Methodological challenges in modelling the choice of mode for a new travel alternative using binary stated choice data - the case of high speed rail in Norway

Stefan Flügel[a], Askill H. Halse[b], Juan de Dios Ortúzar[c] and Luis I. Rizzi[c]

[a] *Department of Economics and Resource Management,*
Norwegian University of Life Sciences (NMBU), P.O. Box 5003, NO-1432 Ås, Norway

[b] *Institute of Transport Economics (TOI), Gaustadalleen 21, NO-0349 Oslo, Norway*

[c] *Department of Transport Engineering and Logistics,*
Centre for Sustainable Urban Development (CEDEUS), Casilla 306, Cod. 105, Correo 22, Santiago, Chile

Abstract:

Binary stated choices between traveller's current travel mode and a not-yet-existing mode might be used to build a forecasting model with all (current and future) travel alternatives. One challenge with this approach is the identification of the most appropriate inter-alternative error structure of the forecasting model.

By critically assessing the practise of translating estimated group scale parameters into nest parameters, we illustrate the inherent limitations of such binary choice data. To overcome some of the problems, we use information from both stated and revealed choice data and propose a model with a cross-nested logit specification, which is estimated on the pooled data set.

1. Introduction

A large-scale study on the feasibility and social benefits of high-speed rail (HSR) in Norway was recently carried out (Jernbaneverket 2012). The estimated market potential of HSR is naturally a crucial element in this quest, as the predicted ridership has a direct effect on expected revenues, user benefits and greenhouse gas reductions. The demand forecasting model (Atkins 2012) was based on a stated choice (SC) study where respondents faced customized surveys based on their current mode choice (revealed choice, RC). The survey included binary choice experiments (CE) between the respondents' current modes and a new HSR alternative (Figure 1 shows a schematic illustration). A similar approach was used in an independent market study conducted by the Institute of Transport Economics, TØI (Flügel and Halse 2012).
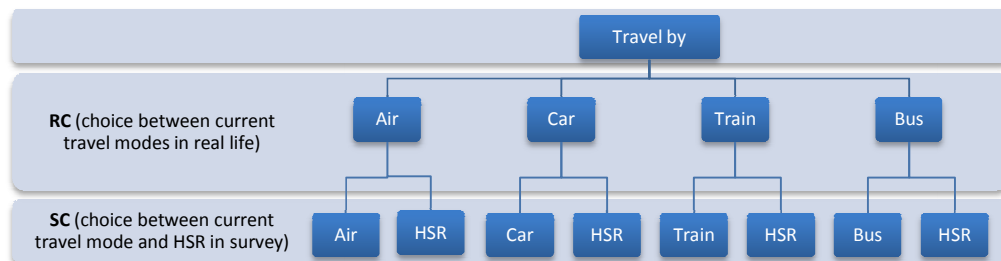

Figure 1. Decision structure in recent Norwegian HSR-studies (RC: revealed choice; SC: stated choice)

The main advantage of binary CE (instead of CE with a full choice set) is the simplification of the respondent's choice task. In a travel mode choice context, CE often entails a rather high degree of complexity because of the large number of attributes typically required to characterise each alternative. Lowering the overall number of attributes is likely to increase respondents' ability to choose between alternatives (Caussade *et al* 2005). In a pivot design, where respondents are typically instructed to recall the last trip they made, it is quite natural to discard the rejected travel alternatives letting the respondent focus on the current travel mode and the hypothetical new alternative.

However, while it is desirable to reduce the respondent's choice set from an experimental design point of view (in our case: providing personal specific choice sets consisting of respondent's current mode and HSR), one would like to build a forecasting model that allows considering the whole future choice set and which applies to all future decision makers, independent of their chosen mode at the time the CE surveys were conducted. This applies, in particular, to HSR implementation scenarios that usually involve long-term predictions. Changes in many level-of-service (LoS) variables of, potentially, all travel modes are possible not only because of the long time horizon, but also because a HSR implementation is likely to affect the competitive structure of the whole travel market. Therefore, it seems unduly restrictive to limit choice sets and to condition model parameters for choice predictions in the forecasting year (e.g. in 2024, the earliest possible year for a HSR-implementation in Norway) on current RC

choices (data from year 2010 in our case). Consequently, a model with a generic choice set and utility functions, independent of the original self-selection of travellers to travel modes is necessary.

Of course, aiming for a generic forecasting model based on binary stated choices (with only one alternative, HSR, being part of every respondent's choice set) is not optimal, as it does not allow considering directly how current car users, say, react to the LoS of other current modes (air, bus and traditional train). When specifying transport specific coefficients in the utility function, one needs to assume that, for example, the current car user's marginal utility (MU) of in-vehicle-time (IVT) by car is representative of everyone's MU for IVT by car. Challenges in finding an appropriate deterministic utility function are not, however, the focus of this paper; moreover, we will assume - unless specified differently - that we can find deterministic utility functions (up to a scale parameter) that fit all *user groups* (defined on the basis current mode choice) "equally well".

For estimation, the different binary choice datasets are typically merged and a mode choice model with a common set of coefficients for HSR is estimated. In this procedure, different scale parameters (so called *group scale* parameters), that are inversely proportional to the error variances associated with each experiment, ought to be estimated to account for the fact that they might actually differ (Louviere *et al* 2000).

While the group scale parameters facilitate the estimation of a common deterministic utility function based on user-specific binary choices, it is not obvious how these parameters may be carried over to a forecasting model with a full choice set. In particular, setting up a nested logit (NL) model by naively treating group scale parameters as structural (nest) parameters, as done by Atkins (2012) in the official assessment study for HSR in Norway, involves several pitfalls:

(i) The group scale parameters only reflect the relative utility scale in choices between the different binary choice tasks (i.e. HSR versus one of the current modes) but not the utility scale difference between existing travel modes. In most cases, this means that the scale at the upper level of the nesting structure and the correlation structure among current modes has to be assumed implicitly (see sections 3.1. and 3.3); we will discuss how RC data between current modes might be utilized here (see section 4).

(iii) The group scale parameters do not only reflect "similarity" of transport modes, (i.e. the degree to which two or more alternatives share unobserved features, which is the classical interpretation of nest parameters, see Ortúzar and Willumsen, 2011, section 7.4.2). They might also include other error sources – in particular unobserved taste heterogeneity – that are associated with characteristics of the user groups rather than of the modes. We will discuss this in more detail in section 3.2., and using an empirical example, we will also show that results change after accounting for unobserved taste heterogeneity with random coefficients models (section 3.4.).

(ii) In many instances a NL model might not be flexible enough to account for the correlation structure suggested by the various group scale parameters. We propose the cross-nested logit (CNL) model as a more flexible structure for this purpose.

As the paper is mostly concerned with error variance differences (utility scale differences) between various user groups, travel modes and datasets, it is important to stress that the error term is, as usual, conditioned by the specification of the deterministic part of utility (i.e. the selection of explanatory variables and their functional form). For instance, when talking about correlation (or "similarity") of travel modes, we are always relating to those parts of the utility function that are not accounted for by the explanatory variables. Indeed, correlation patterns in the error term are nothing desirable in itself and one would ideally strive for a multinomial logit (MNL) model by including all the variables that might explain correlation among travel alternatives. However, this is often not possible in practise (some variables are unobservable, others are just too expensive to collect). Thus, a question often asked to the researcher refers to the most appropriate correlation (nesting) structure in the forecasting model.

The main contribution of this paper is an in-depth discussion about the most appropriate (alternative specific) correlation structure in a case where the deterministic component of the utility functions and the scale parameters are estimated from (user group specific) binary SC data. While obviously the (relative) size of the scale parameters will depend on the chosen systematic utility function, most of the discussion in this paper can be made at a general level without being specific about a particular systematic utility function.

2. Alternative Model Forms

2.1. Multinomial logit (MNL) model

We describe a standard discrete choice set up where traveller n chooses between different transport modes i belonging to a (personal specific) choice set C_n, according to the following choice rule:

(1) $\quad\quad U_{in} = \varepsilon_{in} \varepsilon \ \ \varepsilon_{in} > U_{jn} = \varepsilon_{jn} \varepsilon \varepsilon_{jn} \quad\quad \varepsilon or\ all\ j \in C_n$.

where the deterministic component of utility V_{in} is a function of attributes $\underline{X}$ and a set of parameters $\underline{\beta}$ to be estimated; in a MNL model, the random terms ε_{in} are assumed to distribute IID-Gumbel with mean zero and variance given by: $\sigma^2 = \frac{\pi^2}{\lambda^2 6}$ where $\lambda > 0$ is the scale parameter of the distribution. With this, the MNL model choice probabilities are given by (McFadden 1974, Ortúzar and Willumsen, 2011, Chapter 7):

(2) $\quad\quad P_{in} = Pr\big(V_{in} + \varepsilon_{in} > V_{jn} + \varepsilon_{jn}\big) = \dfrac{e^{\lambda V_{in}}}{\sum_{j \in C_n} e^{\lambda V_{jn}}}$.

The IID-assumption in equation (2) implies proportional substitution patterns across alternatives as the utility of any two alternatives is uncorrelated. Note that the scale factor λ cannot be estimated separately from the parameters β in the deterministic component of utility, so it has to be normalized (see the discussion on identifiability by Walker 2001).

2.2. Heteroskedastic logit (HL) model

The analyst might wish to allow for different error variances for different subgroups in the data. For this, the scale factor (which is inversely proportional to the error variance) can be assumed to be non-generic allowing for different *group scale* parameters, v_{g_n}; in this case the choice probabilities of this HL[1] model become:

$$(3) \qquad P_{in} = \frac{e^{v_{g_n} V_{in}}}{\sum_{j \in C_{g_n}} e^{v_{g_n} V_{jn}}} \; for \; all \; i \in C_{g_n}$$

Note that not all group scale parameters can be estimated simultaneously. For identification, one of them has to be fixed (typically at the value 1). The group scale parameters affect the resulting choice probabilities as illustrated in Figure 2.
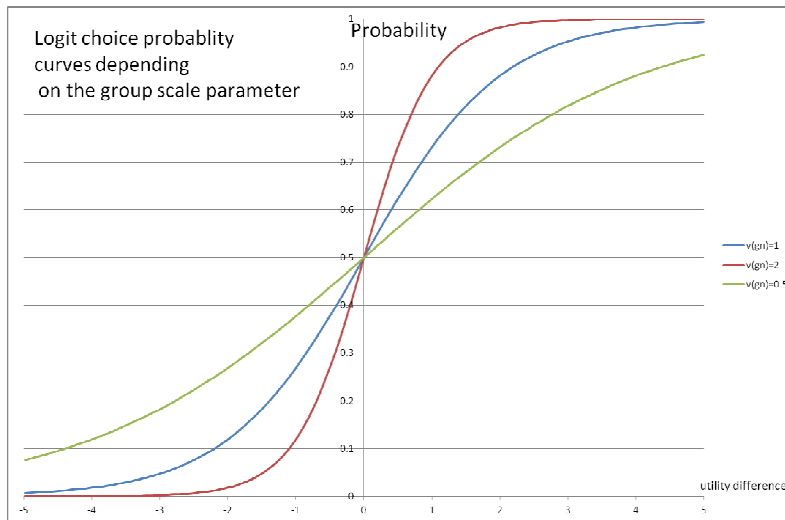

Figure 2: Illustration of effect of group scale parameter on choice probability

---

[1] Note that this model is not identical to the "heteroscedastic extreme value model" (Bhat 1995), which is called "Heteroskedastic Logit" by Train (2009, page 92), as the error variance of each alternative varies. In the HL model described here, the error variance varies for every subgroup but is the same for all alternatives in a subgroup.

If the group scale parameters are different in value, the IID-assumption only applies within the subgroup but is relaxed for a joint sample that combines choice data from different user groups.

2.3. Nested logit (NL) model

In a NL model, alternatives are allocated into non-overlapping nests, m, that contain alternatives $i = 1, ..., J_m$. NL models (Williams 1977; Daly and Zachary 1978) can also be derived from the family of GEV-models (McFadden 1978). Using GEV-notation, the choice probability for the NL model is given as:

$$(4) \quad P_{in} = \frac{(\Sigma_{j=1}^{J_m} e^{\mu_m V_{jn}})^{\frac{\mu}{\mu_m}}}{\Sigma_{m=1}^{M} (\Sigma_{j=1}^{J_m} e^{\mu_m V_{jn}})^{\frac{\mu}{\mu_m}}} \frac{e^{\mu_m V_{in}}}{(\Sigma_{j=1}^{J_m} e^{\mu_m V_{jn}})}$$

where μ_m are scale parameters applied to the alternatives in nest m. We refer to them as *nest* or *structural* parameters. Similar to the group scale parameters v_{g_n} (but arising from a different perspective), the nest parameters are inversely related to the corresponding error variance. A restriction of NL model is that nests cannot overlap, that is, each alternative can only enter one nest.

The overall scale of utility, μ, here interpreted as the scale for the choice between nests, is an arbitrary positive number and only the ratio $\frac{\mu}{\mu_m}$ has a behavioural interpretation. It can be shown (e.g. Bhat 1997) that the correlation between the utilities of two alternatives i and j is given by:

$$(5) \qquad Corr(U_i, U_j) = \left(1 - (\frac{\mu}{\mu_m})^2 \right) d_{ij}$$

where d_{ij} is one when i and j belong to nest m and zero otherwise.

A low error variance in nest m (i.e. μ_m relatively larger than μ) implies a large correlation among utilities between the nested alternatives. For GEV-conditions to hold, we need $\mu_m \geq \mu > 0$. This implies that the utility of the nested alternatives must be positively correlated. This has to be taken into account when setting up a nested structure. The choice probability in (4) has a nice two-fold interpretation as the product of the probability of choosing between nests (choice at the 'upper level') and the probability of choosing between alternatives in the chosen nest (choice at the 'lower level').

2.4. Cross-nested logit (CNL) models

A GEV-model that allows alternatives to enter several nests is the CNL (Williams 1977; Vovsha 1997; Bierlaire 2006) Choice probabilities in the CNL model are given by (Abbe *et al.* 2007, page 797)[2]:

$$
(6) \qquad P_{in} = \sum_{m=1}^{M} \frac{\left(\sum_{j=1}^{J_m} \alpha_{jm}^{\frac{\mu_m}{\mu}} e^{\mu_m V_{jn}} \right)^{\frac{\mu}{\mu_m}}}{\sum_{m=1}^{M} \left(\sum_{j=1}^{J_m} \alpha_{jm}^{\frac{\mu_m}{\mu}} e^{\mu_m V_{jn}} \right)^{\frac{\mu}{\mu_m}}} \frac{\alpha_{jm}^{\frac{\mu_m}{\mu}} e^{\mu_m V_{in}}}{\left(\sum_{j=1}^{J_m} \alpha_{jm}^{\frac{\mu_m}{\mu}} e^{\mu_m V_{jn}} \right)}
$$

Bierlaire (2006, page 293) also derives the following conditions to be met by a CNL model:

1. $\mu_m \geq \mu > 0$ for all $m = 1, \ldots, M$

2. $\alpha_{im} \geq 0$ for all $j = 1, \ldots, J$, $m = 1, \ldots, M$

3. $\sum_{m=1}^{M} \alpha_{jm} > 0$ for all $j = 1, \ldots, J$.

Following Train (2009), we refer to the α-parameters as *allocation parameters*[3]. The NL-model is a special case of CNL-model where all α_{jm} are zero except for the nest *m* the alternative is included in. The exact correlation structure of CNL models (Abbe *et al.* 2007[4]) is much more involved than that for the NL (5). The following is an approximation proposed by Papola (2004):

$$
(7) \qquad Corr\left(U_i, U_j\right) \approx \sum_{m=1}^{M} \alpha_{im}^{\frac{1}{2}} \alpha_{jm}^{\frac{1}{2}} \left(1 - \left(\frac{\mu}{\mu_m}\right)^2 \right)
$$

Equation (7) underlines the fact that the allocation parameters affect the correlation structure of the model. Equations (6) and (7) can be thought of as

---

[2] Equation (6) is the resulting choice probability for the most general formulation of the CNL, but simpler formulations are available (Ben-Akiva and Bierlaire 1999; Wen and Koppelman 2001). BIOGEME 1.8 (and later versions) use the CNL model version in equation (6).

[3] For interpretation and parameter identification, the condition $\sum_{m=1}^{M} \alpha_{jm} = 1$ should be imposed. Then, the allocation parameters are readily interpreted as the portion of an alternative that enters each nest. However, the relationship between α_{im} and μ_m is not obvious. Intuitively, a high correlation between nested alternatives (high μ_m) should go along with a relative high portion of a particular alternative being associated with that nest. However, we are not aware of suggestions for possible functional relationship between α_{im} and μ_m.

[4] See equation (20) in Abbe *et al* (2007, page 800) for the exact formula of correlation between two alternatives in a CNL.

'weighted averages' of (4) and (5) respectively, where averages are taken over nests and the allocation parameters represent weights.

3. Deriving a NL Model from a HC Model on Binary SC Data

In this section we will critically assess the procedure of translating group scale parameters obtained from an estimated HL model into nest parameters of a NL forecasting model; this approach was applied by Atkins (2012a) in the Norwegian HSR assessment study, so the discussion has practical relevance. In this context the group scale parameters stem from the different subsets of travellers (using various travel modes in practice) being subject to different SC experiments (asking them to choose between their current mode and HSR, see Figure 1).

We assume that no RC data is available to replicate the situation in Atkins (2012) that only used SC data in their estimation model. In section 4 we will discuss the use of RC data as a supplement.

3.1. Mathematical conditions

As mentioned above, group scale parameters (v_{g_n}) and nest parameters (μ_m) are both inversely proportional to their related error variances. The difference lies in which error variance is considered. Group scale parameters relate to the error variance in choices between (non-nested) alternatives of a particular user group. Nest parameters, instead, relate to the choice between alternatives in one nest (independent of the user type). An interesting question is under which conditions the two types of scale parameters may be equal and have the same behavioural implications. To answer this, we examine under what conditions the resulting choice probabilities (P_{in}) in (4) and (3) would be equivalent; that is, under which conditions a NL model can be written as a HL model with scale parameters related to groups with (possibly) different choice sets.

This is shown formally in Box 1.

$$\frac{(\sum_{j=1}^{J_m} e^{\mu_m V_{jn}})^{\frac{\mu}{\mu_m}}}{\sum_{m=1}^{M}(\sum_{j=1}^{J_m} e^{\mu_m V_{jn}})^{\frac{\mu}{\mu_m}}} * \frac{e^{\mu_m V_{in}}}{\sum_{j=1}^{J_m} e^{\mu_m V_{jn}}} \rightarrow \frac{e^{v_{g_n} V_{in}}}{\sum_{j \in C_{g_n}} e^{v_{g_n} V_{jn}}}$$

if and only if:

I) $C_{g_n} = 1, \dots, J_m$

II) $\mu_m = v_{g_n}$

and III) $\dfrac{(\sum_{j=1}^{J_m} e^{\mu_m V_{jn}})^{\frac{\mu}{\mu_m}}}{\sum_{m=1}^{M}(\sum_{j=1}^{J_m} e^{\mu_m V_{jn}})^{\frac{\mu}{\mu_m}}} = \begin{cases} 1 & for\ i \in C_{g_n} \\ 0 & otherwise \end{cases}$

Box 1: Mathematical conditions for NL model equalling a HL model with user-group specific choice sets

This implies that the estimated group scale parameters v_{g_n} could only be used as nest parameters (in a mathematical sense) if alternatives were nested according to the group-specific choice sets (condition I) and if the choice between nests was deterministic (condition III). This puts hard/impractical restrictions to the methodological correctness of a naive translation and, indeed, strong assumptions are required in practical approaches (see section 3.3.).

The fundamental reason for the immediate mathematical incompatibility between group scale and nest parameters goes back to the fact that the former are user group specific while the latter are travel mode specific. This point is essential also for the interpretation of scale parameters discussed in the next section.

3.2. Source and interpretation of scale parameters

The inverse proportionality of the scale parameters to the error variances implies that the (classical) sources of the error term in discrete choice models (Manski 1973, Ortúzar and Willumsen 2011), that is, unobserved attributes, taste heterogeneity, measurement errors and use of instrumental/proxy variables, are the possible main sources of the utility scale parameters.

We recall that the NL and HL models are both more flexible than the MNL model as they relax the IID assumption of the error terms (which is often a restrictive assumption in practise). The relaxation of the IID assumption by the NL model is based upon the fact that the error term of the utility functions of different travel modes are correlated. Travel modes with (significant) positive correlation should be candidates to be nested together. The idea is to account for non-proportional substitution patterns caused by the correlated error terms. The typical interpretation is that travel modes that are closer substitutes (those nested together) share unobserved attributes (Williams 1977). Relaxing the IID assumption and accounting for the patterns of unobserved attributes can be important, as illustrated by the well-known blue bus/red bus paradox (Mayberry 1973, Ortúzar and Willumsen, page 214).

The relaxation of the IID assumption in the HL model stems instead from different error variances associated with different subsets of the data. Various reasons for error variances (and thereby scale parameters) to differ are possible. An obvious candidate is the potentially variable impact of unobserved attributes between travel modes involved in the different choice sets. Common sense suggests, for example, that HSR should share more unobserved attributes with the traditional train than with car, in which case the binary choices between train and HSR are less affected by those unobserved factors. On the other hand, the choice between HSR and car is likely to be affected by several unobserved attributes that differ between the modes (i.e. the varying utility associated with having a car available at the point of destination), making the overall impact of the error term more important.

For example, if the unobserved, and varying, "need to have a car at destination" has greater importance, the superior ("observed") LoS of HSR might not impact the choice probabilities between car and HSR that much. Thus, a relatively low

group scale parameter for the car user group should be expected. From this perspective it seems reasonable to use information about group parameter scales to derive a NL forecasting model

However, the scale parameters for the different binary choices might also be high when the taste heterogeneity of users within the subgroup is relatively low. Taste heterogeneity is, to a large degree, unobserved as it involves unobserved factors/preferences relating to the users. The difference with the unobserved attributes discussed in the previous paragraph is that the latter are travel mode specific while taste heterogeneity is person-specific (or user-group specific). Car users, for instance, might have less homogenous preferences of the (observed) LoS than train users (e.g. the subjective Value of Time might vary more among car drivers than among train users). From this perspective, the error variance in the car/HSR choices might be higher than in the train/HSR choices. If (user group specific) taste heterogeneity is the predominant source of the error variance, the estimated scale parameters are not suitable to represent (travel mode specific) nest parameters; the translation of "variance" into "correlation" would not be sound in this case.[5]

3.3. Validity of Practical Approaches

In this section we discuss the validity of (and the necessary assumptions required for) practical approaches to construct a hierarchical forecasting model based on binary SC data. Striving for such a model (instead of applying a simple MNL model ignoring the different sizes of group scale parameters), acknowledges the fact that there might be indeed non-proportional substitution patterns between travel modes worth accounting for when predicting choice probabilities.

The mathematical conditions (section 3.1.) require making some assumptions. A HL model as in equation (3), does not include the respondent's 'choice' about which user group s/he belongs to (this is predefined by the researcher based on the non-modelled RC). Thus, the scale in choices between current modes is unobserved. In the absence of further information, it is necessary to assume the relative scale at the upper level of that hierarchical forecasting model. The choice is restricted by the fact that the scale at the upper level in a NL or CNL model cannot be larger than the scale at the lower level. A sensible choice, resulting in the least complex implicit structure, is to assume that the scale at the upper level equals the lowest estimated group scale parameter yielding a degenerate nest with

---

[5] Another potential source of group scale variability are the different degrees of measurement errors in the subsamples. Arguably this is not an issue in CE where attribute values are directly coded as they are presented in the respondent's screen. The use of proxy variables can also be a source for different scale parameters. This applies when a specified proxy variable is a precise representation of the actual variable for the binary choices of one user group, but an imprecise one for the choices of another user group. We will not discuss this further here but maintain the assumption made at the introduction, that the specification of the deterministic utility function could be done "equally well" for all user groups.

the current travel mode characteristic for the user group with the lowest group scale parameter.

Based on the discussion in section 3.2, it is evident that we have to make sure that the group scale parameters do represent different substitution patterns of transport modes. For now we assume that taste heterogeneity could be controlled for in the systematic utility function and that measurement errors and proxy variables are not an issue. With these assumptions, different error variances in subsamples can indeed be interpreted as representing different substitution patterns (correlation) across travel modes and HSR should be nested with the current travel mode(s) associated with the highest group scale parameter(s). If two (or more) group scale parameters are different from each other, the different degrees of correlation between HSR and the corresponding transport modes should be taken into account with a CNL specification.

To make things more specific, Table 1 discusses four potential cases of estimated group scale parameters. The group scale parameter for car-users is fixed to unity in these examples. If all estimated group scale parameters were close (and insignificantly different) to unity (i.e. case 1 in Table 1), a MNL would be obtained.

Table 1: Possible nesting structures suggested by group scale parameters (SC data only)

	Case 1	Case 2	Case 3	Case 4
Estimated group scale parameter in SC "current mode vs. HSR"	car-users $\equiv 1$; train-users ≈ 1; air-user ≈ 1;	car-users $\equiv 1$; train-users ≈ 1; air-user ≈ 3;	car-users $\equiv 1$; train-users ≈ 3; air-user ≈ 3;	car-users $\equiv 1$; train-users ≈ 2; air-user ≈ 3;
Proposed nest 1	-	car	Car	car
Proposed nest 2	-	train	train, air, HSR	train, HSR
Proposed nest 3	-	air, HSR	-	air, HSR
Proposed structure of forecasting model*	MNL	NL	NL	CNL

*Under the assumption that the overall scale (at the upper level) is one.

As much of the discussion provided here is (implicitly) about the reasonability of translating "variance" into "covariance", it is useful to take a closer look at the covariance structure associated with the estimation and forecasting model. Let **x** and **y** denote the following vectors:

$$(8) \quad x' = \begin{pmatrix} U_{car-user,car} \\ U_{train-user,train} \\ U_{air-user,air} \\ U_{car-user,HSR} \\ U_{train-user,HSR} \\ U_{air-user,HSR} \end{pmatrix}, \quad y' = \begin{pmatrix} U_{car} \\ U_{train} \\ U_{air} \\ U_{HSR} \end{pmatrix}$$

Then, Cov $(x'x)$ is the covariance matrix consistent with the HL model, while Cov $(y'y)$ would be the covariance structure for the proposed forecasting model. For case 1 we would need to translate:

$$(9) \quad Cov(x'x) = \frac{\pi^2}{6} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \text{ into } Cov(y'y) = \frac{\pi^2}{6} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

While being logical, the validity of this translation rests on the aforementioned assumptions regarding the size of the upper level scale parameter, which in this case implies that the scale in choices between current modes is assumed to be equal to the scale in choices between single current modes and HSR, which happens to be the same. If there is only SC survey data, the correctness of this assumption is not testable (without additional data on choice between all alternatives) due to the inherent missing information in binary choice data with only one common travel mode.

In case 2, if the group scale parameter for air-users was estimated as significantly higher than those for the remaining groups, this would indicate that air and HSR are closer substitutes and a NL structure with air and HSR in one nest and two degenerate nests for car and train could be proposed. In that case the following translation would apply:

$$(10) \quad \frac{\pi^2}{6} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & (1/3)^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & (1/3)^2 \end{pmatrix} \text{ into } \frac{\pi^2}{6} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & (1-1/3)^2 \\ 0 & 0 & (1-1/3)^2 & 1 \end{pmatrix}$$

Thus, the variance in the binary choices between air and HSR for current air users would be used to set the covariance between air and HSR for the full forecasting model (via equation 5). This is only valid if the group scale parameters can really be interpreted as accounting for different degrees of similarity (related to unobserved attributes) regarding the transport modes (see the discussion above). Given the above assumption of $\mu = 1$, the group scale parameter estimated as

equal to three can be directly used as the structural parameter in the nest containing air and HSR.

In case 3, if the scale parameter for train-users is estimated as significantly greater than one and insignificantly different from the air-user scale parameter, train, air and HSR might be nested together. Similar to the second case, the following correlation structure in the forecasting model would be proposed:

(11)

$$\frac{\pi^2}{6}\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & (1/3)^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & (1/3)^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & (1/3)^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & (1/3)^2 \end{pmatrix} \text{ into } \frac{\pi^2}{6}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & (1-1/3)^2 & (1-1/3)^2 \\ 0 & (1-1/3)^2 & 1 & (1-1/3)^2 \\ 0 & (1-1/3)^2 & (1-1/3)^2 & 1 \end{pmatrix}$$

Note, that the correlation between the train and air alternatives is derived from the correlation between train and HSR, and air and HSR (assuming that taste heterogeneity is controlled for). This is not obvious and cannot be assessed without choice data between train and air (see the discussion in section 4).

Finally, in case 4, if the train-user scale is greater than one but significantly lower than the air-users scale, the only valid option would be to allow for HSR entering one nest with train and another nest with air. In that case a CNL model would be required[6] and the following correlation structure would be desirable[7]:

(12)

$$\frac{\pi^2}{6}\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & (1/2)^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & (1/3)^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & (1/2)^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & (1/3)^2 \end{pmatrix} \text{ into } \frac{\pi^2}{6}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & (1-1/2)^2 \\ 0 & 0 & 1 & (1-1/3)^2 \\ 0 & (1-1/2)^2 & (1-1/3)^2 & 1 \end{pmatrix}$$

Note that it may be difficult, in application, to find a CNL model that implies this correlation structure, as the choice of allocation parameters in conjunction with the nest parameters is a non-trivial task. This suggests estimating a CNL model from the data rather than to try to infer one such model from a HL model (see section 4).

3.4. Empirical Illustration on Own SC Data

This subsection provides some estimation results that supplement the theoretical discussion of the previous sections. We use data from an independent SC study conducted by the Institute of Transport economics (TØI) in 2010 (Halse 2012,

---

[6] Apart from a CNL model, a fully general mixed logit (ML) model (Train 2009), might be an alternative and provide an even better way to handle this issue at the expense of more complex estimation, interpretation and application.

[7] As for case 3, the correlation assumed for train and air is somewhat arbitrary. It might be reasonable to allow for correlation between air and train as well, but this cannot be directly derived from the given binary data alone.

Flügel and Halse 2012a). Similar to the official assessment study (Jernbarnverket 2012, Atkins 2012), the SC consisted of binary choices and were pivoted on observed RC data (Figure 1). In fact, the RC stem from an on-side, pen-and-pencil study that asked travellers to provide general information about their current mode choice in the main long distance corridors in Norway[8] (Denstadli and Gjerdåker 2011). In the last item, travellers were asked to leave their e-mail address to receive a web-based survey concentrating on high-speed rail.

In the SC-survey, each respondent had to make 14 choices between its current mode of transport (as observed in the on-side study) and a hypothetical HSR. The attributes characterizing the transport modes were: total travel costs, in-vehicle travel time, travel time to station/airport ('access time'), travel time from station/airport ('egress time'), frequency (number of departures per day) and the share of the ride spent in tunnels ('tunnel share'). In the first eight choice tasks ('CE1'), the attributes of the current mode were kept fixed to their reported values, while they varied within certain percentage changes in the last six choice tasks ('CE2') (see details in Halse 2012). CE2 included also an opt-out option ('neither of the two alternatives'), which was, however, seldom chosen and not considered in the models of this paper.

A sample of 893 respondents completed the online SC-study (about 33% of the invited respondents). We focus here on the subsample of leisure trips for which 607 respondents were considered. The general choice behaviour of the subsample is summarised in Table 3.

Table 3: Sample size of user groups and general choice behaviour (leisure trips)

User group as defined by the RC choices	Group size RC	Group size SC*	Percent of SC choices (%)			Percent of respondents always choosing only one mode in SC ("Non-Traders") (%)		
	Final sample	Final sample	Current mode	HSR	Opt-out	Current mode	HSR	Switch between
Car	3833	320	61.6	37.7	0.6	31.3	13.8	55.0
Air	920	76	34.5	64.7	0.8	1.3	31.6	67.1
Train	2867	176	40.8	57.2	2.0	6.3	18.2	75.6
Bus	480	35	41.0	58.5	0.6	0.0	17.1	82.9

*Compared to a representative dataset (Denstadli and Gjerdåker 2011), we have under-sampled current air users somewhat and over-sampled current car and train-users. External weights were used during estimation.

Car drivers are least likely to choose HSR in the SC data and a considerable share of car-users (31.3%) choose car over HSR in each of the 14 SC situations. This indicates that unobserved factors may have affected many of the choices between car and HSR.

Table 4 provides estimation results for HL models on pooled data of different binary SC. As a first benchmark, we include a model where all group scale parameters are fixed to one; in this case the HL model collapses to an MNL

---

[8] Oslo-Trondheim, Oslo-Bergen and Stavanger-Bergen. For the SC-study, only the former two corridors were considered.

model. The difference between SC_HL_1 and SC_HL_2 is that the latter has random coefficients (normally distributed over decision makers) related to the most important level-of-service (LoS) attributes: in-vehicle time, access/egress time, travel cost and (the inverse) of the frequency measured as the number of departures per day. All models were estimated with BIOGEME (Bierlaire 2003, 2008).

Table 4: MNL and HL models on SC data

Model Index	SC_MNL		SC_HL_1		SC_HL_2 (random coefficients*)	
Coefficient	Value	Rob. t-stat (0)	Value	Rob. t-stat (0)	Value	Rob. t-stat (0)
Travel cost (NOK)	-0.00283	-10.45	-0.00189	-6.69	-0.00963	-6.08
sigma cost					0.00551	4.19
Interaction: Dummy "missing income" - travel cost	-0.00709	-3.1	-0.00476	-3.85	-0.00778	-1.27
Interaction: Dummy "did not pay" - travel cost	0.00155	2.87	0.00105	2.92	0.0051	10.32
In-vehicle* (min)	-0.0023	-2.41	-0.00176	-2.98	-0.0155	-5.64
sigma in-vehicle time					0.0119	8.95
Access + egress time** (min)	-0.00169	-0.99	-0.00235	-2.29	-0.00842	-4.82
sigma acc+eg time					0.034	2.75
Dummy (travel time <6h)	0.496	3.06	0.229	2.11	0.145	1.03
1/frequency	-1.15	-3.09	-0.474	-2.68	-2.75	-4.34
sigma 1/ frequency					4.37	5.75
Tunnel share (%)	-0.00232	-0.47	-0.00251	-0.9	-0.0175	-2.74
ASC-HSR	0.248	0.86	0.184	0.95	0.724	1.53
ASC-Car	0	fixed	0	fixed	0	fixed
ASC-Air	-0.819	-2.92	-0.511	-2.98	-1.42	-1.73
ASC-Train	-0.111	-0.5	0.00269	0.02	0.218	0.37
ASC-Bus	0.3	0.96	0.374	2.08	0.331	0.63
Group scale parameters	Value	Rob. T-stat (1)	Value	Rob. T-stat (1)	Value	Rob. T-stat (1)
Car-users	1	fixed	1	fixed	1	fixed
Air-users	1	fixed	1.88	1.89	2.51	2.25
Train-users	1	fixed	2.74	3.43	1.75	1.81
Bus-users	1	fixed	4.35	2.66	2.42	1.77
No. of parameters		12		15		19
No. of observations		8402		8402		8402
No. of respondents		607		607		607
Null-LL		-5822.44		-5822.44		-5822.44
Final-LL		-4677.25		-4572.47		-2612.68
Adjusted rho-square		0.195		0.215		0.548

*Using 500 Halton draws

Comparing first the MNL model with the HL model, we see that the latter has a considerably better final log-likelihood statistic indicating that the inclusion of the group scale parameters improved the estimation on the joint SC data set.

As none of the estimated group scale parameters was below one, the lowest group scale parameter (e.g. the highest error variance) is related to the choices between car and HSR. This seems to fit well with the intuition that car and HSR share the least unobserved attributes with each other (see the discussion above). Surprisingly, the highest group scale parameter appears to be the one corresponding to the SC choices of bus-users. A naive interpretation of this result would indicate that bus and HSR are the closest substitutes and that this should be considered in a forecasting model by nesting bus and HSR in the nest associated with the highest structural parameter. However, from the discussion in 3.2 we should recall that different degrees of unobserved taste heterogeneity in the different subsamples (user groups) should be considered as well.

The estimation results for the random coefficients model (SC_HL_2) show controlling for taste heterogeneity among decision makers lead to considerable changes in the estimated group scale parameters. For example, the group scale parameters for bus and train users are reduced while that for air users is increased. All group scale parameters are not significantly different from two. In the context of finding a plausible structure for a forecasting model, this may suggest a NL model with a (degenerate) nest for the car alternative and a single nest including all public transport options (Figure 3).
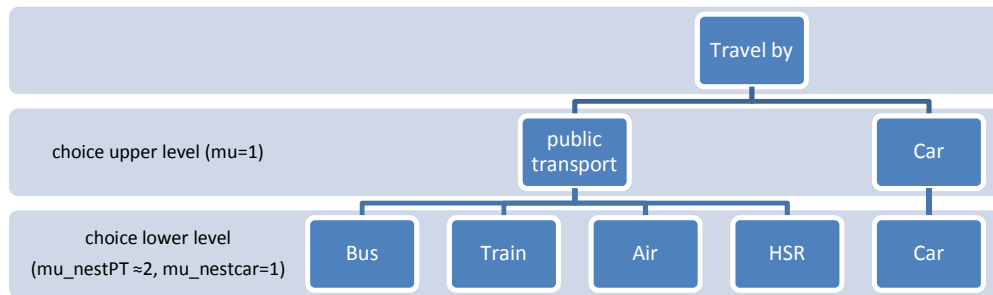

Figure 3: A possible nesting structure for a forecasting model as suggested from SC

While being intuitive and more plausible than what might have been suggested from the results that do not control for unobserved taste heterogeneity, this derivation still rests on two strong implicit assumptions required as a consequence of the missing data issue in the binary stated choices: (i) that the scale between both nests (car and "public transport") equals the scale in the binary choice between car and HSR and (ii) that the correlation in the choices between the current public transport modes (bus, train and air) is derived from the scale in the binary choice between these modes and HSR.

4. Using Additional RC Data among Current Travel Modes

4.1. Motivation

The typical motivation for additional RC data and the joint SC-RC paradigm is the need to ground the SC models in reality (Louviere *et al.* 2000). We will not discuss here the "classical" method of rescaling the SC scale by the RC scale, which became popular after the seminal work of Taka Morikawa (Morikawa 1989; Ben-Akiva and Morikawa 1990) and which is relevant for any kind of SC data (both binary and multinomial). In our context, RC data may provide some of the missing information inherent in binary SC data. In what follows, the focus will be on the correlation structure among current travel modes, which is not "observable" using binary SC data alone.

As RC data involves more than two alternatives in the respondents' choice sets, it is possible and meaningful to estimate hierarchical logit models (NL or CNL) on the RC data. The correlation obtained in a RC model can provide information needed to define plausible correlation structures in the full forecasting model. That is, if estimations on our RC model indicate a nesting structure with one degenerate car nest and another nest including all public transport alternatives (air, train and bus), then the proposed structure from our SC-data in Figure 2 would get empirical support.

4.2. Empirical illustration with RC and SC/RC models

Our RC data includes all relevant travellers in the on-side study (see section 3.4) independent of whether they left or not an e-mail address or whether they were included in the SC study (see sample size in Table 2). Based on the reported geographical information of the trips' start and ending locations, we imported zonal level-of-service data for the related O-D pairs from the Norwegian National Travel Model (Hamre *et al.* 2002[9]).

We tested all possible nesting structures for the four alternatives in the RC dataset (car, bus, train and air), including the same explanatory variables as in the SC data in Table 3[10]. Somewhat surprisingly, the structure shown in Figure 4 had clearly the best fit to our RC data.

An alternative NL model, where air is nested together with bus and train did perform considerably worst (see Table A1 in the appendix for estimation results). Hence, our RC data did not provide immediate support for the correlation structure suggested from the SC-data alone (Figure 3), as air seems to be more highly correlated with car than with bus or train.

---

[9] The LoS data contain representative values for the relevant zone pairs and are, in some instances, not updated, such that the RC data must be considered as rather imprecise.

[10] The tunnel attribute was not available from zonal data and was therefore omitted in the RC dataset. Full choice sets were assumed for all decision makers except for the fact that car was only available to respondents that reported owning a car.
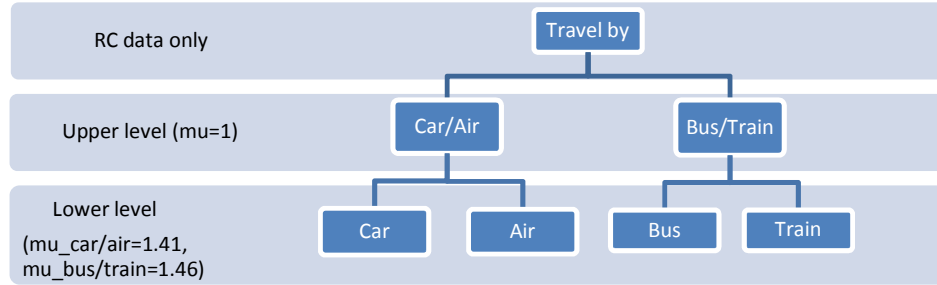
Figure 4: Nesting structure indicated from RC data alone

Combining the information from the RC and SC datasets, the nesting structure shown in Figure 5 might be proposed.
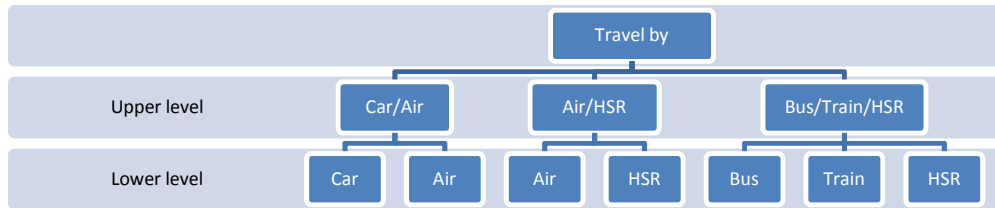

Figure 5: A possible nesting structure for a forecasting model suggested from the information in SC and RC

This structure takes into account the correlation patterns between the current modes as indicated in the RC model and the information from the SC model that HSR is a closer substitute to public transport modes than to car. The nest air/HSR is included (separated from the nest of other public transport modes) to acknowledge the indication from our RC data that the utility for air is not correlated with the utility from bus and train. Both HSR and air enter two nests, thus a CNL model seems to be required.

Two CNL models with the implicit structure in Figure 5 were estimated on the pooled RC/SC data to infer the underlying parameters values (Table 5). The models assume generic coefficients in RC and SC[11]. CNL_1 uses only fixed coefficient while CNL_2 replicates the specification of model SC_HL_2, assuming normal distributed coefficients for the most important LoS-variables.[12]

We can use (7) to approximate the inter-alternative variance-covariance matrix for the two model versions[13] as shown in Box 2.

Table 5: Cross-nested logit models on pooled RC/SC data

---

[11] This is a restrictive assumption and, indeed, seems not to hold for our data as indicated by a likelihood ratio tests (Ortúzar and Willumsen 2011, p. 325). We suspect that the main reason for this is the different measure of attributes in RC and SC; however, it could also be that preferences change when the HSR gets available in the choice sets (see also footnote 13 in section 5).

[12] The estimated parameter for nest bus/train/HSR is not significant different from one which might lead to the suggestion to collapse this nest. However, the value is high, indicating that the related correlation might be important (despite the result not being very reliable).

[13] The actual correlation structure in estimation model CNL_2 may also be affected by the random terms underlying the normal distributed error terms.

Model Index	CNL_1		CNL_2**	
Coefficient	Value	t-stat (0)	Value	t-stat (0)
Travel cost (NOK)	-0.00264	-31.95	-0.00684	-37.66
sigma cost			0.00308	0***
Interaction: Dummy "missing income" - travel cost	-0.00061	-6.25	-0.00061	-2.04
Interaction: Dummy "did not pay" - travel cost	0.001	7.49	0.0029	13.02
In-vehicle* (min)	-0.00094	-5.25	-0.00822	-13.87
sigma in-vehicle time			0.0128	15.84
Access + egress time** (min)	-0.00533	-18.93	-0.0276	-19.13
sigma acc+eg time			0.0193	21.48
Dummy (travel time <6h)	0.389	11.25	0.188	2.16
1/frequency	-0.298	-5.74	-4.9	-7.51
sigma 1/ frequency			6.63	8.81
Tunnel share (%)	-0.00404	-2.84	-0.0181	-6.79
ASC-HSR (SC)	0.945	12.23	2.76	13.86
ASC-Air (SC)	-0.372	-4.44	0.0144	0.09
ASC-Train (SC)	0.204	3.14	1.94	8.49
ASC-Bus (SC)	0.363	4.09	2.41	4.89
ASC-Air (RC)	0.947	10.35	0.41	1.53
ASC-Train (RC)	-0.0641	-1.61	1.66	9
ASC-Bus (RC)	-0.193	-3.54	0.307	1.52
Structural parameters	Value	T-stat (1)	Value	T-stat (1)
Car/Air	1.8	5.3	7.84	12
Air/HSR	1.57	1.4	6.99	2.35
Bus/train/HSR	4.57	10.61	5.88	0.94
Allocation parameters*	Value	T-stat (1)	Value	T-stat (1)
Air to nest Car/Air	0.845	-2.18	0.232	-16.53
Air to nest Air/HSR	0.155	-11.83	0.768	-4.99
HSR to nest Air/HSR	0.527	-12.91	0.595	-6.19
HSR to nest Bus/train/HSR	0.473	-14.36	0.405	-9.09
No. of parameters		22		26
No. of observations		16852		16852
No. of respondents		9057		9057
Null-LL		-16868.4		-16868.4
Final-LL		-11099.5		-9310.01
Adjusted rho-square		0.341		0.447

* The remaining allocation parameters were fixed to 0 or 1 according to Figure 5.
** We used 500 Halton draws.
*** Seemingly some numerical issues were present in the estimation of this standard deviation.

$$y' = \begin{pmatrix} U_{car} \\ U_{bus} \\ U_{train} \\ U_{air} \\ U_{HSR} \end{pmatrix}$$

$$\text{Cov}^{CNL\_1}(y'y) \approx \frac{\pi^2}{6}\begin{pmatrix} 1 & 0 & 0 & 0.64 & 0 \\ 0 & 1 & 0.95 & 0 & 0.65 \\ 0 & 0.95 & 1 & 0 & 0.65 \\ 0.64 & 0 & 0 & 1 & 0.17 \\ 0 & 0.65 & 0.65 & 0.17 & 1 \end{pmatrix}$$

$$\text{Cov}^{CNL\_2}(y'y) \approx \frac{\pi^2}{6}\begin{pmatrix} 1 & 0 & 0 & 0.47 & 0 \\ 0 & 1 & 0.97 & 0 & 0.62 \\ 0 & 0.97 & 1 & 0 & 0.62 \\ 0.47 & 0 & 0 & 1 & 0.66 \\ 0 & 0.62 & 0.62 & 0.66 & 1 \end{pmatrix}$$

Box 2: Correlation pattern suggested by CNL models

The main difference between CNL_1 and CNL_2 is that the latter model (i.e. that controls for unobserved taste heterogeneity) suggests a higher correlation between HSR and air (this corresponds to the comparison between models SC_HL_1 and SC_HL_2 in Table 4). The indicated correlation has approximately the same magnitude as the correlation between HSR, train and bus; something that seems plausible.

It has to be underlined that covariance structure cannot be transformed to other scenarios. They are particular to our data (both RC and SC) and to the specification of our model (i.e. the predefined structure of the CNL model and the chosen deterministic utility function).

With the combined RC/SC modelling, the relative scale between the upper and lower levels is estimated from the data and does not need to be assumed as in the method of translating group scale parameters (from binary SC data) to structural parameters (section 3).

5. Discussion

Modelling the choice of a new alternative (in this case HSR) is a non-trivial task. One important reason for this goes back to the limited data access to revealed choice (RC) data for new travel modes making the collection of stated choice (SC) data a necessity. The papers by Cherchi and Ortúzar (2006; 2011) and Yánez *et al* (2010) have addressed important challenges in the combined analysis of RC and (multinomial) SC data. They discuss how to fit alternative specific constants, to account for taste heterogeneity and to define inter-alternative error structures respectively, and have attempted to provide guidelines on how to cope with these challenges in practice. Our paper acknowledges that analyst's judgment is needed to determine the best way to fit models to SC data for real world forecasting of new alternatives in specific application scenarios. This applies in particular to

situations where the new travel alternative is likely to change the competitive structure of the travel market, as is arguable the case for HSR in Norway.

To the extent that introducing a HSR has the potential to change the correlation structure among current modes, it is not guaranteed that information about the current correlation structure among existing travel modes - as indicated by a RC model - has guaranteed validity for future travel decision making[14].

Despite these caveats it would have been interesting to compare the correlation structure in RC and SC models more rigorously. However, a direct comparison as done by Yáñez *et al* (2010) based on multinonial SC data, is not possible with *binary* SC data, with only one alternative (the new travel mode) being common to all subgroups. Therefore, identifying the most appropriate inter-alternative correlation structure for a forecasting model (i.e. preferring the nesting structure of Figure 5 from those in figures 3 and 4) is somewhat arbitrary and subject to the assumption that the translation from "variance" into "correlation" related to the SC data on HSR is reasonable.

Given the shortcomings of SC binary choice data as discussed in this paper, it seems indispensable to consider having at least three alternatives in choice experiments, even though this is likely to increase the complexity of the choice tasks. Good practise is found in Yáñez *et al*. (2010) where each SP-respondent had to consider four transport modes: the current mode, the new HSR and two other transport modes that were added to the choice experiment on a random basis.

Acknowledgements

---

[14] Moreover, the introduction of HSR in the choice set may change preferences for LoS-attributes (i.e. beta-coefficients). For example, it is possible that travellers get more sensitive to travel time, in particular that of the air alternative, when HSR is available in the choice set. Therefore, it is recommended to test if the assumption of common coefficients in the RC and SC models holds (Ortúzar and Willumsen 2011, p. 325). A potential challenge arises if the measures of attributes (to whom the beta-coefficients apply) are very different in the RC and SC datasets. In our empirical case study, the nature of the RC and SC was different for two reasons: (i) the former consists of real choices and the latter of hypothetical choices; (ii) attributes in the SC set were pivoted around reported respondents' values (and are, therefore, personal-specific) while attributes in the RC set were inferred from zonal data (and, therefore, representative for the O-D zones of the trips). In this case, it is difficult to determine if the higher estimated value of time from the SC model is due to different measurement of the attributes or due to preference changes when HSR is available.

Foundations, and the TEMPO Project financed by the Research Council of Norway.

References

Abbe, E., M. Bierlaire and T. Toledo (2007) Normalization and correlation of cross-nested logit models. *Transportation Research* **41B**, 795-808.

Atkins (2012) Norway High Speed Rail Assessment Study: Phase III - Model Development Report. http://www.jernbaneverket.no/PageFiles/17406/Economic%20and%20Financ ial%20Analysis%20-Final%20Report%20Atkins.pdf. (retrieved 2.12.2014)

Bierlaire, M. (2003) BIOGEME: a free package for the estimation of discrete choice models*, Proceedings of the 3rd Swiss Transportation Research Conference,* Ascona, Switzerland.

Bierlaire, M. (2006) A theoretical analysis of the cross-nested logit model. *Annals of Operations Research* **144**, 287-300.

Bierlaire, M. (2008) An introduction to BIOGEME Version 1.8. http://biogeme.epfl.ch/v18/tutorialv18.pdf (retrieved 27.3.2013)

Ben-Akiva, M. and Bierlaire, M. (1999) Discrete choice methods and their applications to short-term travel decisions. In R. Hall (ed.), *Handbook of Transportation Science*. Kluwer Academic Publishers, Dordrecht.

Ben-Akiva, M. and T. Morikawa (1990) Estimation of switching models from revealed preferences and stated intentions. *Transportation Research* **24A,** 485-495.

Bhat, C.R. (1995) A heteroscedastic extreme value model of intercity travel mode choice. *Transportation Research* **29B,** 471-483.

Bhat, C.R. (1997) Covariance heterogeneity in nested logit models: econometric structure and application to intercity travel. *Transportation Research* **31B,** 11-21.

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

Cherchi, E. and Ortúzar, J. de D. (2006) On fitting mode specific constants in the presence of new options in RP/SP models. *Transportation Research* **40A**, 1-18.

Cherchi, E. and Ortúzar, J. de D. (2011) On the use of mixed RP/SP models in prediction: accounting for random taste heterogeneity. *Transportation Science* **45**, 98-108.

Daly, A. and Zachary, S. (1978) Improved multiple choice models. In D.A. Hensher and M.Q., Dalvi (eds.), *Determinants of Travel Choice*. Saxon House, Sussex.

Denstadli, J.M. and Gjerdåker, A. (2011) Transportmiddelbruk og konkurranseflater i tre hovedkorridorer, Institute of Transport Economics, Oslo (in Norwegian).

Flügel, S. and Halse, A.H. (2012) Non-linear utility specifications in mode choice models for high-speed rail. *Kuhmo Nectar Conference 2012*, Berlin.

Halse, A.H., 2012. Demand for high speed rail in Norway: Survey design and data collection. Working Paper, Institute of Transport Economics, Oslo.

Jernbaneverket (2012) Høyhastighetsutredningen 2010-2012, Konklusjoner og oppsummering av arbeidet i Fase 3, Del1 (in Norwegian). http://www.jernbaneverket.no/PageFiles/17299/Rapport_Del_1.pdf. (retrieved 2.12.2014)

Louviere, J.J., Hensher, D.A. and Swait, J.D. (2000) *Stated Choice Methods: Analysis and Application.* Cambridge University Press, Cambridge.

Manski C.F. (1973) *The Analysis of Qualitative Choice.* Ph.D. Dissertation., Department of Economics, MIT.

Morikawa, T. (1989) *Incorporating Stated Preference Data in Travel Demand Analysis*. Ph.D. Dissertation, Department of Civil Engineering, MIT.

Mayberry, J.P. (1973) Structural requirements for abstract-mode models of passenger transportation. In R.E. Quandt (ed.), *The Demand for Travel: Theory and Measurement*. D.C. Health and Co., Lexington, Mass.

McFadden, D. (1974) Conditional logit analysis of qualitative choice behaviour. In P. Zarembka (ed.), *Frontiers in Econometrics.* Academic Press, New York.

McFadden, D. (1978). Modelling the choice of residential location. In A. Karlquist et al., editor, Spatial interaction theory and residential location, pages 75–96, Amsterdam. North-Holland.

McFadden, D. (1981) Econometric models of probabilistic choice. In C.F. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*. The MIT Press, Cambridge, Mass.

Ortúzar, J. de D. and Willumsen, L.G. (2011) *Modelling Transport*. Fourth Edition, John Wiley and Sons, Chichester.

Papola, A. (2004) Some developments on the cross-nested logit model. *Transportation Research* **38B**, 833–851.

Train, K.E. (2009) *Discrete Choice Methods with Simulation*. Second Edition, Cambridge University Press, Cambridge.

Vovsha, P. (1997) Cross-nested logit model: an application to mode choice in the Tel-Aviv Metropolitan Area. *Transportation Research Record* **1607**, 6–15.

Walker, J.L. (2001) *Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures and Latent Variables.* Ph.D. Thesis, Department of Civil and Environmental Engineering, MIT.

Williams, H.C.W.L (1977) On the formation of travel demand models and economic evaluation measures of user benefits. *Environment and Planning* **9A**, 285-344.

Wen, C.H. and Koppelman, F.S. (2001) The generalized nested logit model. *Transportation Research* **35B**, 627-641.

Yáñez, M.F., Cherchi, E. and Ortúzar J. de D. (2010) Defining inter-alternative error structures for joint RP-SP modelling: some new evidence. *Transportation Research Record* **2175**, 65-73.

Appendix

Table 5: Multinomial and Nested logit model on revealed choice data

Model Index	RC_MNL		RC_NL_1		RC_NL_2	
Coefficient	Value	Rob. T-stat (0)	Value	T-stat (0)	Value	Rob. T-stat (0)
Travel cost (NOK)	-0.0064	-19.17	-0.0064	-22.04	-0.00526	-14.24
Interaction: Dummy "missing income" - travel cost	-3.03E-06	-0.01	-3.03E-06	-0.02	0.000271	1.38
Interaction: Dummy "did not pay" - travel cost	0.000943	2.3	0.000943	2.23	0.000817	2.21
In-vehicle* (min)	-0.00087	-1.42	-0.00087	-1.21	-0.00066	-1.27
Access + egress time** (min)	-0.00678	-6.99	-0.00678	-7.45	-0.00694	-7.63
Dummy (travel time <6h)	1.18	7.37	1.18	6.78	1.04	7.19
1/frequency	-0.591	-2.48	-0.591	-2.51	-0.655	-3.02
ASC-Car	0	fixed	0	fixed	0	fixed
ASC-Air	2.54	8.88	2.54	8.58	1.87	6.86
ASC-Train	-0.486	-3.26	-0.486	-3.34	-0.503	-3.73
ASC-Bus	-1.35	-7.46	-1.35	-6.52	-1.21	-7.53
	Value	Rob. T-stat (1)	Value	T-stat (1)	Value	Rob. T-stat (1)
Nest car/air	1	fixed			1.41	17.44
Nest train/bus	1	fixed			1.46	7.05
Nest car			1	fixed		
Nest air/train/bus			1.00	0.89		
No. of parameters		10		11		12
No. of observations		5406		5406		5406
No. of respondents		5406		5406		5406
Null-LL		-5767.51		-5767.51		-5767.51
Final-LL		-2360.81		-2360.81		-2337.68
Adjusted rho-square		0.589		0.589		0.593