The Journal of Choice Modelling 7 (2013) 1–12

Contents lists available at SciVerse ScienceDirect

The Journal of Choice Modelling

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

Linking response quality to survey engagement: A combined random scale and latent variable approach

Stephane Hess [a,*], Amanda Stathopoulos [b]

[a] *Institute for Transport Studies, University of Leeds, United Kingdom*
[b] *Transport and Mobility Laboratory, EPFL - Ecole Polytechnique Fédérale de Lausanne, Switzerland*

ARTICLE INFO

Article history:
Received 24 March 2013
Accepted 24 March 2013
Available online 31 May 2013

Keywords:
Latent variables
Survey engagement
Random scale
Stated choice
Hybrid model

ABSTRACT

Recent interest in the topic of random scale heterogeneity in discrete choice data has led to the development of *specialised* tools such as the G-MNL model, as well as repeated claims that studies which fail to separate scale heterogeneity from heterogeneity in individual coefficients are likely to produce biased results. Contrary to this, Hess and Rose (2012) show that separate identification of the two components is not in fact possible in a random coefficients model using a typical linear in parameters specification, and that any gains in performance are potentially just the result of more flexible distributional assumptions. On the other hand, linking scale heterogeneity to measured characteristics of the respondents is likely to yield only limited insights, while using respondent reported measures of survey understanding or analyst captured measures such as survey response time puts an analyst at risk of measurement error and endogeneity bias. The contribution made in this paper is to put forward a hybrid model in which survey engagement is treated as a latent variable which is used to model the values of a number of indicators of survey engagement in a measurement model component, as well as explaining scale heterogeneity within the choice model. This model overcomes some of the shortcomings of earlier work, permitting us to link part of the heterogeneity across respondents to differences in scale, while also allowing us to make use of indicators of survey engagement without risk of endogeneity bias. Results from an empirical application show a strong link between the two model components as well as arguably more reasonable substantive outputs for the choice model component.

1. Introduction

In recent years, there has been extensive interest in the notion that a significant share of the heterogeneity retrieved in random coefficients models relates to variations in scale across respondents, rather than differences in individual sensitivities (see e.g. Louviere et al., 1999, 2002; Swait and Bernardino, 2000; Louviere and Eagle, 2006; Louviere and Meyer, 2008; Louviere et al., 2008). This has generated a desire to separate the two components in estimation, and has motivated the development of specialised model structures, most notably the G-MNL model first proposed by Keane (2006) and operationalised by Fiebig et al. (2010) and Greene and Hensher (2010).

Even early on there was a recognition that dissecting the two components of heterogeneity could be problematic in practice (Louviere et al., 2002). Nevertheless, the discussions and developments in the above mentioned papers have led to

* Corresponding author. Tel.: +44 113 34 36611.
E-mail addresses: s.hess@its.leeds.ac.uk (S. Hess), amanda.stathopoulos@epfl.ch (A. Stathopoulos).

1755-5345/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jocm.2013.03.005

a hype of activity in the area, with repeated claims that result from standard MMNL specifications are not reliable as scale heterogeneity is not "filtered out". However, recent discussions in Hess and Rose (2012) have highlighted important flaws in these claims.

In a standard specification, we have that the modelled utility component is given by $V=\beta'x$, where x is a vector of attributes, and β is a vector of estimated parameters. We allow for x to include sufficient dummy terms to estimate $J-1$ alternative specific constants with J alternatives. In a random coefficients framework, some or all of the elements in β are allowed to follow a random distribution across respondents. Hess and Rose (2012) make the case that such a model directly allows for scale heterogeneity, as long as all elements in β (including constants) are randomly distributed, with the full covariance matrix being estimated.

More importantly, Hess and Rose also show that it is not in fact possible to separately identify the two components of random heterogeneity using a typical linear in parameters specification.[1] In a simple specification aiming to separate out scale heterogeneity, we would have that $V=\theta\beta'x$, where θ is a random scalar which multiplies all elements in β.[2] Hess and Rose show that any gains in model fit obtained by using $V=\theta\beta'x$ instead of $V=\beta'x$ are potentially due to the fact that the distribution of the marginal utility coefficients in the former is more flexible than that in the simple (latter) model. This issue applies to existing uses of the G-MNL model, with both Fiebig et al. (2010) and Greene and Hensher (2010) relying on a specification where a Lognormal θ is multiplied by a Normal β, and contrasting this with a MMNL specification using a Normal β. Hess and Rose show that when an appropriate specification of β is used in the simple MMNL model, namely one that gives the same flexibility as $\theta\beta$ in the *scale heterogeneity* model, the multiplication by θ becomes obsolete. This highlights the inability to separately identify the two components of heterogeneity and puts any advantages of structures such as the G-MNL model down to the flexibility of the parameter distributions.

Despite the above issues, the study of scale heterogeneity remains an interesting topic of research, and conceptually, it would still be desirable to understand the role of scale heterogeneity in overall findings, as well as the key drivers behind it. A substantial body of work has looked at the impact that exogenous variables may have in driving scale heterogeneity, primarily focussing on characteristics of the choice scenarios at hand, often in the context of task complexity (cf. DeShazo and Fermo, 2002; Arentze et al., 2003; Caussade et al., 2005; Swait and Adamowicz, 2001). Relating scale heterogeneity to how information is presented in choice tasks is supported by research hypothesising a link between decision processes and task requirements (Johnson and Payne, 1985) and evidence that people adapt decision strategies to the context, trading accuracy against effort (Payne et al., 1992, 1993).

On the other hand, research has shown that what is commonly described as the capacity-difficulty gap matters more than any absolute definition of complexity (cf. Heiner, 1983). The emphasis is hence not on the impact of the task environment, but on the mental capacity of the respondent, and his/her engagement with the survey. Scale heterogeneity could thus be the result of some respondents not understanding the tasks at hand, not being able to relate to the scenarios faced, or not taking the experiment seriously. The topic of respondent engagement is especially relevant given the increasing reliance on data collected through online surveys, where the analyst has little or no way of guaranteeing that respondents pay adequate attention to the questionnaire. While differences in survey engagement and the resulting scale heterogeneity may be related in part to measured characteristics of the respondent, there remains substantial scope for residual random variation, and this in turn has partly stimulated the above interest in random scale approaches.[3]

Even before the development of random scale heterogeneity models, and more recently their criticism in Hess and Rose (2012), analysts have aimed to *capture* respondent engagement and account for it deterministically. The issue is that survey engagement itself is difficult or impossible to quantify, and analysts have instead relied on proxies. Lundhede et al. (2009) explicitly compare subjective descriptions to externally defined proxies for decision certainty leading to scale differences. Effort may also be proxied by the time required to complete a task as discussed by Klein and Yadav (1989), and Rose and Black (2006) test this by including interactions between response times and random parameter estimates. In other work, Breffle and Morey (2000), Scarpa et al. (2003), and Feit (2009) find that experience with the studied choice context can lead to higher scale.

As already mentioned above, measuring survey engagement is a difficult task. Many surveys collect responses to questions about survey complexity, realism, and understanding. These can give an indication of how well given respondents can understand the survey, relate to the tasks at hand, and how seriously they may have taken the experiment. Similarly, computer based surveys also typically collect data on the time taken to complete the survey. Either type of *indicator* is however arguably not a measure of engagement but a function thereof, and their use as explanatory variables is thus likely to only allow an analyst to capture part of the scale heterogeneity across respondents. More importantly, the likely correlation between these indicators and other unobserved factors can lead to endogeneity bias, while respondent answers to qualitative statements are arguably also subject to measurement error. The situation is analogous to the more general use

[1] Separate estimation could be possible with a specification that is not linear in the parameters of the utility function, but such a specification is rarely used.

[2] It should be noted that in the G-MNL model, a more complex specification is used, where θ has a differential impact on the means and variances in β, but the same arguments apply.

[3] The current paper does not explore the connection between different survey modes and response scale. Similarly, it is possible that systematic respondent features relating to the degree of engagement have a different impact for different data collection methods due to self-selection. These issues are beyond the scope of the present paper.