 23, 24) impose some form of homogeneity on model parameters outside the multiplication by a random scale term. At least one other paper (18) does not explicitly report which coefficients are treated randomly aside from scale. Homogeneity restriction can be reasonable and appropriate in any given setting. However, as discussed above and emphasized by Hess & Rose (2012), keeping some of the model coefficients fixed inside the multiplication by a random scale term not only leads to reduced distributional flexibility but can also overstate the variance of the random scale term (as this will then *capture* some of the heterogeneity not allowed for in individual coefficients). The wide-spread practice of restricting heterogeneity seems to have arisen from the attempt to estimate scale heterogeneity, which is not identified unless other types of heterogeneity are assumed away. The authors seem not to be aware that MMNL models with full heterogeneity and full covariance among coefficients can be estimated with standard software and do not entail the restrictions embodied in the G-MNL and S-MNL versions of mixed logit.

One paper (5) goes further than most in the specification of the models by allowing for a full covariance matrix between all estimated parameters, both in the MMNL and G-MNL models. Both models thus allow for the different types of heterogeneity. The authors interpret the better fit for G-MNL as evidence that *"the improvement in statistical fit provided by allowing for scale heterogeneity is substantial"* when the improvements should rather be interpreted as a result of different distributional assumptions between MMNL and G-MNL. Another paper (1) also allows for a full correlation structure inside a G-MNL model but offers no comparisons with other structures. The authors however also reach conclusions that scale heterogeneity exists, when we know that it is not possible to disentangle different types of heterogeneity.

A number of papers published after Hess & Rose (2012) attempt to take their findings on board and do not seek to disentangle the different sources of random heterogeneity and rather exploit the G-MNL model to explain scale heterogeneity linked to socio-demographics, by incorporating covariates in the specification of the scale term. Examples include papers 2, 6 and 7. This however seems to create a new source of possible confounding. Indeed, the scale parameters in G-MNL models is typically defined to follow a Lognormal distributions, and the above examples incorporate the socio-demographic interactions on the mean of the logarithm of the scale parameter, e.g. through setting $\mu_n = exp(\mu_{ln-\sigma} + \sigma_{ln-\sigma}\xi + \lambda z_n)$ where ξ is standard Normal and z_n are characteristics of respondent *n*. With a lognormal distribution, the level of random heterogeneity is a function of both the mean and the standard deviation of the logarithm of the parameter, and the model thus also links random 'scale' heterogeneity to covariates, where the findings are potentially again affected by other random heterogeneity. If an analyst seeks to link scale heterogeneity purely to covariates, then approaches other than G-MNL may be more suitable, or alternatively, the random component in $exp(\mu_{ln-\sigma} + \sigma_{ln-\sigma}\xi + \lambda z_n)$ should be set to zero, making the model collapse to a heteroscedastic MMNL model.

Confusion also arises for latent class structures.

One paper (9) compares a SALC model to a MNL 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')".* However, the analysis does not actually provide evidence of different error variances among consumers. The comparison is between a model that assumes no heterogeneity (MNL) and a model (SALC) that allows, but does not disentangle, heterogeneity in scale and preferences. The correct statement is *"there are some people showing different preferences and some of these differences could be due to different error variances."*

Another paper (4) uses only a SALC model in the context of early career teachers' preferences for support. The paper makes statements that *"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"* and *"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."* The statement about the 'general' latent class models is incorrect, as pointed out in Section 2.3.5, and while the SALC model explicitly allows for scale heterogeneity, it similarly cannot disentangle the two sources of heterogeneity, and the interpretations offered in the paper in relation to the scale classes are thus potentially biased.

While another paper (25) acknowledges the *"potentially confounding role between 'scale-classes' and perfectly correlated 'preference-classes', which is an analogue of the confounding that Hess and Rose*

(2012) describe for continuous mixing models", they nevertheless state that their work can *"successfully isolate the three different behavioral issues, and, in fact, we find evidence of all three in the data"*, the third being attribute non-attendance.

A useful additional barometer of the state-of-play comes from an overview of the 250 abstracts submitted to the 5th International Choice Modelling Conference. Numerous abstracts, all of them in health, discuss the issue of scale heterogeneity. This is of course a valid research topic, but we remain concerned about confusion reflected in the following statements (which are only a subset and which we leave unattributed so as to maintain confidentiality of the process).

Several authors give the impression that the MMNL model imposes an absence of scale heterogeneity and that G-MNL is more general, contrary to the discussions in Section 2. Key quotes in this context are:

- *"Despite the importance of the MXL 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 has led to the introduction of generalised multinomial logit model (GML) that accounts for both preference and scale heterogeneity"*
- *"The structure of these models can be further enriched [by using G-MNL], allowing for scale heterogeneity and different distributional assumptions for the parameters"*

where the final comment above also seems to imply that standard MMNL somehow restricts the choice in terms of distributional assumptions.

Similar confusion also remains for latent class models, with:

- *"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). The standard latent class logit model assumes homogenous scale within each class."*

It is firstly not clear at all whether assuming homoscedasticity would lead to bias. More importantly, the final point suggests that a standard latent class model imposes an absence of scale heterogeneity, when it clearly does not (see Section 2.3.5).

The issue extends beyond published papers and extends to the publication process itself. One motivation for us writing this paper is a concern, brought to us by several authors, that reviewers have asked for revisions that seemed unnecessarily limiting. Reviewers have told authors, for example that:

- *"There is an emerging literature on the confounding of preference and scale heterogeneity in mixed logit (and other) models."* Actually, this confounding occurs only in some types of mixed logit models, namely, those that do not allow full correlation among coefficients. Also, this comment neglects the important fact that models that allow for random scale but not other forms of correlation can confound the various sources of correlation, biasing the estimates of scale heterogeneity. Models that allow for random scale but not other sources of correlation remove one form of confounding by creating another.

- *"Failure to account for these differences [in scale] leads to bias in model estimates and incorrect predictions."* The correct statement is *" ... may lead to bias..."* Exploring random scale can be useful, but there is no reason to believe that random scale always needs to be included in a model. As discussed above, models without random scale fit the data better than models with random scale in the majority of datasets examined by Keane and Wasi. More importantly, statements regarding the form that all models must take are generally detrimental to meaningful research.
- *"The generalized mixed logit model accommodates scale heterogeneity."* This statement is correct but limiting, since many forms of mixed logit accommodate scale heterogeneity. Importantly, mixed logits with full covariance among utility coefficients accommodate other sources of correlation in addition to random scale, which G-MNL does not.
- *"Given that under some circumstances the MIXL model can be seen as a special case of the GMNL model, why not use only the GMNL approach?"* As stated above, mixed logit with *uncorrelated* coefficients is a special case of G-MNL, but G-MNL is a special case of mixed logit with correlated coefficients. So if this question is to be asked at all, it needs to be reworded as: "why not use only mixed logits with correlated coefficients?" There are many software packages available for these models. However, as we have said, more general models are not necessarily better given the data and goals of the research. In Keane and Wasi's comparison, models without any correlation fit the data best for the majority of datasets; G-MNL provided the best fit for only one of the ten datasets, and in that one case the next-best model fit nearly as well.
- *"[P]ublished papers that estimate mixed logit models and claim that preference heterogeneity exists ... make the strong assumption of scale homogeneity".* The correct statement should refer to ``... papers that estimate mixed logit models with uncorrelated coefficients ...".

References

Baba, B. Kallas, Z., Costa-Font, M., Gil, J.M & Realini, C.E. 2016, Impact of hedonic evaluation on consumers' preferences for beef attributes including its enrichment with n-3 and CLA fatty acids, Meat Science 111, pp. 9–17.

Brown, P., Panattoni, L., Cameron, L., Knox, S., Ashton, T., Tenbensel, T., Windsor, J., 2015. Hospital sector choice and support for public hospital care in New Zealand: Results from a labeled discrete choice survey. Journal of health economics 43, 118-127.

Börger, T. 2016, Are Fast Responses More Random? Testing the Effect of Response Time on Scale in an Online Choice Experiment, Environ Resource Econ 65(2), pp. 389-413.

Burke, P.F., Aubusson, P.J., Schuck, S.R., Buchanan, J.D., Prescott, A.E. 2015, How do early career teachers value different types of support? A scale adjusted latent class choice model, Teaching and Teacher Education 47, pp. 241-253.

Czajkowski, M. Bartczak, A. Giergiczny, G., Navrud, S. & Żylicz, T. 2014a, Providing preference-based support for forest ecosystem service management, Forest Policy and Economics 39, pp. 1–12

Czajkowskia, M., Kadzielaa, T. & Hanley, N. 2014b, We want to sort! Assessing households' preferences for sorting waste, Resource and Energy Economics 36, pp. 290–306

Czajkowski, M., Hanley, N. & LaRiviere, J. 2016, Controlling for the Effects of Information in a Public Goods Discrete Choice Model, Environmental and Resource Economics 63 (3), pp 523–544.

Doiron, D., Hall, J., Kenny, P., Street, D. J., 2014. Job preferences of students and new graduates in nursing. Applied Economics 46 (9), 924-939.

Erdem, S., Thompson, C., 2014. Prioritising health service innovation investments using public preferences: a discrete choice experiment. BMC health services research 14 (1), 360.

Fiebig, D., M. Keane, J. Louviere, N. Wasi, 2010, "The Generalized Multinomial Logit Model: accounting for Scale and Coefficient Heterogeneity," *Marketing Science*, Vol. 29, pp. 393-421.

Fiebig, D. G., Viney, R., Knox, S., Haas, M., Street, D. J., Hole, A. R., Weisberg, E., Bateson, D., 2015. Consideration sets and their role in modelling doctor recommendations about contraceptives. Health economics.

Hall, J., Kenny, P., Hossain, I., Street, D. J., Knox, S. A., 2013. Providing informal care in terminal illness an analysis of preferences for support using a discrete choice experiment. Medical Decision Making, 0272989X13500719.

Hess, S., and J. Rose, 2012, "Can Scale and Coefficient Heterogeneity be Separated in Random Coefficients Models?" *Transportation*, Vol. 39, No. 6, pp. 1225-1239.

Ida, T., 2014. A quasi-hyperbolic discounting approach to smoking behavior. Health economics review 4 (1), 1-11.

Kaambwa, B., Lancsar, E., McCaffrey, N., Chen, G., Gill, L., Cameron, I. D., Crotty, M., Ratcliffe, J., 2015. Investigating consumers' and informal carers' views and preferences for consumer directed care: A discrete choice experiment. Social Science & Medicine 140, 81-94.

Keane, M., and N. Wasi, 2013, "Comparing Alternative Models of Heterogeneity in Consumer Choice Behavior," *Journal of Applied Econometrics*, Vol. 28, No. 6, pp. 1018-1045.

Knox, S. A., Viney, R. C., Gu, Y., Hole, A. R., Fiebig, D. G., Street, D. J., Haas, M. R., Weisberg, E., Bateson, D., 2013. The effect of adverse information and positive promotion on women's preferences for prescribed contraceptive products. Social Science & Medicine 83, 70-80.

Islam, T. and N. Meade 2013, Impact of Attribute Preferences and Attitudinal Constructs on Adoption Timing: The Case of Solar Photo-Voltaic (PV) Cells for Household Level Electricity Generation, Energy Policy, 55, 521-530.

Lagarde, M., Pagaiya, N., Tangcharoensathian, V., Blaauw, D., 2013. One size does not fit all: Investigating doctors'stated preference heterogeneity for job incentives to inform policy in thailand. Health economics 22 (12), 1452-1469.

Li, J., Scott, A., McGrail, M., Humphreys, J., Witt, J., 2014a. Retaining rural doctors: Doctors' preferences for rural medical workforce incentives. Social Science & Medicine 121, 56-64.

Li, X., Clark, C.D., Jensen, K.L., Yen, S.T. 2014b, Will consumers follow climate leaders? The effect of manufacturer participation in a voluntary environmental program on consumer preferences, Environ Econ Policy Stud (16), pp. 69–87.

Magidson, J., and J. Vermunt, 2005. "Technical Guide to Latent Gold Software 4.5". Statistical Innovation.

Milte, R., Ratcliffe, J., Chen, G., Lancsar, E., Miller, M., Crotty, M., 2014. Cognitive overload? an exploration of the potential impact of cognitive functioning in discrete choice experiments with older people in health care. Value in Health 17 (5), 655-659.

Kragt, M.E. 2013, The Effects of Changing Cost Vectors on Choices and Scale Heterogeneity, Environ Resource Econ 54, pp. 201–221.

Rezende, C.E., Kahn, J.R., Passareli, L. & Vásquez, W.F. 2015, An economic valuation of mangrove restoration in Brazil, Ecological Economics 120, pp. 296–302

Rockers, P. C., Jaskiewicz, W., Kruk, M. E., Phathammavong, O., Vangkonevilay, P., Paphassarang, C., Phachanh, I. T., Wurts, L., Tulenko, K., et al., 2013. Differences in preferences for rural job postings between nursing students and practicing nurses: evidence from a discrete choice experiment in Lao People's democratic republic. Hum Resour Health 11 (1),

Scott, A., Witt, J., Duffield, C., Kalb, G., 2015. What do nurses and midwives value about their jobs? results from a discrete choice experiment. Journal of health services research & policy 20 (1), 31-38.

Scott, A., Witt, J., Humphreys, J., Joyce, C., Kalb, G., Jeon, S.-H., McGrail, M., 2013. Getting doctors into the bush: General practitioners' preferences for rural location. Social Science & Medicine 96, 33-44.

Thiene, M., Scarpa, R. & Louviere, J.J. 2015, Addressing Preference Heterogeneity, Multiple Scales and Attribute Attendance with a Correlated Finite Mixing Model of Tap Water Choice, Environ Resource Econ 62, pp. 637–656.