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

Stephane Hess*   Amanda Stathopoulos†

Abstract

Surveys aimed at conducting choice modelling analyses routinely include questions about perceptions and attitudes. While researchers have occasionally included responses to such questions in their models in a deterministic fashion, it is well known that this can lead to endogeneity bias. This has led to a growing popularity for latent variable approaches that jointly model the response to stated choice tasks and attitudinal questions. Such hybrid frameworks have proved a fruitful approach for explaining differences across individual respondents in sensitivities. At the same time, a separate stream of research has started to openly question the nature of the taste heterogeneity retrieved in Mixed Logit analyses, showing that at least part of the heterogeneity retrieved in such models is in fact scale heterogeneity, i.e. variation in absolute rather than relative sensitivities. In the present paper, we combine these two approaches. Specifically, we hypothesise that differences across respondents in survey engagement, understanding or attention result in differences in response quality, expressed as scale heterogeneity. We model this through a random scale approach that interacts with a latent variable model. Here, we find clear evidence of a link between this latent variable, model scale, and the response to various questions about survey realism and complexity. The resulting model is able to better represent the observed choices, while also leading to noticeable differences in the retrieved heterogeneity patterns.

Keywords: latent variables; survey engagement; random scale; stated choice; choice experiment

*Institute for Transport Studies, University of Leeds, s.hess@its.leeds.ac.uk, Tel: +44 (0)113 34 36611
†Department of Statistical and Economic Sciences, University of Trieste, Italy, amandairini.blombergstathopoulos@phd.units.it
1 Introduction

Although a growing majority of choice modelling studies are now based on the use of stated preference (SP) data, concerns remain about response quality in hypothetical scenarios (see e.g. Carlsson and Martinsson, 2001; Hensher, 2010). In particular, critics question the assumption that the behaviour observed in such surveys is consistent with that in real world choice scenarios. An increasing reliance on data collected through online surveys, where the analyst has little or no way to guarantee respondents pay adequate attention to the questionnaire, is likely to further compound this issue. The question is whether respondents are able and/or willing to engage carefully with each of the choice scenarios they are faced with. In practice, low response quality would equate to observed choices which are difficult to explain on the basis of the set of attributes used to describe the choice tasks. This would manifest itself through low model scale, i.e. a larger impact for the random component of the choice models used to analyse the data. This could be the result of respondents not understanding the tasks at hand, not being able to relate to the scenarios faced, or not taking the experiment seriously.

A significant body of research has looked at the impact of the choice environment on response quality. Such work builds on early applied research hypothesising a link between decision processes and task requirements. Johnson and Payne (1985) indicate that the cognitive decision process that respondents will use in experiments are contingent upon task conditions. A body of experimental research suggests that people adapt decision strategies to the context, trading accuracy against effort (Payne et al., 1992, 1993). People will dedicate effort both in view of task requirements (evaluating the difficulty) and the actual benefit that can be obtained by choosing well (relevance or realism). In line with these insights, an important strand of recent literature looks at the impact of complexity on choice behaviour (Arentze et al., 2003; Rose et al., 2009; Caussade et al., 2005; Hensher, 2006; Swait and Adamowicz, 2001). An essential feature of these surveys is the exogeneity of the treatment of complexity effects; that is, complexity is assumed to depend on how information is presented in the choice sets. Results from this body of research show an impact of the number of alternatives, attributes and attribute levels on response quality in the shape of model scale\(^1\).

At odds with the focus on exogenous task-complexity in the above mentioned studies, a different view focuses on personal features or abilities of

\(^{1}\text{If say an increase in complexity has a decreasing effect on the scale parameter, this would equate to a greater influence of the error on choices (i.e. a less deterministic choice process).} \)
respondents. Research has shown that what is commonly described as the
capacity-difficulty gap matters more than any absolute definition of complex-
ity (cf. Heiner, 1983). From this point of view, complexity should arguably
not be be defined by the analyst, as is typically the case in applied research.
The emphasis is thus not on the impact of the task environment, but on the
mental capacity of the respondent, and his/her engagement with the survey.

In comparison with the above discussion, the literature in this area is
much more limited. An early model by Garbarino and Edell (1997) finds that
alternatives that are more difficult to evaluate cause respondents to generate
negative affect and decreases the choice probability compared to ‘easier’ al-
ternatives. The effect is particularly pronounced for respondents with lower
(perceived) skills. DeShazo and Ferino (2004) formulate a rationally-adaptiv-
model, assuming that when faced with overwhelming complexity, individuals
adjust their limited cognitive ability to assess only a limited dimension of the
task. Furthermore, the amount of effort dedicated to solve real-life problems
such as choosing among complex pricing schemes may depend on subjective
engagement.

Bonsall et al. (2007) hypothesise that survey engagement is linked to
increased effort to make more optimal choices. On the other hand, disen-
gagement with a hypothetical task is associated with decreased effort or the
use of simplified choice rules. Huneke et al. (2004) find that motivation and
experience with the decision context leads to increased effort in a complex
multi-attribute choice. Additionally, both externally defined and subjective
complexity is shown to increase the probability of selecting status quo alter-
natives, where these are typically viewed as easier to assess and comprehend
(see e.g. Meyerhoff and Liebe, 2009; Moon et al., 2005).

In other work, Brefle and Morey (2000), Scarpa et al. (2003), and Feit
(2009) find, using different approaches, that experience with the studied
choice context can lead to higher scale. 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). This is
quite separate from discussions about the impact of survey duration where
the authors control for scale changes in responses to a succession of tasks to
control for fatigue. Here, Bradley and Daly (1994) find evidence for a drop
in scale following the fifth task while contrasting findings are given in an
extensive recent study by Hess et al. (2010) where fatigue does not appear
across choice tasks.

Looking instead at the time a given respondent takes to complete the sur-
vey is arguably a better indicator of engagement since it seizes an individual-
level effort that the respondent dedicates to the task. Here, Rose and Black (2006) include interactions between response times and random parameter estimates and find a link between response time and the mean and variance of random parameters.

Measuring survey engagement is a difficult task facing analysts in this context. Many surveys collect responses to questions about survey complexity, realism, and understanding. These can give an indication of how well a given respondent can understand the survey and relate to the tasks faced with, and how seriously they may have taken the experiment but are arguably not a measure of engagement but a function thereof. Similarly, computer based surveys also typically collect data on the time taken to complete the survey. This may once again give an indication of engagement, although response time is arguably also not a measure of survey engagement, but is itself a direct result of the degree of survey engagement.

Directly including such indicators in model estimation, in the form of responses to attitudinal questions, questions about experience, or data on response time, has contributed towards establishing links between respondent features, attitudes, engagement and response behaviour. However, such a simple deterministic treatment is hazardous, given the likely correlation between these indicators and other unobserved factors, possibly leading to biased results. The situation is analogous to the more general use of subjective attitudinal data as explanators in discrete choice models. Here, there has been growing interest of late in latent attitude methods, with extensive discussions in Ben-Akiva et al. (1999), Ben-Akiva et al. (2002), Bolduc and Alvarez-Daziano (2010), and Ashok et al. (2002), and applications for example in Johansson et al. (2006), Fosgerau and Bjørner (2006), Daly et al. (2010), Hess and Beharry-Borg (2010), and Yáñez et al. (2010). These models recognise the fact that an analyst does not observe attitudes but only captures responses to attitudinal questions alongside the stated choice behaviour. In these models, a latent variable is thus constructed that helps explain both the response to attitudinal questions and the behaviour in the stated choice component of the survey. While there is a growing literature on the impact of latent variables in mode, vehicle and route choice literature, the incorporated variables concern attitudes and perceptions on behalf of individuals of unmeasured attributes, i.e. quality, safety or environmental benefit of a chosen alternative. Instead, as discussed above, when considering attitudes/perceptions toward the survey instrument and its overall content, modellers typically rely on external and naive approaches.

In the present paper, we propose to treat respondents’ engagement with a survey as a latent variable. This latent variable is used to model the values
of a number of indicators that are proxies for survey engagement. Moreover, the latent variable is used to explain variations across respondents in model scale. This latent specification of the engagement variable avoids the risk of endogeneity bias that would arise in a deterministic treatment. Additionally, the fact that the scale parameter in the resulting model now has a random component links our paper to recent work on random scale heterogeneity. This emerging body of work makes the case for an alternative interpretation of the heterogeneity typically retrieved in random coefficients models by showing that a sizeable part of the variations may in fact be due to heterogeneity in absolute sensitivities rather than relative sensitivities (see e.g. Swait and Bernardino, 2000; Louviere et al., 2002). A number of modelling approaches have been proposed to accommodate such random scale heterogeneity (cf. Breffle and Morey, 2000; Hess and Rose, 2009; Fiebig et al., 2010). However, while these models are able to capture scale heterogeneity caused by differences in survey engagement, no attempts are made to make the specific linkage with respondent-specific survey engagement. The model used in the present paper attempts to explain such scale heterogeneity on the basis of differential survey engagement, captured by a latent variable that also drives the values of the various deterministic indicators of engagement.

The remainder of this paper is organised as follows. The following section outlines the methodology put forward in this paper. This is followed in Section 3 by a discussion of the survey work conducted for this analysis. Section 4 presents the findings 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 $t$, made up of a modelled component $V_{int}$ and a remaining random component $\varepsilon_{int}$, which follows a type I extreme value distribution with variance $\pi^2/6$. We thus have:

$$U_{int} = V_{int} + \varepsilon_{int},$$

(1)

The modelled component $V_{int}$ is a function of the estimated vector of sensitivities $\beta$ and the attributes of the alternative $i$, such that:

$$V_{int} = f(\beta, x_{int}),$$

(2)

where typically, a linear in parameters specification is used, i.e. $f(\beta, x_{int}) = \beta'x_{int}$. 
It is well known that model scale is inversely proportional to the variance of the error term and that the scale \( \mu \) and the vector of coefficients \( \beta \) are not separately identifiable. In other words, we are actually estimating \( \mu \beta \), where an increase in scale of the model can be accommodate either by an increase in \( \mu \) or an increase in \( \beta \), and where for identification purposes we thus typically set \( \mu = 1 \).

In the present context however, we hypothesise that scale varies across respondents as a function of survey engagement, and thus rewrite our modelled utility function as:

\[
V_{\text{int}} = \mu_n \beta' x_{\text{int}}
\]  

(3)

In work looking at a deterministic treatment of scale heterogeneity, we would have that \( \mu_n = g (\gamma' z_n) \), i.e. scale is a function of an individual specific vector of attributes \( z_n \) and an estimated vector of parameters \( \gamma \) (cf. Feit, 2009). Similarly, scale may be choice task rather than respondent specific, i.e. \( \mu_t = g (\gamma' z_t) \) where \( z_t \) contains characteristics of the current choice task (cf. Caussade et al., 2005).

It may similarly be tempting to include respondent reported measures of survey engagement, or other proxies for survey engagement such as survey completion time (cf. Rose and Black, 2006). This approach however not only ignores possible measurement error but also puts the analyst at risk of producing biased results due to correlation between these indicators and the unobserved components of utility (cf. Ben-Akiva et al., 2002). On the other hand, a purely random treatment of scale heterogeneity (cf. Breffle and Morey, 2000; Hess and Rose, 2009; Fiebig et al., 2010), i.e. \( \mu_n \sim h (\mu_n) \) with appropriate normalisation, does not allow us to make use of additional self reported information. This is the motivation for the approach set out below.

In the present work, we acknowledge that any responses to questions relating to survey complexity or understanding are a function of survey engagement rather than a direct measure of survey engagement. The same applies to survey completion time. As a result, we treat respondent engagement as a latent variable. In particular, we have that:

\[
V_{\text{int}} = e^{\tau \alpha_n \beta' x_{\text{int}}},
\]

(4)

where we once again work on the basis of a linear interaction between \( \beta \) and \( x_{\text{int}} \). In this model, the parameter \( \tau \) measures the impact of the latent variable \( \alpha_n \) on the scale of the model.

In itself, Equation 4 is no different from a random scale model if no additional attempts are made to explain \( \alpha_n \). As a first step, we now define
\[ \alpha_n = l(z_n, \gamma) + \eta_n, \] (5)

where \( l(z_n, \gamma) \) represents the deterministic part of \( \alpha_n \), with \( z_n \) being a vector of socio-demographic variables of respondent \( n \), and \( \gamma \) being a vector of estimated parameters. The term \( \eta_n \) is a random disturbance, which we assume follows a Normal distribution across respondents, with a mean of \( \mu_\alpha \) and a standard deviation of \( \sigma_\alpha \).

In order to make use of any additional information such as survey completion time or responses to questions relating to survey complexity or understanding, we now combine our choice model with a measurement model in an integrated choice and latent variable framework (cf. Ben-Akiva et al., 1999; Ashok et al., 2002; Ben-Akiva et al., 2002; Bolduc and Alvarez-Daziano, 2010). In a model looking at the impact of latent variables, we make use of a number of indicators that serve as proxies for these latent variables, typically in the form of responses to attitudinal questions. The value of these indicators is then modelled jointly with the actual choice processes, based on the assumption that both processes are at least in part influenced by latent attitudes. This approach thus integrates choice models with latent variable models resulting in an improvement in the understanding of preferences as well as explanatory power. A main benefit of using a latent variable approach is to overcome the bias inherent in direct incorporation of indicators of attitudes (or other subjective measures) in the utility function. What is more, problems with measurement errors can be overcome by looking at a set of factors that have their origin in a latent variable, rather than a simple one-to-one correspondence.

The latent variable modelling framework is divided into two sets of equations. The structural equations refer to cause-and-effect relations between observed features of the choice and decision-maker which define latent variables. The measurement equations refers to regressions from a set of indicator variables to dependent variables.

The latent variable equation (Equation 5) and the utility function (Equation 4) give the structural equations. The final component of the model is given by the measurement equations for the indicator variables. In particular, we have that the value for the \( k^{th} \) indicator for respondent \( n \) is modelled as:

\[ I_{kn} = \delta I_k + \zeta I_k \cdot \alpha_n + v_{kn}, \] (6)

where \( \delta I_k \) is a constant for the \( k^{th} \) indicator, \( \zeta I_k \) is the estimated effect of
the latent variable $\alpha_n$ on this indicator, and $\nu_{kn}$ is a normally distributed disturbance, with a mean of zero and a standard deviation of $\sigma_{I_k}$.

As mentioned above, the indicators are typically responses to attitudinal questions, with a finite number of possible values (e.g. scale of 1 to 5). The use of a continuous specification despite the discrete nature of the outcomes for the indicator variables is common practice, where the use of an ordered model is an important area for future developments (cf. Daly et al., 2010).

To avoid the estimation of unnecessary parameters, the mean of each indicator can be subtracted from the original indicator variables prior to model estimation, thus meaning that all indicators are centred on zero, obviating the need to estimate $\delta_{I_k} \forall k$. Finally, for identification reasons, the standard deviation of the latent variable $\alpha$ needs to be fixed, i.e. we set $\sigma_\alpha = 1$.

The log-likelihood (LL) function for this model is composed of a number of different components. Firstly, let $L(y_n \mid \beta, \tau, \alpha_n)$ give the likelihood of the observed sequence of choices for respondent $n$ ($y_n$), conditional on the vector of taste coefficients $\beta$, the interaction parameter $\tau$, and the latent variable $\alpha_n$, which itself is a function of $\gamma$. This likelihood will thus be a product of discrete choice probabilities, with the specific form depending on model assumptions.

Next, let $L(I_n \mid \zeta_I, \sigma_I, \alpha_n)$ give the probability of observing the specific responses given by respondent $n$ to the various attitudinal questions, conditional on the parameter vector $\zeta_I$, the vector of standard deviations $\sigma_I$ (grouping together $\sigma_{I_k}$, $k = 1, \ldots K$), and the latent variable $\alpha_n$, which itself is a function of $\gamma$. It can be seen that this probability is given by a product of Normal density functions, i.e.

$$L(I_n \mid \zeta_I, \sigma_I, \alpha_n) = \prod_{k=1}^{K} \phi(I_{kn}),$$

(7)

where:

$$\phi(I_{kn}) = \frac{1}{\sigma_{I_k} \sqrt{2\pi}} e^{-\frac{(I_{kn}-\zeta_{I_k} \alpha_n)^2}{2\sigma_{I_k}^2}}.$$  

(8)

In combination, the LL function is thus given by:

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

(9)

where this is integrated over the distribution of $\eta$, the random component in the latent variable, and where $n = 1, ..., N$ is the index over respondents.
The dependencies of $L(y_n \mid \cdot)$ and $L(I_n \mid \cdot)$ on the various parameters are not shown explicitly for ease of notation. In addition, it should be noted that while the likelihood for a given 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 $\alpha_n$.

In addition to the parameters for the standard model, the use of this model thus entails the estimation of the interaction parameter $\tau$, the parameters of the measurement equations $\zeta_{Ik} \forall k$, the socio-demographic interaction terms $\gamma$ used in the structural equation for the latent variable, and the standard deviations of the normally distributed $v_{kn}$ terms (having normalised the standard deviation of $\eta_n$, i.e. $\sigma_\alpha$ to 1). If we have additional random heterogeneity in the $\beta$ coefficients, additional layers of integration need to be added, and we would have:

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

where $\beta \sim m(\beta \mid \Omega)$ with $\Omega$ being a vector of parameters to be identified.

An illustration of the proposed model structure is given in Figure 1 where observed components are shown in rectangles and unobserved components are shown in ellipses. As shown in 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. The latent engagement variable drives the response to the indicator questions, while it also affects the scale which enters the utility function and thus affects the probabilities and hence the choice.

3 Survey work

The data employed is drawn 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) experiment 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 of any delay by personal text message. 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 (see Bliemer and Rose, 2009, for an in-depth discussion of design techniques). The design was generated in Ngene\(^2\). A final sample of 368 respondents 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 \textit{do not agree} (1) to \textit{fully agree} (5). Specifically, the three questions used the following wording:

\begin{itemize}
  \item \textit{I1: “The scenarios I was presented with were realistic”}
  \item \textit{I2: “I was able to fully understand the tasks I was faced with”}
  \item \textit{I3: “I was able to make choices as in a real world scenario”}
\end{itemize}

The survey also collected responses to ten questions relating to attitudes and perceptions. Exploratory factor analysis was carried out on the above questions as well as these attitudinal questions to assess potential correspondence with the latent engagement factor, as discussed in Appendix A. Along with \textit{I1} to \textit{I3}, the sole attitudinal indicator selected based on the factor loadings was the following:

\begin{itemize}
  \item \textit{I4: “When evaluating a public transport service, I take into account all service characteristics”}
\end{itemize}

A summary of the distribution of responses to these four questions is shown in Figure 2.

4 \textbf{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.

\(^2\) www.choice-metrics.com
4 Empirical analysis

4.1 Model specification and estimation

Four different models were estimated in this analysis. The first two models had an underlying MNL structure, with the difference between the models being the incorporation of the latent engagement variable in the second model (MNL&LV). The third and fourth models are models that additionally allow for random taste heterogeneity in the individual marginal utility coefficients, where once again, the fourth model (MMNL&LV) additionally makes use of the latent engagement variable.

4.1.1 Specification of underlying utility function

The specification for the utility function of the two MNL models is shown in Equation 11. Here, we estimate alternative specific constants (ASC) for the first two alternatives ($\delta_1$ and $\delta_2$), with the ASC for the third alternative being normalised to zero. We estimate linear effect of increases in travel time ($\mu_{TT}$), while strong evidence of non-linearity in fare sensitivity led us to use the natural logarithm of fare, with the associated coefficient $\mu_{L-FARE}$. Additional effects relate to increases in the rate of having to stand ($\mu_{CROWDING}$), the rate of trains being delayed ($\mu_{RELIABILITY}$), the expected delay ($\mu_{EXP\cdot DELAY}$), with the attribute obtained by multiplying the rate of delays with the average delays, and the provision of a delay information service ($\mu_{INFO}$).

$$V_{int} = \delta_i + \mu_{TT} \cdot x_{TT,i} + \mu_{L-FARE} \cdot x_{L-FARE,i} + \mu_{CROWDING} \cdot x_{CROWDING,i} + \mu_{RELIABILITY} \cdot x_{RELIABILITY,i} + \mu_{EXP\cdot DELAY} \cdot x_{EXP\cdot DELAY,i} + \mu_{INFO} \cdot x_{INFO,i} + \Delta_{TT,\text{female}} \cdot x_{TT,i} \cdot z_{\text{female,n}} + \Delta_{CROWDING,\text{age > 50}} \cdot x_{CROWDING,i} \cdot z_{\text{age > 50,n}} + \Delta_{TT,\text{car available}} \cdot x_{TT,i} \cdot z_{\text{car available,n}} + \Delta_{L-FARE,\text{no car available}} \cdot x_{L-FARE,i} \cdot z_{\text{no car available,n}} + \Delta_{INFO,\text{no car available}} \cdot x_{INFO,i} \cdot z_{\text{no car available,n}}$$

(11)

Additionally, five socio-demographic interactions were found to be significant, with shifts in sensitivity to travel time for female respondents and re-
spondents who have a car available to them ($\Delta_{TT,\text{female}}$ and $\Delta_{TT,\text{car available}}$ respectively), shifts in sensitivity to crowding for respondents aged 50 or over ($\Delta_{\text{CROWDING, age > 50}}$), and shifts in sensitivity to the log of fare and the provision of an information service for respondents who do not have a car available to them ($\Delta_{\text{L.FARE, no car available}}$ and $\Delta_{\text{INFO, no car available}}$ respectively). Efforts to include an income effect were not successful. In each case the additional multiplier indicates whether a specific socio-demographic condition applies, e.g., $z_{\text{female, n}}$ is 1 if and only if respondent $n$ is female.

In the Mixed Multinomial Logit (MMNL) models, we tested for random taste heterogeneity in all six marginal utility coefficients, but found no additional heterogeneity in the cost coefficient after making use of the log transform. A fixed cost coefficient was thus used, along with randomly distributed coefficients for travel time, crowding, the rate of delays, the expected delay, and the provision of a free information service. For the sake of simplicity, and given the exploratory nature of this paper, a multivariate Normal distribution was used for these five coefficients. In particular, the utility function was rewritten as:

$$V_{int} = \delta_i + \mu_{TT} \cdot x_{TT,i} + \mu_{L\text{-FARE}} \cdot x_{L\text{-FARE},i} + \mu_{\text{CROWDING}} \cdot x_{\text{CROWDING},i} + \mu_{\text{RELIABILITY}} \cdot x_{\text{RELIABILITY},i} + \mu_{\text{EXP. DELAY}} \cdot x_{\text{EXP. DELAY},i} + \mu_{\text{INFO}} \cdot x_{\text{INFO},i} + \Delta_{TT,\text{female}} \cdot x_{TT,i} \cdot z_{\text{female,n}} + \Delta_{\text{CROWDING, age > 50}} \cdot x_{\text{CROWDING},i} \cdot z_{\text{age > 50,n}} + \Delta_{TT,\text{car available}} \cdot x_{TT,i} \cdot z_{\text{car available,n}} + \Delta_{\text{L.FARE, no car available}} \cdot x_{\text{L.FARE},i} \cdot z_{\text{no car available,n}} + \Delta_{\text{INFO, no car available}} \cdot x_{\text{INFO},i} \cdot z_{\text{no car available,n}} + s_{1,1} \cdot \xi_1 \cdot x_{TT,i} + (s_{2,1} \cdot \xi_1 + s_{2,2} \cdot \xi_2) \cdot x_{\text{CROWDING},i} + (s_{3,1} \cdot \xi_1 + s_{3,2} \cdot \xi_2 + s_{3,3} \cdot \xi_3) \cdot x_{\text{RELIABILITY},i} + (s_{4,1} \cdot \xi_1 + s_{4,2} \cdot \xi_2 + s_{4,3} \cdot \xi_3 + s_{4,4} \cdot \xi_4) \cdot x_{\text{EXP. DELAY},i} + (s_{5,1} \cdot \xi_1 + s_{5,2} \cdot \xi_2 + s_{5,3} \cdot \xi_3 + s_{5,4} \cdot \xi_4 + s_{5,5} \cdot \xi_5) \cdot x_{\text{INFO},i},$$

(12)

where $\xi_1, \ldots, \xi_5$ are normally distributed random terms with a mean of zero.
and a standard deviation of 1, and where the 15 additional parameters \( (s_{j,k}) \) are the diagonal and off-diagonal terms for the Cholesky matrix.

### 4.1.2 Specification of latent variable

For the specification of the latent engagement variable \( \alpha_n \) in Equation 5, we conducted an extensive specification search to look into possible significant 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\(^3\), whether a respondent has a university degree, and whether the respondent currently commutes by train. This thus gives us:

\[
\alpha_n = \gamma_{\text{female}} \cdot z_{\text{female},n} + \gamma_{\text{age \ 35-50}} \cdot z_{35\leq \text{age} \leq 50,n} + \gamma_{\text{graduate}} \cdot z_{\text{graduate},n} + \gamma_{\text{train}} \cdot z_{\text{train},n} + \eta_n, \tag{13}
\]

where \( \eta_n \sim N(\mu_\alpha, \sigma_\alpha) \), with \( \sigma_\alpha = 1 \).

### 4.1.3 Specification of measurement model

The measurement model employed five indicators as dependent variables. In addition to the respondent reported indicators \( I_1 \) to \( I_4 \) discussed in Section 3, we made use of the time \( (RT) \) that a respondent took to complete the survey, given our expectation that this was strongly linked to survey engagement. After normalising the mean of the five indicator variables to zero, our specification for the measurement model in Equations 6 thus becomes:

\[
\begin{align*}
I_1 &= \zeta_{I_1} \cdot \alpha_n + v_{1n} \\
I_2 &= \zeta_{I_2} \cdot \alpha_n + v_{2n} \\
I_3 &= \zeta_{I_3} \cdot \alpha_n + v_{3n} \\
I_4 &= \zeta_{I_4} \cdot \alpha_n + v_{4n} \\
RT &= \zeta_{RT} \cdot \alpha_n + v_{5n},
\end{align*}
\tag{14}
\]

where \( v_{1n}, \ldots, v_{5n} \) are normally distributed random variables with a mean of zero and estimated standard deviations \( (\sigma_{I_1}, \sigma_{I_2}, \sigma_{I_3}, \sigma_{I_4}, \text{ and } \sigma_{RT}) \).

\(^3\) Note this is different from the age interaction in the utility functions.
4.1.4 Combined model specification and estimation

Using the utility specification from Equation 11, the log-likelihood function for the first MNL model (without the latent engagement variable) is given by:

\[
LL(\beta) = \sum_{n=1}^{N} \sum_{t=1}^{T} \ln \left( \frac{e^{V_{cnt}}}{\sum_{j=1}^{J} e^{V_{jnt}}} \right),
\]

(15)

where \( V_{cnt} \) refers to the modelled utility of the alternative (out of \( J \)) chosen by respondent \( n \) (out of \( N \)) in choice situation \( t \) (out of \( T \)). Here, \( \beta \) groups together the six main marginal utility coefficients, the five socio-demographic interaction terms, and the two ASCs.

Using the specification from Equation 14, we can then see (with the help of Equation 7) that the likelihood of the observed values for the five indicator variables is given by:

\[
L(I_n | \zeta_I, \sigma_I, \alpha_n) = \prod_{k=1}^{5} \phi(I_{kn}),
\]

(16)

where \( \zeta_I = (\zeta_{I1}, \zeta_{I2}, \zeta_{I3}, \zeta_{I4}, \zeta_{RT}) \), \( \sigma_I = (\sigma_{I1}, \sigma_{I2}, \sigma_{I3}, \sigma_{I4}, \sigma_{RT}) \), and where, for convenience, we have used \( I_5 \) to refer to \( RT \).

In the model incorporating the latent engagement variable, the probability of observing the sequence of choices made by respondent \( n \) is now given by:

\[
L(y_n | \beta, \tau, \alpha_n) = \prod_{t=1}^{T} \left( \frac{e^{\mu_n V_{cnt}}}{\sum_{j=1}^{J} e^{\mu_n V_{jnt}}} \right),
\]

(17)

which is thus conditional on the values for the latent variable \( \alpha_n \), given that \( \mu_n = e^{\tau \alpha_n} \). In turn, with the help of Equation 9, we can see that the function to maximise is now given by:

\[
LL(\beta, \gamma, \tau, \zeta_I, \sigma_I) = \sum_{n=1}^{N} \ln \int_{\eta} L(y_n | \beta, \tau, \alpha_n) L(I_n | \zeta_I, \sigma_I, \alpha_n) g(\eta) \, d\eta,
\]

(18)

where we use integration over the random component of the latent variable, i.e. \( \eta \), and where this integration is carried out at the level of an individual respondent rather than an individual choice, thus accommodating the panel nature of the data (Revelt and Train, 1998).
In the simple MMNL specification without the latent engagement variable, we have that $\beta$ is distributed according to $m(\beta | \Omega)$, allowing for some fixed elements in $\beta$, namely the ASCs, the log-fare coefficient, and the five socio-demographic interaction terms. The probability of observing the sequence of choices made by respondent $n$ is now given by:

$$L(y_n | \Omega) = \int \prod_{t=1}^{T} \left( \frac{e^{V_{cnt}}}{\sum_{j=1}^{J} e^{V_{jnt}}} \right) m(\beta | \Omega) \, d\beta,$$

with modelled utilities $V_{jnt}$ given by Equation 12, and the overall function to maximise is simply given by $\sum_{n=1}^{N} \ln (L(y_n | \Omega))$.

In the MMNL specification additionally incorporating the latent engagement variable, this now becomes conditional on $\alpha_n$, and we obtain:

$$L(y_n | \Omega, \tau, \alpha_n) = \int \prod_{t=1}^{T} \left( \frac{e^{\mu_n V_{cnt}}}{\sum_{j=1}^{J} e^{\mu_n V_{jnt}}} \right) m(\beta | \Omega) \, d\beta. \tag{20}$$

The combined function to maximise is now given by:

$$LL(\Omega, \gamma, \tau, \zeta_I, \sigma_I)$$

$$= \sum_{n=1}^{N} \ln \int_{\eta} \int_{\beta} L(y_n | \Omega, \tau, \alpha_n) L(I_n | \zeta_I, \sigma_I, \alpha_n) g(\eta) m(\beta | \Omega) \, d\beta \, d\eta,$$

$$\tag{21}$$

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. The estimation of the choice model and measurement model is carried out simultaneously (as shown in Equation 18 and Equation 21) for reasons of efficiency.

### 4.2 Results

#### 4.2.1 Model fit

As a first comparison, Table 1 reports the model fit for the four model structures. The models with latent variables are not directly comparable to those without, given the joint maximisation of the choice model and the measurement model. As a result, we also show the portion of the log-likelihood that corresponds to the choice model component only, conditional on the final
4 Empirical analysis

parameter estimates. This corresponds to:

\[
LL(y_1, \ldots, y_N | \Omega^*, \gamma^*, \tau^*) = \sum_{n=1}^{N} \ln \int_{\eta} \int_{\beta} L(y_n | \Omega^*, \tau^*, \alpha_n) g(\eta) m(\beta | \Omega^*) d\beta d\eta.
\]  

(22)

where \(\Omega^*, \gamma^*\) and \(\tau^*\) give the final parameter estimates for \(\Omega, \gamma\) and \(\tau\) from maximising Equation 21, and where the role of \(\gamma^*\) in computing \(\alpha_n\) is not shown explicitly. Similarly, it is possible to compute the portion of the log-likelihood that corresponds to the measurement equations component only, conditional on the final parameter estimates. This corresponds to:

\[
LL(I_1, \ldots, I_n | \gamma^*, \zeta^*, \sigma^*_I) = \sum_{n=1}^{N} \ln \int_{\eta} L(I_n | \zeta^*, \sigma^*_I, \alpha_n) g(\eta) d\eta,
\]

(23)

with \(\zeta^*_I\) and \(\sigma^*_I\) giving the final estimates for \(\zeta_I\) and \(\sigma_I\) from maximising Equation 21. Here, it can be seen that, after integrating out \(\eta\), the two components are uncorrelated, such that the summation of Equation 22 and Equation 23 indeed yields the final estimate obtained from maximising Equation 21. In other words, the combined log-likelihood from the two components of sequential estimation at the final estimates from simultaneous estimation yields the same LL value; the actual LL obtained with sequential estimation would however be different.

Comparing the simple MNL model with the MNL model with the additional latent variable, we see an improvement in the LL relating to the choice model component only by 13.12 units, which is highly significant coming at the cost of just one additional parameter, \(\tau\). A very similar improvement in fit is obtained when moving from the simple MMNL model to the MMNL model with the latent variable, with a gain by 12.44 units, again for the estimation of one additional parameter.

The simple MMNL model offers an improvement over its MNL counterpart by 161.88 units, at the cost of 15 additional parameters, which is significant beyond the 99% level. A similar improvement is obtained when comparing the MMNL and MNL models with the additional latent variable, with gains by 156.88 units in the overall model fit, and 161.20 units in the choice model component only, each time at the cost of 15 additional parameters, where these improvements are highly significant. These results thus show that both the incorporation of random taste heterogeneity and the inclusion of scale variations as a function of the latent engagement variable lead to important gains in model fit, where these are rather stable whether
the two types of heterogeneity are accommodated jointly or separately. It is also worth noting that the fit for the measurement component in the MMNL model with the latent variable is slightly lower than in the MNL model with the latent variable. This could be an indication that the weight of the measurement model in simultaneous estimation is reduced in the presence of additional random taste heterogeneity.

4.2.2 Estimation results

The main estimation results for the four models are reported in three separate tables. Table 2 shows the estimation results for the choice model component of the four models, Table 3 shows the results for the latent variable model, and Table 4 shows the results for the measurement model. For Table 3 and Table 4, results are only shown for the two models which incorporate the additional latent variable. Throughout, the notation from earlier sections is used.

Starting with the MNL model, we observe a significant propensity to choose the status quo alternative, as well as some left-to-right reading effects given the positive estimate for $\delta_2$. All six main attribute effects are highly significant and of the expected sign. We additionally observe that female respondents have a higher sensitivity to travel time, that respondents over the age of 50 are more sensitive to crowding (significant at the 91% level), that respondents who have a car available tend to have higher travel time sensitivity, while respondents with no car available tend to be more sensitive (likely correlation with income), and also more sensitive (significant at the 94% level) to the provision of a delay information service (which is not surprising as they are more likely to be public transport users).

With 61% of respondents being female, 31.5% being aged over 50, and 49% usually having a car available, we can note that the average respondent has a sensitivity to expected delay that is just over twice as high as the sensitivity to travel time, and that the sensitivity to an increase in the rate of having to stand by 1 train out of 10 is 50% higher than the sensitivity to an increase in the number of delayed trains by 1 out of 10.

For the MNL&LV model, we observe a positive and significant estimate for $\tau$, which indicates that, with $\mu_n = e^{\tau\alpha_n}$, an increase in the latent variable $\alpha_n$ leads to an increase in scale for respondent $n$. The results for the choice model component of the MNL model are only in part affected by the inclusion of the latent engagement variable. This is somewhat expected, given that this variable is interacted with a multiplicative scale factor that affects all model parameters in the same manner. Nevertheless, we do ob-
serve an increase in the travel time coefficient ($\mu_{TT}$) by 18% while the other main effects coefficients remain unaffected. Changes are however also observed for all five socio-demographic interaction terms. Here, we see reductions in $\Delta_{TT,\text{female}}$, $\Delta_{\text{CROWDING, age}>50}$, and $\Delta_{TT,\text{car available}}$, but increases in $\Delta_{\text{LFARE, no car available}}$ and $\Delta_{\text{INFO, no car available}}$.

Looking next at the latent variable model in Table 3, we observe a significant negative mean for $\alpha_n$. Of the four socio-demographic variables, only $\gamma_{\text{graduate}}$ is significant above the 95% level, showing that respondents with a university degree have a higher value for the latent variable. The same applies to female respondents and respondents in the 35 – 50 age group, with effects significant only at the 82% and 88% level respectively. In the MNL&LV model, there is no significant difference in $\alpha_n$ for train users and car users. Overall, these findings are in line with the recognition by Ben-Akiva et al. (1999) that it can be difficult to find good causal variables for the latent variables.

Table 4 however shows significant positive estimates for all five interaction parameters in the measurement equations model, i.e. $\zeta_{I_1}$, $\zeta_{I_2}$, $\zeta_{I_3}$, $\zeta_{I_4}$, and $\zeta_{RT}$. These estimates indicate that an increase in the latent variable also has a positive impact on the response to the four indicator questions, as well as on survey completion time. This would suggest that respondents who agree with the various indicator statements, thus showing engagement with and understanding of the survey are also those respondents who have higher scale, as is the case for respondents who take longer to complete the survey, which is arguably a reflection of more careful study of each choice scenario. This justifies the interpretation of the variable as a latent engagement variable.

Directly comparing the four first indicator effects, we observe an especially strong impact of the latent variable on the statement that decisions in the choice experiment were close to real world choices, although a similarly strong effect is also observed for realism and understanding. Whether a respondent states that all attributes are considered (i.e. $I_4$) has a weaker link to the latent variable. The estimate for the impact on response time, i.e. $\zeta_{RT}$, cannot be compared to $\zeta_{I_1}$ to $\zeta_{I_4}$, given the differences in meaning and scale of measurement.

Turning our attention next to the simple MMNL model, we still observe significant estimates for all five main effects. Similarly, a majority of the Cholesky terms are significant at usual confidence levels. Table 5 shows the resulting standard deviations, coefficients of variation, and inter-coefficient correlations, with $t$-ratios calculated using the Delta method. Here, it is important to note that these results do not incorporate the additional socio-demographic interactions but relate solely to the random components of taste.
heterogeneity. We see that the standard deviations for all five coefficients are statistically significant, and that higher levels of heterogeneity arise for $\beta_{\text{RELIABILITY}}$ and $\beta_{\text{INFO}}$. In terms of correlations, we observe significant positive correlations between $\beta_{\text{TT}}$ and $\beta_{\text{INFO}}$, and between $\beta_{\text{CROWDING}}$ and $\beta_{\text{RELIABILITY}}$, with both results being consistent with intuition. On the other hand, we surprisingly observe negative correlation between the sensitivity to the rate of delays ($\beta_{\text{RELIABILITY}}$) and the expected extent of delays ($\beta_{\text{EXP. DELAY}}$).

For the MMNL model incorporating the latent variable $\alpha_n$, Table 5 shows the heterogeneity patterns after factoring out scale heterogeneity, i.e. looking only at the $\beta$ component in $\mu_n \beta$. Here, we observe drops in heterogeneity for all five random coefficients, with the biggest change being for the travel time and expected delay coefficients. These decreases are consistent with the notion that a portion of the heterogeneity in the simple MMNL model was in fact caused by scale heterogeneity, where this is now captured by the interaction between $\tau$ and the latent variable $\alpha_n$, in $\mu_n = e^{\tau \alpha_n}$. We also note changes in the correlation structure, but these increases are not all negative as would maybe be the expectation (unaccounted scale heterogeneity in the simple MMNL model would lead to more positive correlations). We now observe negative correlation between the travel time coefficient and the coefficient on crowding, suggesting that time sensitive respondents are less concerned about comfort. We more surprisingly observe negative correlation between the travel time and expected delay coefficients, while the increase in the positive correlation between the travel time sensitivity and the desire for a delay information service is consistent with intuition. We also observe increases in the positive correlation between the sensitivity to crowding and the rate of delays, both being attributes that are linked to convenience. Finally, the surprising negative correlation between $\beta_{\text{RELIABILITY}}$ and $\beta_{\text{EXP. DELAY}}$ from the simple MMNL model is now reduced by half, with an accompanying drop in significance.

The estimate for $\tau$ in the MMNL&LV model is essentially the same as in the MNL&LV model, again showing that increases in the latent variable lead to increases in scale. However, looking at Table 3, the MMNL&LV model is relatively more successful in explaining part of the latent variable model through socio-demographic interactions. Here, we note clear evidence that female respondents have a higher value for the latent engagement variable, and thus also higher scale. The estimate for $\gamma_{\text{age 35-50}}$ drops in value and significance compared to MNL&LV, but there is now some evidence that train users show higher engagement ($\gamma_{\text{train}}$ being positive and significant at the 90% level). In terms of the measurement equations estimates (cf. Table
we see a small increase in the intercept terms for the first four indicators ($\zeta_1$, $\zeta_2$, $\zeta_3$, $\zeta_4$), with no other notable changes.

4.2.3 Sample level WTP distributions

In this subsection, we incorporate the socio-demographic interactions in a study of the sample level WTP distributions. For the MNL and MNL&LV models, this consists of calculating point values for each of the six taste coefficients, taking into account the socio-demographic interactions. The non-linear treatment of the fare sensitivity also means that the fare coefficient for each respondent needs to be divided by the fare for the chosen alternative in each observation, giving different values across the 10 choices for each respondent, and hence 3,680 point values. From these point values, we then calculate respondent specific WTP measures. For the MMNL and MMNL&LV models, we additionally take into account the random variation in five of the six coefficients (noting again that no random heterogeneity was included for $\beta_{L-FARE}$). We thus obtain multiple draws from the coefficient (and hence WTP) distributions for each respondent, and the overall distributions are computed across all draws. The random scale component introduced by the $e^{ran}$ multiplier in the MNL&LV and MMNL&LV models has no impact on the WTP patterns given that all coefficients are affected in the same way. As a result, any differences between MNL and MNL&LV, and between MMNL and MMNL&LV, 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 6, showing the mean WTP across respondents, along with the standard deviation and resulting 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 are available in the design. The ratio between the WTP to reduce expected delay and reduce travel time is of the order of around 2 in the MNL and MNL&LV models, and higher in the MMNL and MMNL&LV models, consistent with previous stated preference results (cf. Hollander, 2005).

As expected, the degree of heterogeneity is significantly larger in the MMNL and MMNL&LV models than in the MNL and MNL&LV models, where in the latter, any heterogeneity is accommodated solely through socio-demographic interactions. The increase in the heterogeneity through the
5 Conclusions

incorporation of random variations is lowest for the value of travel time, and highest for the sensitivity to the rate of delays and the desire for the provision of a delay information service. Additionally, we observe modest increases in the mean values for the valuation of travel time, the WTP to reduce crowding, and the WTP for a delay information service. There is a small reduction in the WTP for reductions in the rate of delays, but a major increase in the WTP for reductions in expected delays. As previously indicated in the discussion of the main estimation results, the WTP results for the two MNL models are very similar. What is more interesting are the differences between the MMNL and MMNL&LV models. Here, we observe a small reduction in the mean WTP values (except for average delay). More importantly however, the reduction in the standard deviations is larger than that in the means for the WTP for travel time, rate of delays and expected delays, leading to reductions in the degree of WTP heterogeneity. The heterogeneity in the WTP for reduced rates of crowding stays stable, while there is a modest increase in the heterogeneity in the WTP for the delay information service - this is a result of the reduced estimate for \( \Delta_{\text{INFO,no car available}} \), and the resulting lower sample level mean value for the desire for a delay information service (where the decrease in the standard deviation is smaller). Overall, and in conjunction with the correlation patterns in Table 5, these results show that the inclusion of a random scale effect linked to survey engagement can lead to differences not just in the heterogeneity patterns for individual coefficients but also the WTP distributions, even though such a multiplicative scale factors affects all coefficients in the same way.

5 Conclusions

There is growing evidence indicating that the evaluation of goods and services is influenced by factors beyond the attributes of the offered alternatives. A series of overlooked features in choice modelling include acknowledging how the choice is approached and perceived by the respondent and the level of engagement in replying to survey questions. An assumption is typically made that all respondents are equally certain about their own preferences. A case can clearly be made for variation across respondents in the quality of their response as a result of decision uncertainty, possibly linked to complexity, or as a result of reduced engagement with the survey. Such respondent-based sources of analytical uncertainty are generally not considered and controlled for in random utility choice modelling.

Another hotly debated issue concerns the characterisation of respondent heterogeneity. Discrete choice models by their nature confound different
sources of heterogeneity. This may lead an analyst to attribute heterogeneity to conventional taste variation when it is in reality due to different choice processing or heterogeneity in scale.

The present paper looks particularly at how variations across respondents in the degree to which they engage with a survey will lead to scale heterogeneity. Here, we propose a joint latent variable random scale heterogeneity approach to explore the impact of survey engagement on response variability. Past studies looking at the relationship between respondent engagement and model scale have generally relied on direct approaches, e.g. using respondent stated measures of difficulty and linking these to scale heterogeneity. Aside from the subjective nature of such measures, there is a clear risk of bias caused by endogeneity given the likely correlation between such responses and other unobserved factors.

To overcome some of the problems associated with the use of direct modelling approaches this paper proposes the use of a latent variable model in which a latent engagement variable helps drive the response both to attitudinal questions relating to survey engagement and the stated choice component of the survey. While in the majority of latent variable work to date, the latent variables have been used to accommodate heterogeneity in relative sensitivities, in this paper, we make the link between the latent variable and scale heterogeneity. The proposed model overcomes the biases inherent in incorporating attitudes or engagement indicators directly in a utility modelling framework.

The results from our model indicate that increases in the latent engagement variable lead to stronger agreement with statements relating to whether scenarios were realistic, whether they were fully understood, and whether choices could be made as in real life. A higher value for the latent variable also increases the likelihood of respondents saying that they generally take into account all service characteristics when evaluating a public transport alternative. Finally, increases in the latent engagement variable are also correlated with longer survey response times. At the choice modelling end, increases in the latent variable have a significant and positive impact on model scale, thus supporting our interpretation of the latent variable as a measure of respondent engagement. Additionally, we found a link between various socio-demographic factors and the latent engagement variable, with more positive values for female respondents and respondents with a university degree. When working with models that allow for additional conventional heterogeneity, we find that the incorporation of the latent variable leads to substantially different patterns of heterogeneity, with reductions in heterogeneity for all random coefficients, and resulting reductions in heterogeneity
in the WTP distributions, as well as more plausible correlation patterns.

The model presented in this paper provides us with a tool allowing for variations in scale linked to differences in survey engagement, making use of additional data collected from respondents to act as proxies for engagement, but crucially without using such data in a direct, potentially bias-causing way. Future research needs to test the model framework in a wider variety of choice settings and data structures, and another area for additional work concerns the impact of using different mixture distributions for the version of the model that incorporates heterogeneity in the underlying utility coefficients.

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 $\alpha_n$. 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.

As a second step, exploratory factor analysis was carried out to assess which of the three involvement and ten attitude statements collected in the survey were possibly linked to the underlying $\alpha_n$ factor. The prior assumption of high correspondence between the involvement indicators and a strong link to the $\alpha_n$ was confirmed by factor analysis where models hypothesising 2-6 factors for the 13 indicators were compared based on the $\chi^2$ statistic of overall fit. Factor analysis was carried out in R using varimax rotation of the factors. The factor loadings for $I_1$-$I_3$ 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 (I₄). Again, Cronbach’s alpha indicated a good degree of association among the four indicators (α = 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 (χ² = 1.71 with a p value of 0.425 and 2 df). All indicators had significant loadings on the latent variable ranging from 0.17 (with a t-ratio of 4.14) for I₄ up to a loading of 1 (t-ratio of 24) for I₃. The four indicators accounted for 60% of the variance in αn.

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 individual survey-time (measured in minutes). The final model specification also included several socio-demographic variables to control their influence on the latent engagement variable.

References


Fig. 1: Structure of latent engagement model

Fig. 2: Respondent replies to four indicator statements
Tab. 1: Performance of different models

<table>
<thead>
<tr>
<th>Model</th>
<th>Overall LL</th>
<th>Par.</th>
<th>Choice Component LL</th>
<th>Par.</th>
<th>Measurement Model LL</th>
<th>Par.</th>
</tr>
</thead>
<tbody>
<tr>
<td>MNL</td>
<td>-3,343.20</td>
<td>13</td>
<td>-3,343.20</td>
<td>13</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>MNL&amp;LV</td>
<td>-5,299.52</td>
<td>29</td>
<td>-3,330.08</td>
<td>14</td>
<td>-1,899.44</td>
<td>15</td>
</tr>
<tr>
<td>MMNL</td>
<td>-5,181.32</td>
<td>28</td>
<td>-5,181.52</td>
<td>28</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>MMNL&amp;LV</td>
<td>-5,072.64</td>
<td>44</td>
<td>-3,168.88</td>
<td>29</td>
<td>-1,903.76</td>
<td>15</td>
</tr>
</tbody>
</table>

Tab. 2: Estimation results: choice model component

<table>
<thead>
<tr>
<th>Model</th>
<th>MNL</th>
<th>MNL&amp;LV</th>
<th>MMNL</th>
<th>MMNL&amp;LV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>est.</td>
<td>t-rat.</td>
<td>est.</td>
<td>t-rat.</td>
</tr>
<tr>
<td>( \delta_1 )</td>
<td>0.4217</td>
<td>8.98</td>
<td>0.4122</td>
<td>8.92</td>
</tr>
<tr>
<td>( \delta_2 )</td>
<td>0.1815</td>
<td>3.72</td>
<td>0.1801</td>
<td>3.75</td>
</tr>
<tr>
<td>( \mu_{TT} )</td>
<td>-2.1307</td>
<td>-3.07</td>
<td>-2.1366</td>
<td>-3.52</td>
</tr>
<tr>
<td>( \mu_{FARE} )</td>
<td>-5.1682</td>
<td>-19.21</td>
<td>-5.0882</td>
<td>-19.03</td>
</tr>
<tr>
<td>( \mu_{CROWDING} )</td>
<td>-2.1533</td>
<td>-9.93</td>
<td>-2.1366</td>
<td>-9.90</td>
</tr>
<tr>
<td>( \mu_{RELiability} )</td>
<td>-1.6980</td>
<td>-6.78</td>
<td>-1.6946</td>
<td>-6.86</td>
</tr>
<tr>
<td>( \mu_{EXP _DELAY} )</td>
<td>-5.9846</td>
<td>-6.44</td>
<td>-5.9921</td>
<td>-6.60</td>
</tr>
<tr>
<td>( \mu_{INFO} )</td>
<td>0.2907</td>
<td>4.30</td>
<td>0.2740</td>
<td>4.52</td>
</tr>
<tr>
<td>( \Delta_{TT, female} )</td>
<td>-1.3384</td>
<td>-6.67</td>
<td>-1.3350</td>
<td>-6.28</td>
</tr>
<tr>
<td>( \Delta_{CROWDING, age &gt; 50} )</td>
<td>-1.2557</td>
<td>-1.68</td>
<td>-1.0839</td>
<td>-1.49</td>
</tr>
<tr>
<td>( \Delta_{TT, car available} )</td>
<td>-1.3318</td>
<td>-3.19</td>
<td>-1.1236</td>
<td>-2.71</td>
</tr>
<tr>
<td>( \Delta_{FARE, no car available} )</td>
<td>-1.6782</td>
<td>-4.56</td>
<td>-1.8572</td>
<td>-5.07</td>
</tr>
<tr>
<td>( \Delta_{INFO, no car available} )</td>
<td>0.1687</td>
<td>1.87</td>
<td>0.1891</td>
<td>2.14</td>
</tr>
<tr>
<td>( s_{1,1} )</td>
<td>-</td>
<td>-</td>
<td>3.3278</td>
<td>6.74</td>
</tr>
<tr>
<td>( s_{2,1} )</td>
<td>-</td>
<td>-</td>
<td>0.0358</td>
<td>0.65</td>
</tr>
<tr>
<td>( s_{2,2} )</td>
<td>-</td>
<td>-</td>
<td>4.3601</td>
<td>9.66</td>
</tr>
<tr>
<td>( s_{3,1} )</td>
<td>-</td>
<td>-</td>
<td>1.2946</td>
<td>1.55</td>
</tr>
<tr>
<td>( s_{3,2} )</td>
<td>-</td>
<td>-</td>
<td>2.1671</td>
<td>3.15</td>
</tr>
<tr>
<td>( s_{3,3} )</td>
<td>-</td>
<td>-</td>
<td>-4.1115</td>
<td>-5.95</td>
</tr>
<tr>
<td>( s_{4,1} )</td>
<td>-</td>
<td>-</td>
<td>-0.4608</td>
<td>-0.11</td>
</tr>
<tr>
<td>( s_{4,2} )</td>
<td>-</td>
<td>-</td>
<td>1.9652</td>
<td>0.75</td>
</tr>
<tr>
<td>( s_{4,3} )</td>
<td>-</td>
<td>-</td>
<td>7.6407</td>
<td>2.36</td>
</tr>
<tr>
<td>( s_{4,4} )</td>
<td>-</td>
<td>-</td>
<td>10.5600</td>
<td>4.79</td>
</tr>
<tr>
<td>( s_{5,1} )</td>
<td>-</td>
<td>-</td>
<td>0.3448</td>
<td>2.15</td>
</tr>
<tr>
<td>( s_{5,2} )</td>
<td>-</td>
<td>-</td>
<td>0.0186</td>
<td>0.15</td>
</tr>
<tr>
<td>( s_{5,3} )</td>
<td>-</td>
<td>-</td>
<td>0.3332</td>
<td>1.89</td>
</tr>
<tr>
<td>( s_{5,4} )</td>
<td>-</td>
<td>-</td>
<td>-0.5799</td>
<td>-2.73</td>
</tr>
<tr>
<td>( s_{5,5} )</td>
<td>-</td>
<td>-</td>
<td>0.6719</td>
<td>3.74</td>
</tr>
<tr>
<td>( \tau )</td>
<td>-</td>
<td>-</td>
<td>0.2000</td>
<td>4.84</td>
</tr>
</tbody>
</table>
Tab. 3: Estimation results: latent variable model

<table>
<thead>
<tr>
<th></th>
<th>MNL&amp;LV</th>
<th>MMNL&amp;LV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>est.</td>
<td>t-rat.</td>
</tr>
<tr>
<td>( \mu_{\alpha} )</td>
<td>-0.2459</td>
<td>-2.10</td>
</tr>
<tr>
<td>( \gamma_{\text{female}} )</td>
<td>0.1443</td>
<td>1.33</td>
</tr>
<tr>
<td>( \gamma_{\text{age 35-50}} )</td>
<td>0.1767</td>
<td>1.57</td>
</tr>
<tr>
<td>( \gamma_{\text{graduate}} )</td>
<td>0.2661</td>
<td>2.54</td>
</tr>
<tr>
<td>( \gamma_{\text{train}} )</td>
<td>-0.0149</td>
<td>-0.14</td>
</tr>
</tbody>
</table>

Tab. 4: Estimation results: measurement equations

<table>
<thead>
<tr>
<th></th>
<th>MNL&amp;LV</th>
<th>MMNL&amp;LV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>est.</td>
<td>t-rat.</td>
</tr>
<tr>
<td>( \zeta_1 )</td>
<td>0.8382</td>
<td>20.55</td>
</tr>
<tr>
<td>( \zeta_2 )</td>
<td>0.9088</td>
<td>21.63</td>
</tr>
<tr>
<td>( \zeta_3 )</td>
<td>0.9865</td>
<td>30.30</td>
</tr>
<tr>
<td>( \zeta_4 )</td>
<td>0.1695</td>
<td>4.13</td>
</tr>
<tr>
<td>( \zeta_{RT} )</td>
<td>0.0547</td>
<td>1.99</td>
</tr>
<tr>
<td>( \sigma_{I_1} )</td>
<td>0.5701</td>
<td>22.40</td>
</tr>
<tr>
<td>( \sigma_{I_2} )</td>
<td>0.5476</td>
<td>20.22</td>
</tr>
<tr>
<td>( \sigma_{I_3} )</td>
<td>0.2599</td>
<td>5.73</td>
</tr>
<tr>
<td>( \sigma_{I_4} )</td>
<td>0.7688</td>
<td>27.06</td>
</tr>
<tr>
<td>( \sigma_{RT} )</td>
<td>0.5176</td>
<td>27.11</td>
</tr>
</tbody>
</table>
Tab. 5: Analysis of random coefficients in MMNL and MMNL&LV models, after factoring out scale heterogeneity

<table>
<thead>
<tr>
<th></th>
<th>MMNL</th>
<th>MMNL&amp;LV</th>
<th>change</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>est.</td>
<td>t-rat.</td>
<td>est.</td>
</tr>
<tr>
<td>$\sigma_{TT}$</td>
<td>3.328</td>
<td>6.74</td>
<td>2.6859</td>
</tr>
<tr>
<td>$\sigma_{CROWDING}$</td>
<td>4.3601</td>
<td>9.64</td>
<td>4.2077</td>
</tr>
<tr>
<td>$\sigma_{RELIABILITY}$</td>
<td>4.8234</td>
<td>8.44</td>
<td>4.2601</td>
</tr>
<tr>
<td>$\sigma_{EXP. DELAY}$</td>
<td>13.1897</td>
<td>4.43</td>
<td>11.4954</td>
</tr>
<tr>
<td>$\sigma_{INFO}$</td>
<td>1.0158</td>
<td>11.01</td>
<td>0.9824</td>
</tr>
<tr>
<td>$cv_{TT}$</td>
<td>1.6138</td>
<td>3.08</td>
<td>1.4631</td>
</tr>
<tr>
<td>$cv_{CROWDING}$</td>
<td>1.4868</td>
<td>6.81</td>
<td>1.4573</td>
</tr>
<tr>
<td>$cv_{RELIABILITY}$</td>
<td>2.3585</td>
<td>4.39</td>
<td>2.2224</td>
</tr>
<tr>
<td>$cv_{EXP. DELAY}$</td>
<td>1.2035</td>
<td>4.34</td>
<td>1.0359</td>
</tr>
<tr>
<td>$cv_{INFO}$</td>
<td>2.5635</td>
<td>3.51</td>
<td>2.3017</td>
</tr>
<tr>
<td>corr.($\beta_{TT}, \beta_{CROWDING}$)</td>
<td>0.0082</td>
<td>0.05</td>
<td>-0.2876</td>
</tr>
<tr>
<td>corr.($\beta_{TT}, \beta_{RELIABILITY}$)</td>
<td>0.2674</td>
<td>1.55</td>
<td>0.2934</td>
</tr>
<tr>
<td>corr.($\beta_{TT}, \beta_{EXP. DELAY}$)</td>
<td>-0.0330</td>
<td>-0.11</td>
<td>-0.2862</td>
</tr>
<tr>
<td>corr.($\beta_{TT}, \beta_{INFO}$)</td>
<td>0.3396</td>
<td>2.19</td>
<td>0.4703</td>
</tr>
<tr>
<td>corr.($\beta_{CROWDING}, \beta_{RELIABILITY}$)</td>
<td>0.4515</td>
<td>3.31</td>
<td>0.6796</td>
</tr>
<tr>
<td>corr.($\beta_{CROWDING}, \beta_{EXP. DELAY}$)</td>
<td>0.1487</td>
<td>0.74</td>
<td>-0.1298</td>
</tr>
<tr>
<td>corr.($\beta_{CROWDING}, \beta_{INFO}$)</td>
<td>0.0211</td>
<td>0.18</td>
<td>-0.0103</td>
</tr>
<tr>
<td>corr.($\beta_{RELIABILITY}, \beta_{EXP. DELAY}$)</td>
<td>-0.4362</td>
<td>-2.58</td>
<td>-0.2282</td>
</tr>
<tr>
<td>corr.($\beta_{RELIABILITY}, \beta_{INFO}$)</td>
<td>-0.1973</td>
<td>-1.48</td>
<td>-0.1923</td>
</tr>
<tr>
<td>corr.($\beta_{EXP. DELAY}, \beta_{INFO}$)</td>
<td>-0.2648</td>
<td>-1.32</td>
<td>-0.2475</td>
</tr>
</tbody>
</table>
Tab. 6: Sample level WTP distributions

<table>
<thead>
<tr>
<th></th>
<th>MNL</th>
<th>MNL&amp;LV vs MNL</th>
<th>MMNL</th>
<th>MMNL vs MNL</th>
<th>MMNL&amp;LV vs MMNL &amp; LV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>cv</td>
<td>mean</td>
<td>s.d.</td>
</tr>
<tr>
<td>time (£/hr)</td>
<td>1.3447</td>
<td>1.9293</td>
<td>1.43</td>
<td>0.43%</td>
<td>0.27%</td>
</tr>
<tr>
<td>crowding (£/10%)</td>
<td>0.1180</td>
<td>0.1804</td>
<td>1.52</td>
<td>-2.41%</td>
<td>-2.50%</td>
</tr>
<tr>
<td>delays (£/10%)</td>
<td>0.0792</td>
<td>0.1148</td>
<td>1.45</td>
<td>0.15%</td>
<td>0.84%</td>
</tr>
<tr>
<td>delay duration (£/hr)</td>
<td>2.7918</td>
<td>4.0457</td>
<td>1.45</td>
<td>0.48%</td>
<td>1.16%</td>
</tr>
<tr>
<td>info system (£)</td>
<td>0.1579</td>
<td>0.2114</td>
<td>1.34</td>
<td>3.69%</td>
<td>3.61%</td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>cv</td>
<td>mean</td>
<td>s.d.</td>
</tr>
<tr>
<td>time (£/hr)</td>
<td>1.4671</td>
<td>2.9678</td>
<td>2.02</td>
<td>9.10%</td>
<td>53.83%</td>
</tr>
<tr>
<td>crowding (£/10%)</td>
<td>0.1252</td>
<td>0.3358</td>
<td>2.68</td>
<td>5.24%</td>
<td>86.14%</td>
</tr>
<tr>
<td>delays (£/10%)</td>
<td>0.0753</td>
<td>0.3297</td>
<td>4.38</td>
<td>-4.92%</td>
<td>187.27%</td>
</tr>
<tr>
<td>delay duration (£/hr)</td>
<td>4.0789</td>
<td>10.2892</td>
<td>2.52</td>
<td>46.10%</td>
<td>154.33%</td>
</tr>
<tr>
<td>info system (£)</td>
<td>0.1683</td>
<td>0.6925</td>
<td>4.11</td>
<td>6.64%</td>
<td>227.62%</td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>cv</td>
<td>mean</td>
<td>s.d.</td>
</tr>
<tr>
<td>time (£/hr)</td>
<td>1.4018</td>
<td>2.6305</td>
<td>1.88</td>
<td>3.79%</td>
<td>36.29%</td>
</tr>
<tr>
<td>crowding (£/10%)</td>
<td>0.1233</td>
<td>0.3305</td>
<td>2.68</td>
<td>6.27%</td>
<td>87.89%</td>
</tr>
<tr>
<td>delays (£/10%)</td>
<td>0.0718</td>
<td>0.2977</td>
<td>4.15</td>
<td>-9.45%</td>
<td>157.23%</td>
</tr>
<tr>
<td>delay duration (£/hr)</td>
<td>4.1143</td>
<td>9.5498</td>
<td>2.32</td>
<td>46.67%</td>
<td>133.33%</td>
</tr>
<tr>
<td>info system (£)</td>
<td>0.1612</td>
<td>0.6799</td>
<td>4.22</td>
<td>-1.52%</td>
<td>210.48%</td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>cv</td>
<td>mean</td>
<td>s.d.</td>
</tr>
<tr>
<td>time (£/hr)</td>
<td>1.4487</td>
<td>1.9293</td>
<td>1.43</td>
<td>-0.43%</td>
<td>0.27%</td>
</tr>
<tr>
<td>crowding (£/10%)</td>
<td>0.1180</td>
<td>0.1804</td>
<td>1.52</td>
<td>-2.41%</td>
<td>-2.50%</td>
</tr>
<tr>
<td>delays (£/10%)</td>
<td>0.0792</td>
<td>0.1148</td>
<td>1.45</td>
<td>0.15%</td>
<td>0.84%</td>
</tr>
<tr>
<td>delay duration (£/hr)</td>
<td>2.7918</td>
<td>4.0457</td>
<td>1.45</td>
<td>0.48%</td>
<td>1.16%</td>
</tr>
<tr>
<td>info system (£)</td>
<td>0.1579</td>
<td>0.2114</td>
<td>1.34</td>
<td>3.69%</td>
<td>3.61%</td>
</tr>
</tbody>
</table>