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

Stephane Hess[*] *Amanda Stathopoulos*[†]

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.

Keywords: latent variables; survey engagement; random scale; stated choice; hybrid model

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

[†]Transport and Mobility Laboratory, EPFL - Ecole Polytechnique Fédérale de Lausanne, amanda.stathopoulos@epfl.ch

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 a hype of activity in the area, with repeated claims that results 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)

[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.

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].

[3] The current paper does not explore the connection between different survey modes

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 of subjective attitudinal data as explanatory variables in discrete choice models, a practice that is increasingly being abandoned in favour of latent variable models (see e.g. Ben-Akiva et al., 1999; Ashok et al., 2002; Ben-Akiva et al., 2002; Bolduc and Alvarez-Daziano, 2010).

An analyst wishing to account for scale heterogeneity caused by potentially different levels of survey engagement across respondents is thus faced with three issues. Firstly, a deterministic treatment relying on exogenous indicators is unlikely to capture all heterogeneity. Secondly, relying on proxies for survey engagement as explanators of scale heterogeneity puts the analyst at risk of endogeneity bias and measurement error, while arguably still not allowing all heterogeneity to be captured. Thirdly, a purely random

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.

approach cannot be used given the discussions in Hess and Rose (2012)[4].

This paper's contribution is to jointly address these three issues. We use a hybrid model structure in which the actual level of engagement that a respondent has with the survey is treated as a latent variable. This latent variable is used inside the choice model component to explain the scale heterogeneity across respondents, and inside a separate measurement equations component to explain the answers to questions relating to survey understanding and complexity, as well as survey response time. As the random term explaining the scale heterogeneity in the choice model is also used in the measurement equations component of the overall structure, we avoid the problems highlighted by Hess and Rose (2012) and are able to separate out scale heterogeneity from heterogeneity in individual coefficients, subject to some further caveats discussed in the next section. We use respondent provided answers to survey understanding and complexity as dependent variables, with the same applying to survey response time. This avoids the issues with endogeneity bias and measurement errors that would affect an approach using the indicators as explanatory variables. Finally, the latent variable has both random and deterministic components (in the form of interactions with respondent covariates), meaning that we recognise the advantage of explaining part of the level of engagement in a deterministic manner, while also allowing for a remaining random component.

The remainder of this paper is organised as follows. The next section outlines the methodology. This is followed in Section 3 by a discussion of the survey work conducted for this analysis. Section 4 presents the results of the empirical analysis, and Section 5 briefly summarises the key findings of the paper.

2 Modelling methodology

Let U_{int} be the utility of alternative i for respondent n in choice situation t ($t = 1, \ldots, T$), made up of a modelled component V_{int} and a remaining random component ε_{int}, which follows a type I extreme value distribution. We hypothesise that scale varies across respondents as a function of survey engagement, and, using a linear in attributes specification of the utility

[4] At this point, it should be noted that Fiebig et al. (2010) also discuss how their *scale* component can be decomposed into a random and deterministic component. However, the issues raised by Hess and Rose still arise for the random component, and any respondent reported or analyst captured measures of engagement must not be used in such a decomposition, which would once again treat them as explanatory rather than dependent variables.

function, we have that:

$$V_{int} = \theta_n \beta_n' x_{int}, \tag{1}$$

where θ_n is the scale parameter for respondent n, β_n is a vector of taste coefficients, and x_{int} is a vector of attributes for alternative i as faced by respondent n in choice task t.

As discussed by Hess and Rose (2012), a separate treatment of random heterogeneity in θ_n is not feasible. On the other hand, deterministically linking scale heterogeneity to measured respondent characteristics or characteristics of the choice task is unlikely to capture the full range of effects. Finally, while respondent reported measures of survey engagement, or analyst captured proxies for survey engagement (e.g. survey completion time) can contain valuable information, they must not be used as explanatory variables for decomposing scale, either in a deterministic or random framework. Indeed, they are not only subject to measurement error, but correlation with other unobserved effects can lead to endogeneity bias.

In the present work, we take account of these issues and recognise that survey engagement itself is not observed (i.e. it is latent) and that either type of indicator (whether respondent reported or analyst captured) is simply a function of this underlying level of engagement. Consequently, we treat respondent engagement as a latent variable, say:

$$\alpha_n = l\left(z_n, \gamma\right) + \eta_n, \tag{2}$$

where $l\left(z_n, \gamma\right)$ represents the deterministic part of α_n, with z_n being a vector of measured covariates related to respondent n, and γ being a vector of estimated parameters. The variable η_n is a random disturbance, which we assume follows a Normal distribution across respondents, with a mean of μ_α and a standard deviation of σ_α. For normalisation, we set $\mu_\alpha = 0$ and $\sigma_\alpha = 1$.

We then replace Equation 1 by:

$$V_{int} = e^{\tau \alpha_n} \beta_n' x_{int}, \tag{3}$$

i.e. we have that $\theta_n = e^{\tau \alpha_n}$, where the estimated parameter τ measures the impact of the latent variable α_n on the scale of utility, and where the exponential ensures positive scale throughout.

In itself, Equation 3 is no different from a random scale model which would be subject to the issues highlighted by Hess and Rose (2012). We now use a standard hybrid model approach by including an additional model component that firstly allows us to address these issues, while also enabling us to make use of additional information such as respondent reported measures of survey engagement or analyst captured survey completion time without exposing ourselves to the risks of measurement error and endogeneity bias.

In particular, let us assume that we have a set of K such indicator variables, which may contain a mixture of ordered indicators (e.g. on a Likert-scale) and continuous indicators. We then aim to explain the observed values for I_{kn}, $k = 1, \ldots, K$ on the basis of α_n.

If I_k is an ordered response, the most appropriate approach would be an ordered model, such as the ordered logit, with the likelihood of the observed value given by:

$$L_{I_{kn}} = \sum_{s=1}^{S} \delta_{(I_{kn}=s)} \left[\frac{e^{\varsigma_{k,s} - \zeta_k \alpha_n}}{1 + e^{\varsigma_{k,s} - \zeta_k \alpha_n}} - \frac{e^{\varsigma_{k,s-1} - \zeta_k \alpha_n}}{1 + e^{\varsigma_{k,s-1} - \zeta_k \alpha_n}} \right], \tag{4}$$

where $\delta_{(I_{kn}=s)}$ is 1 if $I_{kn} = s$ and 0 otherwise, S is the number of levels, $\varsigma_{k,s}$ are estimated threshold parameters, and ζ_k measures the impact of α_n on indicator I_{kn}. For normalisation, we set $\varsigma_{k,S}$ to $+\infty$, and $\varsigma_{k,0}$ to $-\infty$, such that the probability for an indicator value of 1 is given by $\frac{e^{\varsigma_{k,1} - \zeta_k \alpha_n}}{1+e^{\varsigma_{k,1} - \zeta_k \alpha_n}}$ while the probability for an indicator value of S is given by $1 - \frac{e^{\varsigma_{k,S-1} - \zeta_k \alpha_n}}{1+e^{\varsigma_{k,S-1} - \zeta_k \alpha_n}}$.

If on the other hand, I_k is a continuous response, the straightforward method is to work with $I^*_{kn} = I_{kn} - \overline{I_{kn}}$, i.e. centre the indicator on zero, and then using a Normal density, i.e.:

$$L_{I_{kn}} = \frac{1}{\sigma_{I_k}\sqrt{2\pi}} \cdot e^{-\frac{\left(I^*_{kn} - \zeta_k \cdot \alpha_n\right)^2}{2\sigma^2_{I_k}}}, \tag{5}$$

where we estimate σ_{I_k} and ζ_k.

The value of the indicators is then modelled jointly with the actual choice processes, based on the assumption that both processes are at least in part influenced by the latent engagement variable. This approach thus integrates choice models with latent variable models. A main benefit of using a latent variable approach is to overcome the bias inherent in direct incorporation of indicators as explanatory variables in the utility function by instead treating them as dependent variables.

The log-likelihood (LL) function for this model is composed of a number of different components. Firstly, let $L\left(y_n \mid \beta_n, \tau, \alpha_n\right)$ give the likelihood of the observed sequence of choices (y_n) for respondent n, conditional on the vector of taste coefficients β_n, the parameter τ, and the latent variable α_n, which itself is a function of γ. This likelihood will thus be a product of

T discrete choice probabilities, with the specific form depending on model assumptions.

Next, let $L(I_n \mid \zeta_I, \sigma_I, \varsigma_I, \alpha_n)$ give the probability of observing the actual values for the various indicator variables, conditional on the parameter vector ζ_I, the vector of standard deviations σ_I for any continuous indicators, the vector of threshold parameters ς_I for any ordered indicators, and the latent variable α_n, which itself is a function of γ. This likelihood will be given by $L(I_n \mid \cdot) = \prod_{k=1}^{K} L_{I_{kn}}$ where the individual elements in this product potentially make use of a mix of specifications from Equation 4 and Equation 5.

In combination, the LL function across the N respondents is thus given by:

$$LL(\Omega, \gamma, \tau, \zeta_I, \sigma_I, \varsigma_I) = \sum_{n=1}^{N} \ln \int_{\beta} \int_{\eta} L(y_n \mid \cdot)\, L(I_n \mid \cdot)\, g(\eta)\, m(\beta \mid \Omega)\, \mathrm{d}\eta \mathrm{d}\beta, \tag{6}$$

where this is integrated over the distribution of η, the random component in the latent variable, and the randomly distributed vector of taste coefficients β_n, with $\beta_n \sim m(\beta_n \mid \Omega)$, where Ω is a vector of parameters to be estimated. It should be noted that while the likelihood for a single respondent is given by the product of the likelihood of the observed choices and the likelihood of the observed responses to the indicators, the model clearly also incorporates the structural equation of the latent variable, given that both $L(y_n \mid \cdot)$ and $L(I_n \mid \cdot)$ are a function of α_n.

Figure 1 contains an illustration of the proposed model structure, where observed components are shown in rectangles and unobserved components are shown in ellipses. As is clear from the graph, respondent characteristics (socio-demographics) affect both the latent engagement variable and the utility function. The utility is also a function of measured attributes of the alternatives. The latent engagement variable is used to explain the values of the indicators, while it also affects the scale of the utility function in the choice model component.

Before proceeding with the empirical work, it is worth highlighting how the use of the above latent variable structure addresses the key issues discussed earlier, hence illustrating the contribution made by this paper.

Firstly, the model avoids the issues highlighted by Hess and Rose (2012) as the random variable related to scale heterogeneity is used not just in the choice model but also in the separate measurement model explaining the values of the indicators. Crucially, in this second model component, α is used independently of β, which allows the effects to be separated out in the choice model.

Secondly, the model avoids issues with endogeneity bias inherent to earlier applications that have used respondent provided measures of survey engagement or analyst captured proxies such as response time as explanatory variables. This is achieved by treating these variables as dependent rather than explanatory variables, while issues with measurement error are also avoided as the indicators are no longer treated as error free.

Finally, while the model allows the analyst to establish a link between measured characteristics of the respondents (e.g. age and gender) and survey engagement, we recognise that a non-trivial share of survey engagement may relate to intrinsic and unmeasured characteristics, where this is accommodated by the random component in α.

At this point, it is important to recognise that while our proposed approach offers improvements over the simple models criticised by Hess and Rose (2012), there remains the risk that the model may only capture part of the scale heterogeneity in the θ component. Indeed, any scale heterogeneity that cannot be linked to the variation in responses used in the measurement model is likely to still be captured within β. This issue cannot be resolved completely, but further improvements are possible by additional work identifying the most suitable indicators of engagement and carrying out survey work that captures such measures.

3 Survey work

The data used for this study come from a survey looking at commuting by rail and bus, collected through an online panel in the United Kingdom in January 2010. A stated choice (SC) component presented respondents with three work commuting options described by six attributes; travel time, fare, rate of having to stand (out of 10 typical trips), frequency of delays (out of 10 typical trips), average extent of delays, and the availability of an information service alerting travellers to any delay by personal text message (sms). The first of the three alternatives relates to a typical commute trip as reported by the respondent, with attributes held invariant across the 10 choice tasks, while the attributes for the remaining two alternatives are varied according to a D-efficient experimental design pivoted around individual reference values[5]. The design was generated in Ngene[6], and a final sample of 368 respondents

[5] See Bliemer and Rose (2009) for an in-depth discussion of design techniques.

[6] www.choice-metrics.com

was obtained for the present analysis, yielding 3,680 observations.

Given the context of the present study, the survey included several questions probing for subjective descriptions of the level of realism and understanding. In particular, three questions assessed different dimensions of survey involvement. These questions were scored on five-point scales from *do not agree* (1) to *fully agree* (5). Specifically, the three questions used the following wording:

I_1: *"The scenarios I was presented with were realistic"*

I_2: *"I was able to fully understand the tasks I was faced with"*

I_3: *"I was able to make choices as in a real world scenario"*

The answers to these three questions were collected at the end of the survey and thus relate to a respondent's overall impression of the survey. There is however a possibility that respondents may focus especially on understanding/complexity of later tasks rather than earlier tasks. For future work, it would be of interest to contrast this with an approach where answers to these questions are captured at a choice task level.

In addition, we captured respondents' level of agreement with the following statement[7]:

I_4: *"When evaluating a public transport service, I take into account all service characteristics"*

Finally, we also captured the time that a respondent took to complete the survey. A number of limitations arise as we collected the response time for the entire survey, as opposed to the stated choice component alone. This potentially exposes us to the issue that two respondents could have taken the same overall time to complete the survey, but where the split in time between the stated choice component and remaining survey components varied across these two respondents. We are thus implicitly assuming that our measure of engagement relates to the survey as a whole, rather than the stated choice component alone. Future work should investigate the use of response time for the stated choice component alone, or even choice task specific response times. At this point, it is also worth noting that unlike some surveys, the compensation paid to respondents for completing the survey was not linked to the time spent in the survey, meaning that we do not have an incentive incompatibility with this variable.

[7] This is part of a wider set of questions which also included a number of attitudinal statements. Exploratory factor analysis was carried out (see Appendix A) which identified I_4 as the only additional statement of interest for the present study.

4 Empirical analysis

This section presents the findings of our empirical analysis. We first look at model specification and estimation before turning our attention to the empirical results.

4.1 Model specification and estimation

Two different models were estimated in this analysis, namely a simple MMNL model using the stated choice data only[8], and a hybrid structure as outlined in the previous section, explaining both the stated choice data and the five indicators. We now look in turn at the specification of the choice model, the latent variable specification, and the measurement model, before discussing model estimation.

4.1.1 Specification of utility function for choice model

In our final model specification, no significant alternative specific constants were retrieved, and we thus limited ourselves to the effects of six explanatory variables, which are essentially those discussed in Section 3, with the exception that the average delay attribute is multiplied by the rate of delays, leading to an expected delay attribute. It should also be noted that for crowding and the rate of delays, the attributes entered the utility function on a range of 0 to 1 (rather than 0 to 10), while the time attributes were entered in hours rather than minutes.

The models follow the recommendations from Hess and Rose (2012) in that all model parameters were specified to vary randomly across respondents, with a full covariance matrix being estimated for the six elements in β. Furthermore, to avoid discrepancies in flexibility, both β and θ were specified to follow Lognormal distributions. Looking for example at the fare coefficient, we have that $\mu_{\ln(-\beta_{\text{fare}})}$ gives the mean of the underlying Normal distribution (i.e. the logarithm of the negative of the fare coefficient follows a Normal distribution), with $s_{\ln(-\beta_{\text{fare}}),\ln(-\beta_{\text{tt}})}$ and $s_{\ln(-\beta_{\text{fare}})}$ giving the two components of the Cholesky matrix relating to the fare coefficient, the first being off-diagonal, the second being the diagonal element[9].

[8] A comparison with simpler models is not appropriate; indeed, studying the role of scale heterogeneity in the absence of a treatment of heterogeneity in individual coefficients is clearly not possible given the discussions in Hess and Rose (2012).

[9] It should be noted that in order to allow negative covariances, elements within the Cholesky matrix can take on negative values as well as positive values.

Additionally, we incorporated five socio-demographic interactions, introduced as shifts in the means of the underlying Normal distributions. The final specification includes shifts in the sensitivity to travel time for female respondents and respondents who have a car available to them ($\Delta_{\text{female},\mu_{\ln(-\beta_{\text{tt}})}}$ and $\Delta_{\text{car available},\mu_{\ln(-\beta_{\text{tt}})}}$ respectively), shifts in the sensitivity to crowding for respondents aged 50 or over ($\Delta_{\text{age}>50,\mu_{\ln\left(-\beta_{\text{crowding}}\right)}}$), and shifts in the sensitivity to fare and the provision of a delay information service for respondents who do not have a car available to them ($\Delta_{\text{no car available},\mu_{\ln(-\beta_{\text{fare}})}}$ and $\Delta_{\text{no car available},\mu_{\ln\left(\beta_{\text{free delay sms}}\right)}}$ respectively). Efforts to include an income effect were not successful.

4.1.2 Specification of latent variable

For the specification of the latent engagement variable α_n in Equation 2, we conducted an extensive specification search to look into possible socio-demographic interactions. In the final specification, we included interactions with four socio-demographic variables, namely whether a respondent is female, whether a respondent is aged between 35 and 50[10], whether a respondent has a university degree, and whether a respondent currently commutes by train. Associated model parameters are identified as γ_{female}, $\gamma_{\text{age 35-50}}$, $\gamma_{\text{university degree}}$ and $\gamma_{\text{rail traveller}}$. The random component ($\eta_n$) of the latent variable is specified to follow a Normal distribution with a mean of 0 and a standard deviation of 1, where the exponential in Equation 3 ensures that we once again meet the requirements in Hess and Rose (2012) in that both β and θ follow Lognormal distributions in the hybrid model, leading to Lognormal marginal utilities, just as in the MMNL model.

4.1.3 Specification of measurement model

For the measurement model component of the hybrid structure, we use five indicators, namely the respondent provided answers to the four statements described in Section 3, i.e. I_1 to I_4, as well as survey response time. For the first four indicators, an ordered logit specification was used, as detailed in Equation 4. In the base specification, four thresholds were estimated for each of the four indicators (standard normalisation with five levels), but the specific distribution of responses (cf. Figure 1) led to us merging the first two levels for all indicators except I_1). As a further simplification, we found that the estimates for ζ_k in Equation 4 could be constrained to 1 for $k = 1, ..., 4$

[10] Note this is different from the age interaction in the utility functions.

without any significant impact on model fit, where any differential impact of the latent variable on the four indicators is then also incorporated in the estimates for the thresholds. For the final indicator, i.e. survey response time, we used a continuous specification as outlined in Equation 5. Albeit that no extreme outliers were present in the distribution of survey response time, we found that superior performance was obtained by working with the natural logarithm of response time as opposed to its untransformed value.

4.1.4 Log-likelihood and model estimation

All models were coded in Ox (Doornik, 2001), where the simulation based estimation made use of 500 MLHS draws (Hess et al., 2006) per individual and per random component[11].

For the combined model, the log-likelihood function is as outlined in Equation 6, with random variation in both β and α. The first likelihood component $L(y_n \mid \cdot)$ is a product of 10 Logit probabilities. The second likelihood component $L(I_n \mid \cdot)$ is a product of four ordered Logit terms as in Equation 4 (for indicators I_1 to I_4) and one continuous term as in Equation 5 (for survey response time). The distribution $g(\eta)$ is univariate Normal, with a mean fixed at 0 and a standard deviation fixed at 1, while the distribution $m(\beta \mid \Omega)$ is a multivariate Normal, with six elements and a full covariance matrix being estimated. Exponentials are used in the specification of both β and θ to yield Lognormal distributions, with appropriate sign changes for the first five elements in β. The integration over η and β takes place at the level of an individual respondent (rather than observation) such that our specification reflects the repeated choice nature of our data (cf. Revelt and Train, 1998). The estimation of the choice model and measurement model is carried out simultaneously (as shown in Equation 6); it is well known that sequential estimation of hybrid choice models provides only consistency, while simultaneous estimation adds efficiency (cf. Bolduc et al., 2005; Raveau et al., 2010).

For the simple MMNL model, we make use of a version of Equation 6 without the $L(I_n \mid \cdot)$ component, and without integration over η or the multiplication of the utility functions by e^{α_n}.

[11] Sensitivity tests showed no difference in results when increasing the number of draws to $1,000$.

4.2 Results

4.2.1 Estimation results

The estimation results for the two models are presented in two parts. Table 1 presents the parameters for the choice model component only, while Table 2 shows the estimates for the structural equation for the latent variable as well as for the measurement model components. The estimates in Table 2 clearly relate only to the hybrid model as the concerned model components are not used in the simple MMNL model. A comparison of model fit between the two structures is not possible given the differences in the data used in estimation, with the hybrid model jointly explaining choices and indicators. It should also be noted that a number of parameters are not statistically significant, where this relates for example to some of the elements of the Cholesky matrix. However, for the reasons outlined in Hess and Rose (2012), a full covariance matrix needs to be specified in both models, and as such, all parameters are retained.

Looking first at the estimates for the choice model component, it is important to bear in mind that the values shown relate to the parameters of an underlying Normal distribution which applies to the logarithm of the actual coefficients, where, in the case of the first five coefficients, it is the logarithm of the negative of the coefficient. The values for the transformed estimates are looked at later in this section. We note that in both models, the estimate for $\mu_{\ln\left(-\beta_{\text{rate of delays}}\right)}$ is not statistically different from zero, which means that the median of $\beta_{\text{rate of delays}}$ is close to a value of -1. A number of estimates in the Cholesky matrix are similarly not statistically significant, and the actual impact of the estimated values is discussed later in this section when looking at the implied coefficient values.

Looking at the role of the socio-demographic variables, we note that respondents over the age of 50 are more sensitive to crowding, while respondents with no car available tend to be more fare sensitive (likely correlation with income)[12]. We also see that, albeit not statistically significant, respondents who do not have a car available are more sensitive to the provision of a delay information service (which is not surprising as they are more likely to be public transport users).

The final parameter to be considered in the context of the choice model component is τ, which captures the impact of the latent variable α_n on the scale parameter, with $\theta = e^{\tau\alpha_n}$. We note a positive and statistically

[12] The shifts apply to the mean of the underlying Normal distribution, such that increases lead to a more negative estimate for the coefficient, given the sign change.

significant estimate, showing that increases in the latent engagement variable lead to increases in model scale, in line with expectations.

We next turn our attention to the estimates in Table 2, i.e. the parameters of the structural equation for the latent variable, and the parameters of the measurement component of the hybrid model. Looking first at the socio-demographic interactions in Equation 2, we observe a higher value for the latent variable for respondents in the middle age group, and respondents with a university degree. Especially the latter parameter estimate is consistent with intuition when linking it to increased survey understanding.

We finally concentrate on the results for the measurement model. Looking first at the four ordered indicators, the increasing levels for the thresholds, along with the constraint of $\zeta_k = 1$[13] for $k = 1, \ldots, 4$ mean that increases in the latent variable α_n are associated with a higher probability of stronger agreement with the four statements described in Section 3. Similarly, we see a positive estimate for ζ_{I_5}, indicating that increases in the latent variable are also associated with a higher probability of increases in survey response time. The parameter σ_{I_5} meanwhile captures the variance in the logarithm of survey response time.

Overall, these estimates thus show that a respondent with a more positive value for the latent variable α_n is more likely to state that she or he found the survey to be realistic and understandable, was able to make choices as in real life, and generally takes all service characteristics into account. This type of respondent is more likely to be of the middle age group and have a university degree. Additionally, such a respondent is likely to have taken longer to complete the survey, which is arguably a reflection of more careful study of each choice scenario, remembering that no excessively long response times were observed. In making his or her decisions in the stated choice component, such a respondent is also more likely to exhibit behaviour that is more deterministic from the perspective of the analyst, i.e. contains less noise or has higher scale. In conjunction, these observations justify the interpretation of the variable as a latent *engagement* variable.

4.3 Implied distributional patterns for marginal utility coefficients

As a next step, we look at the implied sample level distributions for the marginal utility coefficients. For this, we simulate values for the underlying Normal distributions for each of the six β coefficients, and for every respon-

[13] Remembering that the original estimates were not significantly different from 1.

dent in our sample, using the same draws as those used in estimation, and incorporating the socio-demographic shifts applicable to each given individual. In the hybrid model, we additionally perform this task for θ, once again incorporating any shifts in α applicable to a given individual through the socio-demographic interactions. The resulting draws are then transformed onto a Lognormal scale and Table 3 shows the implied coefficients of variation for β in the MMNL model and the hybrid model. Comparing the heterogeneity in β in the MMNL model to that in $\theta\beta$ in the hybrid model, we note substantial increases in heterogeneity for the travel time and delay sms coefficients, with a large drop in heterogeneity for the rate of delay coefficient, and smaller changes for the remaining three coefficients.

Overall, these findings are an indication that a treatment of scale heterogeneity within a hybrid model can yield substantially different findings in terms of heterogeneity in the choice model component. This would not be possible without the additional measurement model component, given the arguments in Hess and Rose (2012).

It is also of interest to look at the share of the heterogeneity in the marginal utility coefficients that is the result of scale heterogeneity in the hybrid model, where we obtain rates between seven and eight percent, suggesting that the majority of heterogeneity in fact relates to heterogeneity in β. A caveat arises here as some scale heterogeneity may still be captured in β, namely scale variation that is not also reflected in the values for the indicators, a possibility that was highlighted in the earlier discussions.

4.3.1 Sample level WTP distributions

As a final step, we now look at the implied sample level willingness-to-pay (WTP) distributions, making use of the individual-specific sets of draws produced in the previous subsection. The random scale component introduced by the $e^{\tau\alpha_n}$ multiplier in the hybrid model clearly has no impact on the WTP patterns given that all coefficients are affected in the same way. As a result, differences between the two models are purely the result of any impacts that the inclusion of this additional variable has on the remaining model parameters, and in particular differential impacts on individual marginal utility coefficients.

The actual results of these calculations are summarised in Table 4, showing the mean WTP across respondents, along with the coefficient of variation. It should first be noted that the WTP measures coming out of this analysis are relatively low. This is however in line with the low average journey cost reported by respondents and the frequency of journeys, as well as the comparatively small time savings that were presented in the survey.

Comparing the mean WTP estimates between the two models, we see reductions in four of the measures, with increases only for the WTP to avoid delays. The drop is especially marked for the WTP for reductions in expected delay; the ratio between the mean WTP to reduce expected delay and reduce travel time is now of the order of around 3, which is more realistic than the MMNL results, and more consistent with previous stated preference results (cf. Hollander, 2005).

Looking finally at the heterogeneity patterns in the WTP measures, we see reductions in the coefficient of variation for each of the five measures. This suggests an overestimation of the heterogeneity in WTP measures in the MMNL model, which could obviously have detrimental impacts if the model were to be used to provide outputs for policy evaluation.

5 Conclusions

The topic of variations in model scale across respondents has created extensive interest across a number of disciplines in which random utility models are used to study individual behaviour. In recent years, this has led to the development of *specialised modelling* tools that purport to be able to separate out random scale heterogeneity from heterogeneity in individual coefficients, most notably the G-MNL model (Keane, 2006; Fiebig et al., 2010). However, recent work by Hess and Rose (2012) shows that it is not in fact possible to separately identify the two components of heterogeneity within a random coefficients framework using a linear in parameters specification, and that any advantages of models such as G-MNL are simply the result of more flexible distributional assumptions.

An alternative approach to the modelling of scale heterogeneity places the emphasis on a deterministic treatment. In this context, a number of authors have linked scale to observable and externally defined characteristics relating to the choice task (cf. DeShazo and Fermo, 2002; Arentze et al., 2003; Caussade et al., 2005; Swait and Adamowicz, 2001), generally under the guise of complexity. However, an additional, and potentially more important source for scale heterogeneity can be found in differences across respondents in their understanding as well as engagement with the survey. The key issue raised in the present paper is that it is not possible for an analyst to capture exact measures of respondent understanding/engagement. What's more, the indicators typically used to study understanding or engagement are potentially correlated with other unobserved factors that play a role in decision making. In conjunction, this means that such indicators should not be used as explanatory variables for decomposing scale as this would put an analysis at risk of measurement error and endogeneity bias. Additionally, such a deterministic treatment is arguably likely to only account for part of the variation in survey engagement across respondents.

The contribution made by this paper is to jointly address the above issues through formulating a hybrid model structure. Our proposed structure treats survey engagement as a latent variable which is used to explain the values of a number of indicators (questions relating to survey understanding and realism, as well as survey response time), as well as the scale heterogeneity in the choice model. By treating survey engagement as a latent variable and making use of the indicators as dependent rather than explanatory variables, we avoid issues with measurement error and endogeneity bias inherent to more deterministic approaches. We link survey engagement to a number of socio-demographic characteristics, but additionally allow for intrinsic differences that cannot be explained by measured characteristics, through the random component included in the specification of the latent variable, thus overcoming another shortcoming of a deterministic approach. Finally, the key limitation highlighted by Hess and Rose (2012), i.e. the separate identification of heterogeneity in β and θ (using the notation from this paper), is addressed by using the random component of θ inside the measurement model in addition to the choice model.

The empirical results from our analysis seem to indicate a clear link between our two model components. Indeed, we observe that increases in the latent engagement variable lead to a greater probability of agreement with statements relating to survey understanding and realism, heightened probability of longer survey response time, and increased scale within the choice model component. We are able to partly link engagement to measured characteristics of the respondent, with a higher value for respondents aged between 35 and 50, as well as respondents with a university education. While these results are of interest from a behavioural perspective, it is also worth noting that the hybrid model leads to different, and in our view more realistic findings in terms of the implied sample level distribution of individual sensitivities, and crucially also willingness-to-pay measures.

As is the case with most work, it is important to highlight a number of limitations, as well as areas for future research.

Firstly, it should again be noted that the model is likely to only capture part of the scale heterogeneity in the θ component, with scale heterogeneity that cannot be linked to the indicators still being captured within β. An important area for work in this context is be the development of more appropriate indicators of respondent engagement.

Secondly, our survey captured all indicators at the respondent level rather than the choice task level. For survey response time, a further limitation applies in that this variable is measured at the overall survey level rather than the stated choice component alone. Future work should additionally look at collecting values for these indicators at the level of individual choice tasks. This would not only provide additional information, but could also allow a link to be made with choice task characteristics. Additionally, it may increase our ability to separate out the two components of heterogeneity, with scale heterogeneity now having a task level component. On the other hand, issues with separating within respondent heterogeneity from between respondents heterogeneity may arise, and computational cost will clearly increase further still (cf. Hess and Train, 2011). Finally, it would also be of interest to combine and contrast the approach used in the present paper with one that explains scale heterogeneity on the basis of choice task characteristics, such as the work of Caussade et al. (2005). This was not possible with the data at hand where the number of alternatives and attributes was kept fixed across respondents.

Acknowledgments

The first author acknowledges the support of the *Leverhulme* Trust in the form of a *Leverhulme Early Career Fellowship*. The second author acknowledges the support of a Trieste University scholarship (M.U.R. - Progetto Giovani Ricercatori).

A Appendix: Structure of latent variables and indicators

This appendix describes the procedure used to identify the set of indicators that are observable manifestations of the latent variable α_n. Exploratory statistical analysis was employed to assess reliability and internal associations of the indicators used to represent the latent variable. The three indicators, drawn from the survey questions regarding involvement and understanding (I_1, I_2 and I_3), were taken as a point of departure. A control concerning the internal consistency was carried out using Cronbach's alpha based on pairwise correlations between the indicators, I_1, I_2 and I_3. The value for Cronbach's alpha was very high (0.914), indicating a large correspondence between the responses to the three survey questions related to engagement and understanding, thus confirming the reliability of using these as a common construct.

Respondent reported levels of agreement with ten further statements relating to real life commuting behaviour and attitudes (as opposed to questions related to the survey) were also collected. As a second step, *exploratory factor analysis* was then carried out to assess which of the combined set of answers to the thirteen statements were possibly linked to the underlying α_n factor. The prior assumption of high correspondence between the three engagement indicators (I_1 to I_3) and a strong link to α_n was confirmed by factor analysis where models hypothesising $2-6$ factors for the 13 indicators were compared based on the χ^2 statistic of overall fit. Factor analysis was carried out in R using varimax rotation of the factors. The factor loadings for I_1 to I_3 were consistently between 0.82 to 0.96, indicating that these indicators accounted for a very large proportion of variance in the latent variable and had a high degree of communality. The remaining indicator with the consistently highest factor loading was the level of agreement with the statement that a respondent evaluated options based on all trip characteristics (I_4). Again, Cronbach's alpha indicated a good degree of association among the four indicators ($\alpha = 0.811$).

A second round of *confirmatory factor analysis* was carried out in LISREL hypothesising the latent variable to be the underlying factor behind the four indicators. The confirmatory factor analysis (cf. Jöreskog and Sörbom, 1996) based on the suggested measurement model of the four identified indicators showed that the null hypothesis of perfect model fit for the population could not be rejected ($\chi^2 = 1.71$ with a p-value of 0.425 and 2 df). All four indicators had significant loadings on the latent variable ranging from 0.17 (with a t-ratio of 4.14) for I_4 up to a loading of 1 (t-ratio of 24) for I_3. The four indicators accounted for 60% of the variance in α_n.

Aside from the four indicators, and drawing on prior studies concerning the links between survey duration and engagement, a further indicator is included among the measurement equations, namely the logarithm of survey response time.

References

Arentze, T., Borgers, A., Timmermans, H., DelMistro, R., 2003. Transport stated choice responses: effects of task complexity, presentation format and literacy. Transportation Research Part E 39 (3), 229–244.

Ashok, K., Dillon, W. R., Yuan, S., 2002. Extending discrete choice models to incorporate attitudinal and other latent variables. Journal of Marketing Research, 31–46.

Ben-Akiva, M., Walker, J., Bernardino, A., Gopinath, D., Morikawa, T., Polydoropoulou, A., 2002. Integration of choice and latent variable models. In: Mahmassani, H. (Ed.), In Perpetual motion: Travel behaviour research opportunities and application challenges. Pergamon, Ch. 13, pp. 431–470.

Ben-Akiva, M., Walker, J., McFadden, D., Gärling, T., Gopinath, D., Bolduc, D., Börsch-Supan, A., Delquié, P., Larichev, O., Morikawa, T., Polydoropoulou, A., Rao, V., 1999. Extended framework for modeling choice behavior. Marketing Letters 10 (3), 187–203.

Bliemer, M. C. J., Rose, J. M., 2009. Designing stated choice experiments: State-of-the-art. In: Kitamura, R., Yoshi, T., Yamamoto, T. (Eds.), The Expanding Sphere of Travel Behaviour Research, Selected Papers from the 11th International Conference on Travel Behaviour Research. Emerald Group Publishing, pp. 459–498.

Bolduc, D., Alvarez-Daziano, R., 2010. On estimation of hybrid choice models. In: Hess, S., Daly, A. (Eds.), Choice Modelling: The State-of-the-Art and the State-of-Practice: Proceedings from the Inaugural International Choice Modelling Conference. Emerald Group Publishing, pp. 259–288.

Bolduc, D., Ben-Akiva, M., Walker, J., Michaud, A., 2005. Hybrid choice models with logit kernel: Applicability to large scale models. In: Lee-Gosselin, M., Doherty, S. (Eds.), Integrated Land-Use and Transportation Models: Behavioural Foundations. Elsevier, Oxford, pp. 275–302.

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), 1–20.

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.

DeShazo, J. R., Fermo, G., 2002. Designing choice sets for stated preference methods: the effects of complexity on choice consistency. Journal of Environmental Economics and Management 44 (1), 123–143.

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

Feit, E. M., 2009. Extending the generalized multinomial logit model: Error scale and Decision-Maker characteristics. Working paper series, The Wharton School, Philadelphia.

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.

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.

Heiner, R., 1983. The origin of predictable behavior. American Economic Review 73, 560–595.

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

Hess, S., Train, K., 2011. Recovery of inter- and intra-personal heterogeneity using mixed logit models. Transportation Research Part B 45 (7), 973–990.

Hess, S., Train, K., Polak, J. W., 2006. On the use of a Modified Latin Hypercube Sampling (MLHS) method in the estimation of a Mixed Logit model for vehicle choice. Transportation Research Part B 40 (2), 147–163.

Hollander, Y., 2005. The Attitude of Bus Users to Travel Time Variability. Paper presented at the European Transport Conference, Strasbourg.

Johnson, E. J., Payne, J. W., Apr. 1985. Effort and accuracy in choice. Management Science 31 (4), 395–414.

Jöreskog, K. G., Sörbom, D., 1996. LISREL 8 user's reference guide. Scientific Software International.

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

Klein, N. M., Yadav, M. S., 1989. Context effects on effort and accuracy in choice: An enquiry into adaptive decision making. Journal of Consumer Research: An Interdisciplinary Quarterly 15 (4), 411–21.

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., 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.

Lundhede, T. H., Olsen, S. B., Jacobsen, J. B., Thorsen, B. J., 2009. Handling respondent uncertainty in choice experiments: Evaluating recoding approaches against explicit modelling of uncertainty. Journal of Choice Modelling 2 (2), 118–147.

Payne, J. W., Bettman, J. R., Johnson, E. J., 1992. Behavioral decision research: A constructive processing perspective. Annual Review of Psychology 43 (1), 87–131.

Payne, J. W., Bettman, J. R., Johnson, E. J., 1993. The adaptive decision maker. Cambridge University Press.

Raveau, S., Alvarez-Daziano, R., Yáñez, M., Bolduc, D., Ortúzar, J. de D., 2010. Sequential and simultaneous estimation of hybrid discrete choice models: some new findings. Transportation Research Record 2156, 131–139.

Revelt, D., Train, K., 1998. Mixed logit with repeated choices: Households' choices of appliance efficiency level. Review of Economics and Statistics 80, 647–657.

Rose, J. M., Black, I. G., 2006. Means matter, but variance matter too: Decomposing response latency influences on variance heterogeneity in stated preference experiments. Marketing Letters 17 (4), 295–310.

Scarpa, R., Ruto, E. S. K., Kristjanson, P., Radeny, M., Drucher, A. G., Rege, J. E. O., 2003. Valuing indigenous cattle breeds in kenya: an empirical comparison of stated and revealed preference value estimates. Ecological Economics 45, 409–426.

Swait, J., Adamowicz, W., 2001. The influence of task complexity on consumer choice: A latent class model of decision strategy switching. Journal of Consumer Research 28 (1), 135–148.

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

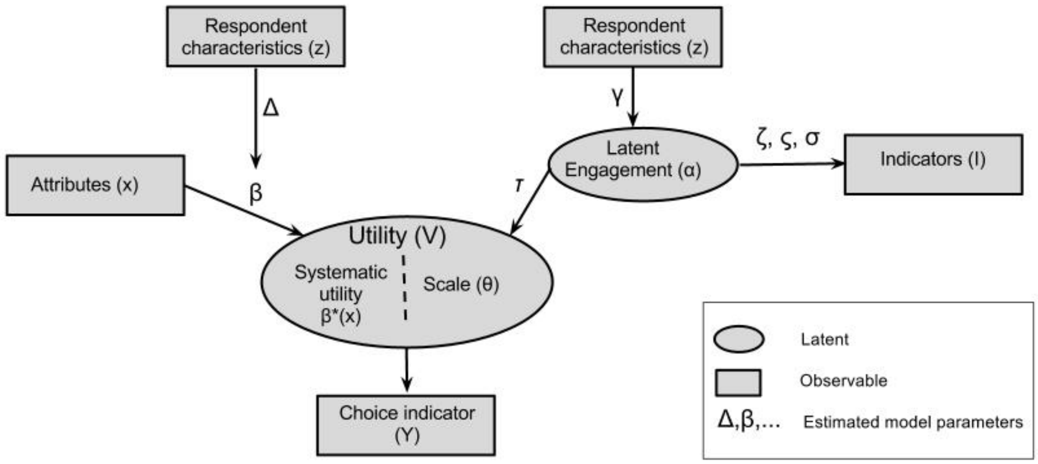

Fig. 1: Structure of latent engagement model

Tab. 1: Estimation results (part I)

	MMNL		Hybrid	
Respondents	368		368	
Observed choices	3,680		3,680	
Observed indicator measurements	0		1,840	
Log-likelihood	-3,020.82		-4,907.77	
	est.	t-rat.	est.	t-rat.
$\mu_{\ln(-\beta_{\text{tt}})}$ (hours)	1.2702	7.10	1.1069	5.42
$\mu_{\ln(-\beta_{\text{fare}})}$ (£)	0.8926	5.82	0.7357	4.77
$\mu_{\ln(-\beta_{\text{crowding}})}$ $(0-1)$	0.5717	2.43	0.5814	3.40
$\mu_{\ln(-\beta_{\text{rate of delays}})}$ $(0-1)$	-0.3172	-0.78	-0.1296	-0.35
$\mu_{\ln(-\beta_{\text{exp. delay}})}$ (hours)	1.6960	5.60	1.4133	4.27
$\mu_{\ln(\beta_{\text{free delay sms}})}$	-1.5652	-4.76	-1.7735	-5.30
$s_{\ln(-\beta_{\text{tt}})}$	1.3600	9.29	1.4191	9.87
$s_{\ln(-\beta_{\text{fare}}),\ln(-\beta_{\text{tt}})}$	0.9919	7.70	0.9641	7.86
$s_{\ln(-\beta_{\text{fare}})}$	-1.3906	-14.17	-1.3022	-14.58
$s_{\ln(-\beta_{\text{crowding}}),\ln(-\beta_{\text{tt}})}$	0.8857	6.59	0.8924	6.68
$s_{\ln(-\beta_{\text{crowding}}),\ln(-\beta_{\text{fare}})}$	-0.2099	-1.99	-0.1243	-1.47
$s_{\ln(-\beta_{\text{crowding}})}$	-1.8154	-13.19	-1.7397	-18.57
$s_{\ln(-\beta_{\text{rate of delays}}),\ln(-\beta_{\text{tt}})}$	1.2407	6.93	1.1190	5.62
$s_{\ln(-\beta_{\text{rate of delays}}),\ln(-\beta_{\text{fare}})}$	0.0706	0.43	0.1553	1.14
$s_{\ln(-\beta_{\text{rate of delays}}),\ln(-\beta_{\text{crowding}})}$	-1.2582	-13.38	-1.2266	-13.94
$s_{\ln(-\beta_{\text{rate of delays}})}$	1.4716	8.69	1.4035	9.59
$s_{\ln(-\beta_{\text{exp. delay}}),\ln(-\beta_{\text{tt}})}$	0.8094	5.40	0.7999	5.05
$s_{\ln(-\beta_{\text{exp. delay}}),\ln(-\beta_{\text{fare}})}$	-0.0755	-0.60	-0.2467	-1.38
$s_{\ln(-\beta_{\text{exp. delay}}),\ln(-\beta_{\text{crowding}})}$	-0.2215	-1.39	0.1213	1.02
$s_{\ln(-\beta_{\text{exp. delay}}),\ln(-\beta_{\text{rate of delays}})}$	-1.2426	-7.98	-1.1779	-6.65
$s_{\ln(-\beta_{\text{exp. delay}})}$	1.0080	7.98	1.0174	9.75
$s_{\ln(\beta_{\text{free delay sms}}),\ln(-\beta_{\text{tt}})}$	-0.5883	-3.08	-0.5248	-2.24
$s_{\ln(\beta_{\text{free delay sms}}),\ln(-\beta_{\text{fare}})}$	-0.1178	-0.91	-0.0412	-0.27
$s_{\ln(\beta_{\text{free delay sms}}),\ln(-\beta_{\text{crowding}})}$	0.2612	2.10	0.1995	1.84
$s_{\ln(\beta_{\text{free delay sms}}),\ln(-\beta_{\text{rate of delays}})}$	-0.4163	-3.17	-0.4886	-3.83
$s_{\ln(\beta_{\text{free delay sms}}),\ln(-\beta_{\text{exp. delay}})}$	1.1916	8.76	1.1762	6.82
$s_{\ln(\beta_{\text{free delay sms}})}$	-0.5024	-2.41	-0.6943	-4.62
$\Delta_{\text{female},\mu_{\ln(-\beta_{\text{tt}})}}$	0.0937	0.64	-0.0062	-0.04
$\Delta_{\text{age} > 50,\mu_{\ln\left(-\beta_{\text{crowding}}\right)}}$	0.7104	2.98	0.7060	4.57
$\Delta_{\text{car available},\mu_{\ln(-\beta_{\text{tt}})}}$	0.0622	0.40	0.0613	0.36
$\Delta_{\text{no car available},\mu_{\ln(-\beta_{\text{fare}})}}$	0.5051	3.67	0.6193	4.74
$\Delta_{\text{no car available},\mu_{\ln\left(\beta_{\text{free delay sms}}\right)}}$	0.1881	0.76	0.2925	1.08
τ	-		0.3797	4.61

Tab. 2: Estimation results (part II)

	est.	t-rat.
γ_{female}	0.1255	0.88
$\gamma_{\text{age 35-50}}$	0.2975	1.97
$\gamma_{\text{university degree}}$	0.4788	3.30
$\gamma_{\text{rail traveller}}$	-0.0970	-0.68
$\varsigma_{I_1,1}$	-2.8801	-11.20
$\varsigma_{I_1,2}$	-2.0679	-9.46
$\varsigma_{I_1,3}$	-0.6212	-3.27
$\varsigma_{I_1,4}$	2.7248	12.01
$\varsigma_{I_2,1\&2}$	-2.6294	-10.83
$\varsigma_{I_2,3}$	-1.4007	-6.92
$\varsigma_{I_2,4}$	1.2233	6.36
$\varsigma_{I_3,1\&2}$	-2.6557	-10.92
$\varsigma_{I_3,3}$	-1.1740	-5.94
$\varsigma_{I_3,4}$	1.7897	8.94
$\varsigma_{I_4,1\&2}$	-2.6208	-10.49
$\varsigma_{I_4,3}$	-0.0606	-0.32
$\varsigma_{I_4,4}$	3.1302	12.59
ζ_{I_5}	0.0822	2.90
σ_{I_5}	0.5113	26.64

Tab. 3: Heterogeneity in individual coefficients

	β (MMNL)	β (HYBRID)	$\theta \cdot \beta$ (HYBRID)	change	part due to θ
β_{tt}	2.32	2.55	2.77	+19.40%	7.97%
β_{fare}	4.34	3.78	4.08	-5.99%	7.28%
β_{crowding}	8.05	6.95	7.49	-6.96%	7.16%
$\beta_{\text{rate of delays}}$	14.16	10.77	11.6	-18.08%	7.16%
$\beta_{\text{exp. delay}}$	5.02	4.71	5.07	+1.00%	7.18%
$\beta_{\text{free delay sms}}$	2.97	3.24	3.5	+17.85%	7.49%

Tab. 4: Sample level WTP distributions

MEAN			
	MMNL	HYBRID	change
travel time reduction (£/hr)	£3.61	£2.98	-17.48%
standing in one fewer train out of 10 (£)	£0.68	£0.67	-1.59%
one fewer train delayed out of 10 (£)	£0.47	£0.54	+15.10%
expected delay reduction (£/hr)	£ 15.73	£8.99	-42.84%
free delay information system (£)	£1.51	£1.37	-9.40%

COEFFICIENT OF VARIATION			
	MMNL	HYBRID	change
travel time reduction (£/hr)	2.75	2.55	-7.22%
standing in one fewer train out of 10 (£)	11.07	9.77	-11.76%
one fewer train delayed out of 10 (£)	20.32	17.55	-13.60%
expected delay reduction (£/hr)	9.04	6.17	-31.76%
free delay information system (£)	20.68	19.82	-4.15%