Development and Application of a Mixed Cross-Nested Logit model

Stephane Hess* Michel Bierlaire† John W. Polak‡

October 30, 2004

Abstract

Research in the area of discrete choice modelling can be split into two broad categories; applications accounting for the prevalence of inter-alternative correlation, and applications concerned with the representation of random inter-agent taste heterogeneity. The difference between these two phenomena is however not as clear-cut as this division might suggest, and there is in fact a high risk of confounding between the two phenomena. In this article, we investigate the potential of Mixed Generalised Extreme Value (GEV) models to simultaneously account for the two phenomena, using a Stated Preference (SP) dataset for mode-choice in Switzerland. Initial results using more basic modelling techniques reveal the presence of both correlation and random taste heterogeneity. The subsequent use of Mixed GEV models on this dataset leads to important gains in performance over the use of the more basic models. However, the results also show that, by simultaneously accounting for correlation and random taste heterogeneity, the scope to retrieve the individual phenomena is reduced. This shows that a failure to account for the potential impacts of either of the two phenomena can lead to erroneous conclusions about the existence of the other phenomenon. This is a strong indication that the use of Mixed GEV models to jointly explain random taste heterogeneity and inter-alternative correlation in a common modelling framework should be encouraged in the case where the nature of the error-structure is not clear a priori.

*Centre for Transport Studies, Imperial College London, stephane.hess@imperial.ac.uk, Tel: +44(0)20 7594 6105, Fax: +44(0)20 7594 6102

†Institute of Mathematics, École Polytechnique Fédérale de Lausanne, michel.bierlaire@epfl.ch, Tel: +41(0)21 693 25 37, Fax: +41(0)21 693 55 70

‡Centre for Transport Studies, Imperial College London, j.polak@imperial.ac.uk, Tel: +44(0)20 7594 6089, Fax: +44(0)20 7594 6102

1 Introduction

Two main streams of model structures can be identified from the existing body of literature on discrete choice models; models concerned with representing the correlation between alternatives in the unobserved utility components, and models concerned with allowing for random variations in tastes across decision-makers.

An appropriate treatment of the correlation structure is crucial especially in the case where a model is used for forecasting of market shares after hypothetical changes to the market structure. In this case, the unrealistic substitution patterns of the Multinomial Logit (MNL) model can lead to very misleading forecasts of demand in the case where heightened correlation exists between some of the alternatives. Acknowledging the potential existence of random taste heterogeneity is similarly important. Indeed, although for reasons of interpretation, it is always preferable to as much as possible attempt to explain the variation in decision-makers' behaviour as a function of socio-demographic characteristics, the limitations of the data (along with inherent randomness involved in decision-making) mean that there is usually some remaining non-quantifiable (random) variation. By not explicitly accounting for such heterogeneity in a model, researchers not only discard valuable information about variations in choice-behaviour, but are also at risk of reaching false conclusions, most notably in the form of biased trade-offs between coefficients (c.f. Hensher & Greene 2003, Hess & Polak 2004*a*).

While the two phenomena of inter-alternative correlation and inter-agent taste heterogeneity have usually been treated in quite separate ways, it should be noted that the differences between these two phenomena are not necessarily that *clear-cut*, and that there is a significant risk of confounding. As an example, in the classic red bus/blue bus problem (c.f. Train 2003), the correlation in the unobserved utility of the two different bus types could in fact be a reflection of the existence of random taste heterogeneity in the preference for buses. Such random differences would clearly induce correlation in the unobserved utility components. As such, accounting for (arbitrary) correlation in the unobserved utility components without acknowledging the effects of random taste heterogeneity can mask the presence of the latter phenomenon. The converse can also be the case; as an example, Hess, Bierlaire & Polak (2004) have recently shown that the presence of unexplained correlated attributes across alternatives can lead to the erroneous conclusion that there are random variations in tastes across decision-makers.

The discussion presented in this article looks at the issues researchers are faced with in the case of choice scenarios where the two phenomena of inter-alternative correlation and random inter-agent taste variation potentially both have an effect on decision-making behaviour. It is in this case crucial to disentangle the two effects. The discussion also applies to the case where only one of the two phenomena is present, but where it is not clear *a priori* whether the error term reflects the presence of random taste heterogeneity or *simple* inter-alternative correlation, as caused for example by unobserved shared attributes.

Two different approaches have classically been used in the joint analysis of these two phenomena; the Multinomial Probit (MNP) model (c.f. Daganzo 1979), and more recently, the Error Components Logit (ECL) formulation of the Mixed Multinomial Logit (MMNL) model (c.f. McFadden & Train 2000). The MNP model rapidly becomes computationally intractable in the case of complex model structures; the ECL model has similar problems, and can also become difficult to formulate due to important identification issues. In this article, we illustrate the potential of an alternative approach, based on the integration of GEV-style choice probabilities over the distribution of taste coefficients, leading to a Mixed GEV model (c.f. Chernew et al. 2003, Bhat & Guo 2004). This model form not only reduces the number of random terms in the models to the number of random taste coefficients, but also avoids some issues of identification that are specific to the *ECL* formulation (c.f. Walker 2001).

The remainder of this article is organised as follows. In the following section, we give an overview of the theory, looking first at closed-form GEV models, and then at Mixed GEV models. Section 3 presents a summary of the empirical analysis conducted to explore the potential of Mixed GEV models and to highlight the issues of confounding discussed above. Finally, the fourth section gives a summary of the findings and presents the conclusions of the research.

2 Methodology

A random utility model is defined by a choice set $\mathcal{C}$ containing J alternatives, and a vector of J random utility functions

$$U = \begin{pmatrix} U_1 \\ \vdots \\ U_J \end{pmatrix} = \begin{pmatrix} V_1 \\ \vdots \\ V_J \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_J \end{pmatrix} = V + \varepsilon. \tag{1}$$

where U and ε are random vectors and $V \in \mathbb{R}^J$. Each V_i is defined by

$$V_i = f(\beta, x_i) \tag{2}$$

where x_i is a vector combining attributes of alternative i and socio-economic attributes of the decision-maker, and β is a vector of (taste-)parameters estimated from the data.

2.1 Closed-form GEV models

The family of Generalised Extreme Value (GEV) models was derived from the random utility paradigm by McFadden (1978). This family of models comprises the basic MNL model (McFadden 1974), as well as the much-used Nested Logit (NL) model (Williams 1977, McFadden 1978, Daly & Zachary 1979).

In a GEV model, the random vector of variables ε in (1) has a Cumulative Distribution Function (CDF) given by

$$F_{\varepsilon_1,\ldots,\varepsilon_J}(x_1,\ldots,x_J) = e^{-G(e^{-x_1},\ldots,e^{-x_J})}, \tag{3}$$

which is such that the marginal distribution of the individual ε terms is Gumbel (type I extreme-value). The choice of functional form for the generating function $G()$ determines the correlation structure in place between the individual ε terms, where $G()$ needs to satisfy four main conditions, as set out by McFadden (1978), and later revised by Ben-Akiva & Francois (1983).

The probability of choosing alternative i within the choice set $\mathcal{C}$ for a given choice maker is given by

$$P(i|V,C) = \frac{y_i G_i(y_1,\ldots,y_J)}{\mu G(y_1,\ldots,y_J)} = \frac{e^{V_i + \log G_i(\ldots)}}{\sum_{j=1}^{J} e^{V_j + \log G_j(\ldots)}}. \tag{4}$$

where J gives the number of available alternatives, $y_i = e^{V_i}$, V_i is the deterministic part of the utility function associated with alternative i, and $G_i = \partial G / \partial y_i$. The factor μ is the scale parameter, which, in the absence of separate population groups, is generally constrained to be equal to 1.

With the most basic choice of generating function

$$G(y) = \sum_{j \in \mathcal{C}} y_j^{\mu}, \tag{5}$$

we obtain the MNL model, in which the substitution patterns are governed by the Independence from Irrelevant Alternatives (*IIA*) assumption.

The corresponding generating function for an NL model with M nests is given by:

$$G(y) = \sum_{m=1}^{M} \left(\sum_{j \in \mathcal{C}_m} y_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}}, \tag{6}$$

where $\mathcal{C}_m$ gives the set of alternatives contained in nest m (with $m = 1, ..., M$), μ_m is the structural parameter for nest m, and where, with this notation, μ_m is constrained to be greater than 1, with the correlation between the unobserved utilities of alternatives sharing nest m being given by $1 - \frac{1}{\mu_m^2}$.

An extension of the NL generating function (equation 6) leads to a model form allowing for cross-nesting, whose generating function is given by:

$$G(y_1, \ldots, y_J) = \sum_{m=1}^{M} \left(\sum_{j \in \mathcal{C}_m} (\alpha_{jm} y_j)^{\mu_m} \right)^{\frac{\mu}{\mu_m}}, \tag{7}$$

where α_{jm} is the allocation parameter for alternative j and nest m.

The history of cross-nested Logit (CNL) models reaches back to the initial developments of the GEV family; first discussions of this structure were given by Williams (1977) and McFadden (1978). This model form has been used and analysed under different names by a number of authors, including Small (1987), Vovsha (1997), Vovsha & Bekhor (1998), Koppelman & Wen (2000), Wen & Koppelman (2001), Ben-Akiva & Bierlaire (2003), Daly & Bierlaire (2003), Bierlaire (2004), and Papola (2004). CNL models allow for ambiguous allocation of alternatives to nests, hence reflecting the different degrees of similarity between them. There are many problems in which this extra flexibility has the potential to offer considerable improvements, even in the case of a relatively low number of nests or alternatives, as illustrated for example by Bierlaire et al. (2001).

2.2 Mixed GEV models

In a *mixed* GEV model, the vector V in equation (1) is itself a random vector. In this case, the probability of choosing alternative i within the choice set $\mathcal{C}$ for a given decision-maker is given by

$$P(i \mid \mathcal{C}) = \int_V P(i|V, \mathcal{C}) \mathrm{d}V, \tag{8}$$

where $P(i|V, \mathcal{C})$ is defined as in equation (4).

Historically, the GEV model used inside the integral in equation (8) has been of MNL form, leading to the MMNL model. Two conceptually different, yet mathematically identical (as illustrated namely by Ben-Akiva & Bierlaire 2003) modelling approaches can arise from this notation; the Random-Coefficients Logit (RCL) model, and the Error-Components Logit (ECL) model.

In the RCL model, some entries of the vector β in equation (2) are specified to be random variables, capturing taste heterogeneity in the population. The choice probability of alternative i is then given by:

$$P(i \mid \mathcal{C}) = \int_{\beta} P(i \mid \mathcal{C}, \beta) f(\beta, \theta) \, \mathrm{d}\beta, \tag{9}$$

where $P(i \mid \mathcal{C}, \beta)$ is the MNL choice-probability of alternative i, conditional on β, and where θ is a vector of parameters of the distribution of the elements contained in the vector β, giving for example the mean and standard deviation across the population. Recent examples of this approach are given by Revelt & Train (1998), Bhat (2000), Hess & Polak (2004*a*,*b*) and Hess, Train & Polak (2004).

In the ECL model, the vector V in equation (1) is defined as

$$V = V(\beta, x) + \xi, \tag{10}$$

where $V(\beta, x) \in \mathbb{R}^J$ and ξ is a random vector of disturbances. In this case, the error term is composed of two parts, and the utility function is given by

$$U = V + \xi + \varepsilon, \tag{11}$$

where the vector ξ is generally assumed to follow a multivariate Normal distribution, with mean zero and covariance matrix Ω, where Ω is usually constrained to be diagonal (c.f. Walker 2001). By allowing some alternatives to share the same error-components, correlation between these alternatives is introduced into the unobserved part of utility. This approach can thus be used to relax the *IIA* property of the MNL model, and it has been shown (McFadden & Train 2000) that, with an appropriate specification of error-components, the ECL structure can theoretically approximate any random utility model (and thus also any GEV-style nesting structure) arbitrarily closely. Another major advantage of this model structure is that the error-components can be used to induce heteroscedasticity. For recent applications of the *ECL* formulation, see for example Bhat (1998) and Brownstone & Train (1999).

The two approaches (RCL and ECL) can be combined straightforwardly, allowing for the joint modelling of random taste heterogeneity and inter-alternative correlation. However, while the MMNL model is very flexible (and more so than the MNP model), important issues of identification need to be dealt with in the specification of the error-component structure (c.f. Walker 2001). Furthermore, although the MMNL model has the theoretical property of being able to approximate other random utility models arbitrarily closely, this may not always be as straightforward in practice (c.f. Garrow 2004). Finally, depending on the correlation structure, the high number of error-components required can lead to high simulation costs. Indeed, the integral in equation (8) does not generally have a closed form, and numerical techniques, typically simulation, are required during the estimation and application of MMNL models (and Mixed GEV models by extension). The development of ever more powerful computers and recent improvements in the efficiency of simulation techniques (c.f. Bhat 1999, Hess, Polak & Daly 2003, Hess, Train & Polak 2004) have significantly reduced the computational overheads of this process, and the number of applications using the RCL model especially has increased rapidly over recent years. Nevertheless, the computational cost of estimating and applying mixed GEV models remains high, when compared to their closed-form counterparts.

While integration over mixing distributions is necessary in the representation of random taste heterogeneity, this is not strictly the case for inter-alternative correlation. Indeed, just as, conditional on a given value of the taste-coefficients, a model allowing for random taste heterogeneity reduces to an MNL model, a model allowing for inter-alternative correlation in addition to random taste heterogeneity can in this case be seen to reduce to a given GEV model (assuming that an appropriate GEV model exists). As such, the correlation structure can be represented with the help of a GEV model, while the random taste heterogeneity is accommodated through integration over the assumed distribution of β. The use of a more complicated GEV model as the integrand leads to a more general type of a Mixed GEV model, of which the RCL model is simply the most basic form. Applications of this approach include for example Chernew et al. (2003) and Bhat & Guo (2004). In such a Mixed GEV model, the number of random terms, and hence the number of dimensions of integration (and thus simulation) is limited to the number of random taste coefficients, whereas, in the ECL model, one additional random term is in principle needed for the representation of each separate nest. It should be noted that the potential runtime-advantage resulting from this difference in dimensions of integration only manifests itself beyond a certain number of nests, as the more complicated form of the integrand in Mixed GEV models initially gives the ECL model a computational advantage. The use of Mixed GEV models does however have another advantage over the use of the ECL model in that it avoids the issues of identification that are specific to this latter model form.

Finally, it should be noted that while the error-components method has historically only been used with an MNL model as the basis, the approach can theoretically also be used when ε is GEV distributed, for example in the case where some correlation is to be captured by the GEV structure, with a remaining amount of correlation (or indeed heteroscedasticity) to be explained by the error-components. This can be useful in the case where existing GEV structures are incapable of capturing the full array of correlation in the data (GEV models are homoscedastic and do not allow to capture all types of correlation structure, c.f. Abbé 2003), while the exclusive reliance on error-components would lead to excessive computational cost or issues of identification. This approach would thus lead to an error-component GEV model. In this article, we concentrate on the use of the random-coefficients GEV model, the analysis of the potential of advanced error-components GEV models (not based on MNL) is an important area for further research.

3 Empirical analysis

3.1 Analytical framework

The data used for our empirical analysis form part of the survey data collected to estimate the hypothetical demand for a new high-speed transit system in Switzerland; the Swiss Metro (c.f. Abay 1999, Bierlaire et al. 2001). The aim is to build a mag-lev underground system operating at speeds up to 500 km/h in partial vacuum, connecting the major urban centres along Switzerland's *Mittelland* corridor; St. Gallen, Zurich, Bern, Lausanne and Geneva[1]. Aside from the problems of funding, technological feasibility and commercial viability, there is an important question about the impact that the development of such a system would have on the environment. Even though the construction of the Swiss Metro (SM) is thus rather unlikely in the near future, the data collected to estimate the demand for the system can give important insights into respondents' evaluation of hypothetical choice alternatives in general, and transport modes in particular. Furthermore, the

[1] For details see www.swissmetro.com

SM alternative can be seen as a proxy for a high-speed rail alternative; in the face of increasingly congested roads and skies, the analysis of the potential demand for such advanced public transport modes is a topic of great interest.

A combined Revealed/Stated Preference (RP/SP) approach was used to collect the data (c.f. Abay 1999). Initial interviews about a specific observed trip were followed by a set of SP experiments based on this specific trip, where both car-travellers and rail-travellers were used in the survey. The SP surveys comprised 9 hypothetical choice scenarios, using the three alternatives of car, rail and SM, where car was only available to car-owners. The main explanatory variables used to describe the alternatives were travel-time, cost/fare and headway (for train and SM alternatives). Two different seating arrangements were used for SM alternatives, corresponding to *1st class rail-travel*, and *business class aircraft seats*. Fares for SM services were obtained by multiplying rail fares by a factor of 1.2, while car running costs were set to 1.20 CHF/km.

The aim of the present article is to illustrate the potential of Mixed GEV models in practice, rather than making policy implications per se. As such, only the SP survey was used, whereas a more policy-oriented analysis would have had to make use of the combined RP/SP survey. Also, the potential scale differences in the error-term between car users and train users were not directly taken into account, where a treatment of these differences would again have been important in a more policy-oriented analysis. A separate analysis revealed some differences in scale between the two groups; allowing for these differences did however not significantly affect the conclusions with regards to the nesting structure or the presence of random taste heterogeneity. Software limitations meant that it was not possible to jointly accommodate scale differences and correlation across repeated choice observations; the latter phenomenon was in this case judged to be more important (c.f. section 3.5). Finally, it should be noted that the sample used in this analysis can be seen as being choice-based (given the selection of respondents on the basis of RP choices). As it was not possible to properly take into account the effects of sampling (the population weights were clearly only known for the two existing modes), the results of this analysis must be seen as applying to the present sample only.

Only commuters and business travellers were included in the analysis, and no distinction was made between these two groups of travellers at this stage, leading to a sample size of 6,870 observations. The main explanatory variables used in the model fitting exercise were cost, travel-time, and headway. Additionally, the impacts of seating arrangements for SM, age for rail travellers (divided into 5 roughly equally sized discrete groups), and season ticket ownership for rail-based alternatives were taken into account in the model. While separate travel-time coefficients coefficients were used for the three different modes, it was not possible to identify significantly different cost-coefficients for the three modes. Similarly, the differences between the estimated season ticket constants for rail and SM were not significant, such that a common coefficient was used. Attempts to account for possible further interactions between socio-demographic variables and taste coefficients were not successful. Additionally, some effort went into experimenting with non-linear specifications for the marginal utilities of the various explanatory variables; however, this did not lead to any significant gains in model performance. For reasons of identification, the ASC of rail was set to zero, in all model types used in the analysis. No significant random heterogeneity was identified in any of the models for either of the three ASCs, such that the three possible normalisation approaches are equivalent in the mixed models, just as they are in the closed-form models. Finally, aside from the season ticket variable, other inertia variables, such as car ownership, license holding and past choices can be expected to have a significant effect on choice behaviour; the analysis of these effects was however beyond the scope of the present analysis.

For the calibration of the various models discussed in this article, the estimation software BIOGEME (c.f. Bierlaire 2003) was used[2]. This estimation tool can be used for all types of closed-form as well as mixed GEV model structures. Furthermore, the program can accommodate non-linear utility functions, and the estimation can be performed so as to account for correlation across repeated choice observations for the same individual.

When estimating models based on mixing distributions, it is of interest to attempt to minimise the computational overhead of the calibration process (c.f. Hess, Train & Polak 2004). This is especially crucial in the case of Mixed GEV models that are based on a more complicated integrand than the simple MNL formula. One such improvement that can lead to important gains in simulation efficiency is the use of quasi-random number sequences instead of pseudo-random number sequences as the basis of the simulation process (c.f. Train 2003). In the present application, one such quasi-random approach, known as the Halton sequence (Halton 1960), was used in conjunction with an iterative drawing procedure. This procedure is

[2]The estimation software, together with examples, and documentation, is available from http://roso.epfl.ch/biogeme; the data and model files for the application presented in this article are available from http://roso.epfl.ch/mbi/biogeme/swissmetro

based on the notion that the first few iterations of the maximisation process are *rough* steps in the general direction of the maximum of the log-likelihood function, requiring a lower degree of precision in the simulation. As such, a comparatively lower number of draws can be used for these initial steps, leading to important reductions in computation time. To this extent, the model was first estimated to a preset convergence level using a very low number of draws. This number of draws was then increased, and the final estimates from the preceding run were used as starting values. This process was repeated until the preset number of $1,000$ draws (per dimension of integration, and per individual) was reached; a sensitivity analysis showed this to be sufficient to obtain stable estimates. At each step in this iterative process (increase in the number of draws), the sequences of Halton draws were newly generated, so as to obtain as uniform a spread as possible with the given number of draws used in a specific run. A trust-region algorithm was used in the estimation, and at each step, a more stringent convergence criterion was used. Overall, this approach is very similar to that proposed by Bastin (2004), except that in our approach, the change in the number of draws is controlled externally (and set prior to estimation), rather than being controlled internally. Furthermore, the approach of Bastin (2004) allows for occasional decreases in the number of draws during the estimation process.

3.2 Multinomial Logit model

As a basis for comparison, a simple MNL model was first fitted to the data; the estimation results for this model are reported in the first part of table 1. As expected, the results show negative marginal utilities for increases in travel-time on all three modes, with similar conclusions for cost and headway increases. The model further shows that older people are relatively more likely to choose rail, while season-ticket holders are more likely to choose rail and SM (when compared to car). Finally, the results show that in terms of the seating arrangements for SM, respondents have a preference for first-class rail seats as opposed to business class aircraft seats. In terms of the implied willingness to pay for travel-time reductions, the results show significantly higher values of travel-time savings (VTTS) for rail, while the value for SM is only marginally higher than that for car.

3.3 Nested Logit model

To account for the potential existence of heightened correlation between some of the alternatives, three separate NL structures were estimated on the data; grouping together car and rail, car and SM and rail and SM respectively. Only the nesting of car and rail, i.e. the grouping of existing modes versus the hypothetical SM alternative, resulted in a structural parameter that is greater than 1 (using the notation from section 2.1). The results of this estimation are shown in the second part of table 1[3]. With this model structure, the nesting parameter (μ_{CR}) takes a value of 2.23, implying a high correlation between the unobserved utilities of the car and rail alternatives of around 0.8. Aside from a difference in scale, the substantive results of the two models are very similar, although the VTTS measures are lower than in the corresponding MNL model, especially so for the car and rail alternatives. This also implies that the results show a clearer difference between the VTTS for car and SM. Finally, in terms of model fit, the results show a very significant increase in Log-Likelihood (LL) by 122.42 units, with one additional parameter. This leads to a likelihood-ratio test value of 244.84, which has an associated χ_1^2 p-value that is identical to zero (3^{-55}).

3.4 Cross-Nested Logit model

As reported in section 3.3, significant correlation could only be retrieved between the car and rail alternatives, leading to a nesting of existing versus hypothetical alternatives. It is however conceivable that such correlation also exists between the rail and the SM alternatives, given that they have the common aspect of being public transport modes. To test for the presence of such correlation, a CNL model was fitted to the data, allowing the rail alternative to belong to a rail-SM nest as well as to the car-rail nest. The results of this estimation process are reported in the third part of table 1 (CNL_A)[4]. The results show that in addition to high correlation between the unobserved utilities of the two existing modes of car and rail, there is also very high correlation between the unobserved parts of the utilities for rail and SM. The allocation parameters $\alpha_{R,CR}$ and $\alpha_{R,SR}$ show the degree of membership of the rail alternative to the nests it shares with car and SM respectively, where the estimates are very similar, with slightly higher allocation to the public transport nest.

[3] The t-test for μ_{CR} is expressed with regards to a base-value of 1.

[4] The t-tests for μ_{CR} and μ_{SR} are expressed with regards to a base-value of 1, while for $\alpha_{R,CR}$ and $\alpha_{R,SR}$, a base-value of 0.5 is used.

The CNL model reduces to the NL model described in section 3.3 when $alpha_{R,CR} = 1$ and $alpha_{R,SR} = 0$. In this scenario, the nesting parameter μ_{SR} becomes obsolete. When μ_{CR} further becomes equal to 1, the model reduces to the MNL model described in section 3.2. Likelihood-ratio tests can thus be used to compare the CNL model to the MNL and NL models, with 3, respectively 2 degrees of freedom (only one α is actually estimated, given that $\alpha_{R,CR} = 1 - \alpha_{R,SR}$). The resulting likelihood ratio test values are 332.42 and 87.58, both of which have p-values that are indistinguishable from zero, for χ^2_3 and χ^2_2 tests respectively. This shows that important gains in model fit can be obtained by accounting for the correlation between the two public transport alternatives; interestingly, this was not possible in the NL model, suggesting that this correlation can only be explained simultaneously with the correlation between the car and rail alternatives.

In terms of substantive results, the estimated coefficients are again all of the expected sign. However, the implied VTTS measures are significantly lower than those reported with the MNL and NL structures, where a similar observation can be made for the willingness to pay for headway reductions. This is the result of an increase in the relative weight of the marginal utility of cost when compared to the MNL and NL structures, and shows the impact of model structure on the relative scale of the various coefficients. This thus suggests that, by accounting for the correlation structure, the cost attribute gains in weight when compared to the other attributes.

The interpretation that should be given to the allocation parameters in a CNL model is not clear, although intuitive interpretations exist for example in the case of route-choice. In the present application, the allocation of the rail alternative was split almost evenly between the car-rail and the rail-SM nest. To establish the impact of these parameters, the model was re-estimated, with both allocation parameters constrained to a value of 0.5. The results of this process are reported in the fourth part of table 1 (CNL_B). The use of constrained allocation parameters leads to a slight drop in the estimated correlation in the two nests. Furthermore, it leads to a 4.6% increase in the estimated VTTS for the rail alternative. Aside from these two changes, the substantive impacts of the additional constraint are relatively minor. The constraint leads to a statistically significant drop in LL by 5.09 units, equating to a likelihood ratio test-value of 10.18, with an associated p-value of 0.0062.

To conclude this section, it should be noted that similar experiments where conducted with a structure allowing car to belong to a car-rail and a car-SM nest; no extra correlation between car and SM could however be identified.

3.5 Mixed Multinomial Logit model

As discussed in the introduction to this article, not allowing for potential random variations in tastes across respondents puts researchers at risk of producing seriously biased results. With this in mind, several experiments were conducted to explore the potential prevalence of random taste heterogeneity in the population of decision-makers. The repeated choice nature of the data was taken into account in these experiments, such that tastes vary across individuals, but not across observations for the same individual (c.f. Train 2003); this leads to efficient estimates, whereas the purely cross-sectional leads only to consistent estimates. Attempts were also made to accommodate other SP panel effects, such as inertia, but none of these was found to have a significant effect.

Significant random taste heterogeneity was identified for five coefficients; the three travel-time coefficients, in addition to the dummy coefficients for age for rail users, and for seating type for SM users. For reasons of simplicity, a Normal distribution was used for all five coefficients. This is a valid assumption for the two dummy coefficients, but can lead to problems with the three travel-time coefficients. Indeed, by using a Normal distribution, researchers in effect make an a priori assumption that the coefficient takes a positive value for some of the respondents. The use of bounded distributions is in this case preferable (c.f. Train & Sonnier 2004, Hess, Bierlaire & Polak 2004). However, in the present application, the Normal distribution led to very good performance, while problems in estimation were encountered when using alternative distributions. Furthermore, with the estimated distributional parameters, the probability of a wrongly specified coefficient was always at an acceptable level.

The results of the estimation are summarised in the first part of table 2 (MMNL_A). The first observation that can be made is that the MMNL model leads to an improvement in LL over the MNL model by 229.63 units, with 5 additional parameters. This equates to a likelihood-ratio test-value of 459.26, giving a χ^2_5 p-value of 0. This illustrates the important gains in model fit that result from accommodating random variations in respondents' tastes. The results further show that the effect of using a Normal distribution for the three travel-time coefficients is benign, with probabilities of a *wrongly* signed coefficient of 1%, 0% and 2% for car, rail and SM respectively. Finally, it should be noted that, with the MMNL model, the estimates of the two ASCs, as well as that of the mean for the age-dummy for rail-travellers, are not significant at the usual 95% level of confidence.

In terms of actual estimation results, the model shows that, while age still has a positive mean effect on the utility of the rail alternative, for about 30% of respondents, this effect is now negative. Tests with bounded distributions led to poor results, suggesting that these results do indeed signal the existence of travellers for which this dummy variable is negative, rather than being simply an effect of using the Normal distribution. A similar observation can be made for the coefficient associated with the type of seating, where the results now indicate that almost 42% of travellers have a preference for aircraft-type business-class seats over first-class rail-seats. These results illustrate the potential of the MMNL model; the closed-form models falsely suggest a consistent positive effect of age and rail-type seats across the population.

In terms of the implied willingness to pay for travel-time reductions, the results show consistently higher VTTS measures for all three modes than was the case in the closed-form models. This shows the important bias that can result from not accounting for random variations in the coefficients involved in trade-off calculations. Although it was not possible to estimate such a coefficient in the present analysis, it should be stressed that the risk of bias becomes even greater in the case of a randomly distributed cost-coefficient. Again, like in the MNL model, the estimated VTTS for car and SM are very similar, while the corresponding measure for rail is significantly higher. It is important to note that the use of fixed coefficients not only leads to a risk of biased results, but also leads to a loss of all information about the variation in the VTTS across respondents. The standard deviations reported in table 2 for the travel-time coefficients are very high, and lead to very wide confidence intervals for the VTTS. As an illustration, the lower and upper 80% quantiles were calculated, leading to lower limits of 42.13, 82.83, and 36.08 CHF/hour for car, rail and SM respectively, with corresponding upper limits of 145.12, 144.32 and 156.10 CHF/hour respectively. This shows that while rail has got the highest associated VTTS, it has the narrowest confidence interval, followed by car and SM. The variation in the VTTS for SM is so important that, while the mean VTTS for SM lies in between those for car and rail, its lower and upper limits are more extreme than those of car and rail respectively. This could be seen as a reflection of the uncertainty involved with the evaluation of a hypothetical mode.

3.6 Mixed Nested Logit model

The results in sections 3.3, 3.4 and 3.5 have shown that important gains in model performance can be obtained both by accounting for the presence of inter-alternative correlation in the unobserved utility terms, and by allowing for a random distribution of tastes across decision-makers. However, as highlighted in the introduction and theoretical part of this article, it is not clear a priori whether these results actually signal the presence of separate phenomena, or whether the two approaches simply explain the same phenomenon in different ways. The aim was now to attempt to jointly model the two phenomena, hence reducing the risk of confounding. For this, a Mixed NL model was fitted to the data.

Whereas, with the MMNL model described in section 3.5, it was possible to retrieve significant random variation for five taste coefficients, this number was reduced to four in the Mixed NL model. Indeed, the standard deviation associated with the marginal utility of travel time for rail alternatives was no longer statistically significant at any reasonable level of significance. This was already the coefficient with the smallest variation in the MMNL model (c.f. table 2), and by accounting for inter-alternative correlation, the error-term in the model decreases, reducing the scope for retrieving random taste heterogeneity further. This signals possible confounding in the simple MMNL model presented in section 3.5. The final estimates for the Mixed NL model are reported in the second part of table 2[5]. The results show that, compared to the NL model reported in table 1, the use of the Mixed NL model leads to a gain in LL by 179.59 units, with 4 additional parameters. This equates to a likelihood ratio test of 359.18, with an associated χ^2_4 p-value of 0. Similarly, the Mixed NL model leads to an improvement in LL by 72.38 units over the MMNL model from section 3.5. To allow for the use of a nested log-likelihood ratio comparison between the Mixed NL and MMNL structures, the MMNL model from section 3.5 had to be re-estimated with a fixed coefficient for the the rail travel-time. The results of this re-estimation are reported in the third part of table 2, showing that, as expected, the use of a fixed coefficient leads to a significant drop in LL by 7.16 units. The use of the Mixed NL model leads to a highly significant improvement in LL by 79.54 units when compared to this re-estimated MMNL model, with a single additional parameter. These results reflect the importance of jointly accommodating the two phenomena of correlation and random taste heterogeneity.

In terms of actual estimation results, the values in table 2 show that the Mixed NL model retrieves a correlation structure between car and rail alternatives that is virtually indistinguishable from that obtained when using the simple NL model reported in table 1. However, the significance level of the nesting parameter is markedly lower. A similar observation can be made

[5] Again, the t-test for μ_{CR} is expressed with regards to a base-value of 1.

for the standard deviations of the randomly distributed coefficients (when compared to the two MMNL models). This drop in significance levels is to be expected, given that the Mixed NL model decomposes the error-term further than the NL and MMNL models. It can also be noted that in the Mixed NL model, the mean VTTS measure for rail and SM are now indistinguishable, whereas, in the NL and MMNL models, the VTTS for rail was markedly higher. This could be seen as an effect of using a fixed travel-time coefficient for rail, when compared to the MMNL model; however, the re-estimated MMNL model uses the same restriction, yet still yields a slightly higher VTTS for rail than for SM. Any other remaining differences between the two models are largely down to a difference in scale.

3.7 Mixed Cross-Nested Logit model

The final model fitted during the analysis was a Mixed CNL model, using the same nesting structure as the CNL model described in section 3.4. In section 3.6, we observed that, by accounting for the correlation between the car and rail alternatives, the scope for retrieving significant amounts of random taste heterogeneity is reduced. When fitting the Mixed CNL model, serious estimation problems were encountered. These related specifically to the ability to retrieve random taste heterogeneity, especially when also accounting for the repeated choice nature of the dataset. These problems reflect the complexity of the model, but could also be a sign of a lack of explanatory power in the data, in such that the error-term cannot be partitioned enough to reproduce a Mixed CNL structure with a high number of random taste coefficients. Eventually, it was possible to estimate a Mixed CNL model with a single randomly distributed taste coefficient, namely the marginal utility of travel-time for the car-alternative. For estimation purposes, the allocation parameters were both constrained to be equal to 0.5. The results of this estimation process are reproduced in the fourth part of table 2[6].

The first observation that can be made from table 2 is that, with one additional parameter, the Mixed CNL model leads to a very significant improvement over the constrained CNL model reported in the second part of table 1; the difference in LL is 80.96, leading to a likelihood-ratio test-value of 161.92, which has an associated χ^2_1 value of zero. This shows that even a single randomly distributed taste coefficient leads to important gains in explanatory power. The Mixed CNL model also has a higher LL than the two

[6] The t-tests for μ_{CR} and μ_{SR} are again expressed with regards to a base-value of 1.

MMNL models, although no nested likelihood-ratio test can be performed for these differences. On the other hand, the LL of the Mixed CNL model is inferior to that of the Mixed NL model by 59.93 units. This suggests that the increased partitioning of the error term resulting from allowing for cross-nesting adds less explanatory power than the partitioning resulting from accounting for the additional levels of random taste heterogeneity in the Mixed NL model. Efforts to explain a larger part of the error term by accounting for further levels of taste heterogeneity in a Mixed CNL framework are ongoing.

The other main observation that can be made from table 2 is that, while the VTTS measures produced by the Mixed CNL model are closer in scale to those produced by the closed-form models than those produced by the other mixed models, the VTTS of SM is now lower than the mean VTTS of the car alternative. This can however be seen as an effect of using a randomly distributed coefficient for the marginal utility of travel-time for car, while a fixed coefficient is used for the other two modes. Finally, it should be noted that, while the estimated value for μ_{SR} is very similar to that obtained with the constrained CNL model in section 3.4, the value estimated for μ_{CR} is markedly higher. This shows that the estimation of the structural parameters is affected by the use of a utility function containing random coefficients. This in turn again suggests some interaction between the part of the error-term linked to random taste heterogeneity and the part linked to inter-alternative correlation.

4 Summary and Conclusions

In this article, we have discussed the issues arising with model specification in the case of a non-trivial error-structure. We have focussed on two separate ways of partitioning the error-term; accounting for (arbitrary) correlation between alternatives in the unobserved utility components, and allowing for a random distribution of tastes across decision-makers. The theoretical discussions presented in this article have highlighted the fact that the distinction between these two phenomena is not clear-cut, and that there exists a significant risk of confounding in the case where researchers account for only one of the two phenomena.

Our empirical analysis has shown that while it is possible to separately model the prevalence of correlation in the choice-set and random taste heterogeneity in the population of decision-makers, and while both approaches lead to very significant gains in model fit, the joint modelling of these two phenomena can be more problematic. Indeed, while the Mixed NL model described in section 3.6 retrieves a near identical nesting structure to that obtained with the simple NL model in section 3.3, random taste heterogeneity can only be retrieved for four taste coefficients, as opposed to five in the simple MMNL model (c.f. section 3.5). Even more severe problems were encountered when using a Mixed CNL model, where random taste heterogeneity could only be retrieved for a single coefficient. Although, in the Mixed CNL model, these problems were at least party due to model complexity, the overall results do highlight the issue of confounding of taste heterogeneity and correlation, complementing similar observations made by Cherchi & Ortuzar (2004) with regards to the ECL model. It should also be noted that the various model fitting exercises described in this article have highlighted the fact that the assumptions made with regards to the error-structure can have significant impacts on substantive results, such as willingness-to-pay indicators.

It should be stressed that the failure to simultaneously account for all heterogeneity and correlation should not be seen as a deficiency of the model, but rather as a sign that the error-term in the model has decreased. Indeed, by accounting for either of the two phenomena, the modeller explains processes that take place in the unobserved part of utility of the alternatives. This is analogous to the case where the specification of the utility function in the most basic of discrete choice models is improved by the inclusion of more explanatory variables. If it were possible to improve the utility specification to the point were all correlation across alternatives is explained in the observed part of utility, the errors would become independent, and it would no longer be possible to explain inter-alternative correlation with the help of a nesting structure. As such, it can often be observed that, while inter-alternative correlation can be retrieved in models using a very basic specification of the observed utility, further refinement of the utility function will lead to problems with retrieving significant nesting effects. This should clearly be seen as desirable, as any correlation is now explained in a deterministic way, through the observed utility function. A similar process occurs in models jointly allowing for random taste heterogeneity and correlation. When only allowing for either of the two phenomena in a model, the impact of the unrepresented phenomenon will at least be partly carried over into the other phenomenon. This in turn shows that, by simultaneously accounting for the two phenomena, the scope for retrieving apparent significant effects of either of the two phenomena is reduced. On the other hand, this however also means that the risk of falsely explaining correlation by random taste heterogeneity, or vice-versa, is reduced. As such, researchers should always strive to simultaneously account for the potential prevalence of both random taste heterogeneity and unexplainable inter-alternative correlation, in the case where the observed utility function is incapable of explaining sufficient amounts of choice behaviour for the remaining error-term to be distributed purely type I extreme-value.

Acknowledgements

The authors would like to express their thanks to Georg Abay and Kay Axhausen for making the Swiss Metro data available for the present analysis. The authors would also like to thank Elisabetta Cherchi, Andrew Daly, Laurie Garrow and Kenneth Train for helpful comments on an earlier version of this article.

References

Abay, G. (1999), *Nachfrageabschaetzung Swissmetro: Eine stated-preference Analyse (in German)*, Berichte des Nationalen Forschungsprogrammes 41: Verkehr und Umwelt, Bericht F1, EDMZ, Bern.

Abbé, E. (2003), *Behavioral modeling: Generalized extreme value for choice models*, MSc Thesis, Institute of Mathematics, Ecole Polytechnique Fédérale de Lausanne.

Bastin, F. (2004), *Trust-Region Algorithms for Nonlinear Stochastic Programming and Mixed Logit Models*, Ph.D. Thesis, University of Namur, Namur, Belgium.

Ben-Akiva, M. & Bierlaire, M. (2003), Discrete choice models with applications to departure time and route choice, *in* R. Hall, ed., 'Handbook of Transportation Science, Second Edition', Kluwer Academic Publishers, Boston, MA., pp. 7–37.

Ben-Akiva, M. & Francois, B. (1983), *Mu-homogenous generalized extreme value model*, Working Paper, Department of Civil Engineering, MIT, Cambridge, MA.

Bhat, C. R. (1998), 'Accommodating flexible substitution patterns in multi-dimensional choice modeling: formulation and application to travel mode and departure time choice', *Transportation Research* **32B**(7), 425–440.

Bhat, C. R. (1999), 'Quasi-random maximum simulated likelihood estimation of the mixed multinomial Logit model', *Transportation Research* **35B**(7), 677–693.

Bhat, C. R. (2000), 'Incorporating observed and unobserved heterogeneity in urban work travel mode choice modeling', *Transportation Science* **34**(2), 228–238.

Bhat, C. R. & Guo, J. (2004), 'A Mixed Spatially Correlated Logit Model: Formulation and Application to Residential Choice Modeling', *Transportation Research* **38B**(2), 147–168.

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

Bierlaire, M. (2004), *A theoretical analysis of the cross-nested Logit model*, Annals of Operations Research, forthcoming.

Bierlaire, M., Axhausen, K. W. & Abay, G. (2001), *The acceptance of modal innovation : The case of Swissmetro*, Proceedings of the 1^{st} Swiss Transport Research Conference, Monte Verita, Ascona, Switzerland.

Brownstone, D. & Train, K. (1999), 'Forecasting new product penetration with flexible substitution patterns', *Journal of Econometrics* **89**, 109–129.

Cherchi, E. & Ortuzar, J. de D. (2004), *Random effect vs. error component in the Mixed Logit model: correlation, heteroscedasticity, preference and response unobserved heterogeneity*, paper presented at the 10^{th} triennial World Conference on Transport Research, Istanbul, Turkey.

Chernew, M., Gowrisankaran, G. & Scanlon, D. (2003), *Learning and the Value of Information: The Case of Health Plan Report Cards*, Submitted to International Economic Review.

Daganzo, C. F. (1979), *Multinomial Probit: The Theory and its Applications to Demand Forecasting*, Academic Press, New York.

Daly, A. & Bierlaire, M. (2003), *A general and operational representation of GEV models*, Technical Report RO-030502, Institute of Mathematics, Operations Research Group ROSO, EPFL, Lausanne, Switzerland.

Daly, A. & Zachary, S. (1979), Improved multiple choice models, *in* D. Hensher & Q. Dalvi, eds, 'Identifying and Measuring the Determinants of Mode Choice', Teakfields, London.

Garrow, L. (2004), *Comparison of models representing correlation and random taste variation: An application to airline passengers' rescheduling behavior*, Published doctoral dissertation, Northwestern University, Evanston, Illinois.

Halton, J. (1960), 'On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals', *Numerische Mathematik* **2**, 84–90.

Hensher, D. & Greene, W. H. (2003), 'The Mixed Logit Model: The State of Practice', *Transportation* **30**(2), 133–176.

Hess, S., Bierlaire, M. & Polak, J. W. (2004), *Estimation of value of travel-time savings using Mixed Logit models*, Transportation Research A, forthcoming.

Hess, S. & Polak, J. W. (2004*a*), *Mixed Logit estimation of parking type choice*, paper presented at the 83^{rd} Annual Meeting of the Transportation Research Board, Washington, DC.

Hess, S. & Polak, J. W. (2004*b*), *Mixed Logit modelling of airport choice in multi-airport regions*, Journal of Air Transport Management, forthcoming.

Hess, S., Polak, J. W. & Daly, A. (2003), *On the performance of shuffled Halton sequences in the estimation of discrete choice models*, paper presented at the European Transport Conference, Strasbourg, France.

Hess, S., Train, K. & Polak, J. W. (2004), *On the use of a Modified Latin Hypercube Sampling (MLHS) Method in the Estimation of a Mixed Logit Model for Vehicle Choice*, Transportation Research B, forthcoming.

Koppelman, F. S. & Wen, C.-H. (2000), 'The Paired Combinatorial Logit model: properties, estimation and application', *Transportation Research* **34B**, 75–89.

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, ed., 'Spatial Interaction Theory and Planning Models', North Holland, Amsterdam, chapter 25, pp. 75–96.

McFadden, D. & Train, K. (2000), 'Mixed MNL Models for discrete response', *Journal of Applied Econometrics* **15**, 447–470.

Papola, A. (2004), 'Some developments on the cross-nested Logit model', *Transportation Research* **38B**(9), 833–854.

Revelt, D. & Train, K. (1998), 'Mixed Logit with repeated choices: households' choices of appliance efficiency level', *Review of Economics and Statistics* **80**(4), 647–657.

Small, K. (1987), 'A discrete choice model for ordered alternatives', *Econometrica* **55**(2), 409–424.

Train, K. (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge, MA.

Train, K. & Sonnier, G. (2004), Mixed logit with bounded distributions of correlated partworths, *in* A. Alberini & R. Scarpa, eds, 'Applications of simulation methods in environmental and resource economics', Kluwer Academic Publishers, Boston, MA.

Vovsha, P. (1997), 'Application of a Cross-Nested Logit model to mode choice in Tel Aviv, Israel, Metropolitan Area', *Transportation Research Record* **1607**, 6–15.

Vovsha, P. & Bekhor, S. (1998), 'The Link-Nested Logit model of route choice: overcoming the route overlapping problem', *Transportation Research Record* **1645**, 133–142.

Walker, J. L. (2001), Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables, PhD thesis, MIT, Cambridge, MA.

Wen, C.-H. & Koppelman, F. S. (2001), 'The Generalized Nested Logit Model', *Transportation Research* **35B**(7), 627–641.

Williams, H. C. W. L. (1977), 'On the Formulation of Travel Demand Models and Economic Evaluation Measures of User Benefit', *Environment and Planning* **9A**(3), 285–344.

	MNL		NL		CNL$_A$		CNL$_B$	
Parameter	Estimate	T-statistic	Estimate	T-statistic	Estimate	T-statistic	Estimate	T-statistic
ASC Car	0.5731	2.88	0.3013	2.56	-0.3472	-3.93	-0.2832	-2.71
ASC SM	0.8661	4.89	0.5965	4.31	0.0071	0.09	0.047	0.42
Age for Rail	0.2639	6.15	0.178	6.3	0.1538	9.17	0.1652	6.71
Airline seats for SM	-0.435	-4.35	-0.2743	-3.05	-0.1725	-3.65	-0.2309	-3.61
Season ticket for Rail/SM	0.9718	5.31	0.7595	7.17	0.6893	7.75	0.6715	6.92
Headway (minutes)	-0.0055	-5.55	-0.0037	-5.5	-0.0028	-5.94	-0.0031	-5.74
Cost (CHF)	-0.0098	-13.98	-0.0071	-11.75	-0.0065	-14.91	-0.007	-13.53
Travel time Car (minutes)	-0.0116	-10.66	-0.0077	-6.67	-0.0062	-10.98	-0.0067	-5.64
Travel time Rail (minutes)	-0.0152	-14.03	-0.0106	-10.13	-0.0079	-12.56	-0.0089	-8.7
Travel time SM (minutes)	-0.0119	-6.55	-0.0085	-5.06	-0.0067	-9.62	-0.0073	-5
μ_{CR}	-		2.23	7.76	2.78	9.27	2.67	7.48
μ_{SR}	-		-		5.91	6.61	4.68	6.9
$\alpha_{R,CR}$	-		-		0.4293	3.76	0.5	-
$\alpha_{R,SR}$	-		-		0.5707	3.76	0.5	-
VTTS Car (CHF/hour)	71.02		65.07		57.23		57.43	
VTTS Rail (CHF/hour)	93.06		89.58		72.92		76.29	
VTTS SM (CHF/hour)	72.86		71.83		61.85		62.57	
Final LL	-5328.24		-5205.82		-5162.03		-5167.12	
$\rho^2(0)$	0.2471		0.2644		0.2706		0.2698	

Table 1: Estimates for MNL, NL and CNL models

	$MMNL_A$		Mixed NL		$MMNL_B$		Mixed CNL	
Parameter	Estimate	T-statistic	Estimate	T-statistic	Estimate	T-statistic	Estimate	T-statistic
ASC Car	0.2758	0.92	0.1807	0.89	0.3268	1.11	-0.2222	-1.71
ASC SM	0.4455	1.63	0.3765	1.81	0.49	1.81	-0.0196	-0.16
Age for Rail (μ)	0.1382	1.4	0.0869	1.29	0.0496	0.53	0.1612	5.94
Age for Rail (σ)	0.2579	6.25	0.2553	7.43	0.3064	8.37	-	-
Airline seats for SM (μ)	-0.3142	-2.09	-0.3376	-2.82	-0.3374	-2.17	-0.2247	-3.54
Airline seats for SM (σ)	1.5553	6.48	0.7717	2.23	1.6191	6.94	-	-
Season ticket for Rail/SM	1.2126	3.66	0.8555	4.42	1.3097	3.88	0.6099	4.56
Headway (minutes)	-0.0072	-5.48	-0.0052	-5.3	-0.0071	-5.43	-0.0031	-5.94
Cost (CHF)	-0.0138	-11.02	-0.0103	-9.07	-0.0139	-11.06	-0.0076	-11.05
Travel time Car (minutes) (μ)	-0.0215	-14.99	-0.016	-11.1	-0.0216	-14.5	-0.0089	-8.77
Travel time Car (minutes) (σ)	0.0092	11.25	0.0074	4.1	0.0094	7.36	0.0039	8.10
Travel time Rail (minutes) (μ)	-0.0261	-15.13	-0.0175	-12.18	-0.0229	-15.09	-0.0100	-10.28
Travel time Rail (minutes) (σ)	0.0055	6.65	-	-	-	-	-	-
Travel time SM (minutes) (μ)	-0.0221	-13.41	-0.0175	-11.89	-0.0221	-13.43	-0.0086	-6.32
Travel time SM (minutes) (σ)	0.0108	8.77	0.0104	8.07	0.012	10.37	-	-
μ_{CR}	-	-	2.28	5.08	-	-	4.10	4.27
μ_{SR}	-	-	-	-	-	-	4.26	6.03
VTTS (CHF/hour)	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Car	93.48	40	93.2	43.11	93.24	40.58	70.25	30.45
Rail	113.48	23.91	101.94	-	98.85	-	78.96	-
SM	96.09	46.96	101.94	60.58	95.4	51.8	67.45	-
Final LL	-5098.61		-5026.23		-5105.77		-5086.16	
$\rho^2(0)$	0.2795		0.2898		0.2785		0.2813	

Table 2: Estimates for MMNL, Mixed NL and Mixed CNL models