Report

Distributional assumptions in mixed logit models

Author(s): Axhausen, Kay W.; Hess, Stephane; Polak, John W.

Publication Date: 2005

Permanent Link: https://doi.org/10.3929/ethz-a-006020898

Rights / License: In Copyright - Non-Commercial Use Permitted
Distributional assumptions in Mixed Logit models

Stephane Hess∗ Kay W. Axhausen† John W. Polak‡

August 1, 2005

Abstract

The characterisation of taste heterogeneity in discrete choice models has been an area of intense research activity in recent years. Although the underlying theory imposes few constraints on the structure of this heterogeneity, in practice most empirical work has adopted an approach based on the use of a relatively limited range of standard probability distributions such as the Normal and Lognormal. However, a number of recent studies have raised doubts about the adequacy of these commonly used distributional assumptions. Against this backdrop, the aim of this paper is to explore the potential of a much wider range of distributional assumptions, including a number offering substantially greater flexibility than the most commonly used forms. Using data from a recent stated preference study undertaken to estimate travellers' valuation of travel time savings, we compare eleven different distributions for characterising travellers tastes with respect to changes in travel time and travel cost. The results demonstrate that the choice of distributional assumption can have a significant impact on estimation results, particularly and predictably, in the inferences that can potentially be drawn regarding extreme values. Overall, our results suggest that for attributes such as travel time and travel cost there are potentially significant advantages in using bounded distributions (such as the Johnson $S_B$) and in estimating these bounds simultaneously with other model parameters.

Word Count: 6,357 words, 2 tables and 5 figures

∗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 for Transport Planning and Systems, ETH Zürich, axhausen@ivt.baug.ethz.ch, Tel: +41(0)1 633 3943, Fax: +41(0)1 633 1057
‡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 and context

Over recent years, researchers and practitioners have increasingly begun using the Mixed Multinomial Logit (MMNL) model for a representation of random variations in tastes across respondents. This not only offers great benefits in terms of providing insights into variations in tastes across respondents, but also potentially avoids bias in the estimated trade-offs in models based on the use of fixed taste coefficients (see for example Algers et al. 1, Hess and Polak 20).

However, while the MMNL model can offer great gains in flexibility, the pitfalls are similarly significant, as stressed for example by Hensher and Greene [17] and Hess, Bierlaire and Polak [19]. Aside from the greater cost in terms of model estimation and application, as discussed for example by Hess, Train and Polak [21], two main issues arise with the use of the MMNL model; the choice of statistical distribution for randomly distributed coefficients, and the economic interpretation of such coefficients. The additional issue of deciding which parameters should be modelled as being randomly distributed across agents can be addressed relatively straightforwardly on the basis of statistical tests [c.f. 7].

This paper looks specifically at the issues of the choice of distribution and the interpretation of the resulting parameter values, and discusses how the two are strongly inter-related. The discussion centres on one specific application of MMNL models, namely the estimation of variations in the valuation of travel time savings (VTTS). Here, it should be noted that the issues described in this paper apply to mixture models in general, and are not constrained to MMNL. For reasons of simplicity, the discussion here centres on the MMNL model, which is the only widely used mixture model. For a discussion of more advanced mixture structures, see for example Hess, Bierlaire and Polak [18].

While the issue of the choice of distribution has been discussed repeatedly in the existing literature, with some examples being the work of Hensher and Greene [17], Sørensen [28] and Train and Sonnier [29], it should be noted that the number of distributions tested in previous work has generally been rather limited. Additionally, the vast majority of actual modelling analyses still rely exclusively on the use of the Normal distribution, while only a handful of alternatives to the Normal distribution have received even modest exposure. From this point of view, there is an urgent need for extensive research into the applicability of hitherto unused (or rarely used) distributions in the context of mixture models. As such, the main aim and contribution of this paper is to conduct a case-study using a high number of
different continuous distributions in a single application, hence highlighting the potential differences in performance as well as substantive results. The work presented in this paper is based entirely on standard continuous distribution functions, and ignores the potential use of empirical distributions. Additionally, any estimation work is based purely on standard techniques, with options such as the conditioning on respondents’ choices [c.f. 27] not explored here.

The remainder of this paper is organised as follows. The next section discusses the state-of-practice in terms of the choice of distribution in MMNL modelling. The importance of the distributional assumptions and the issues faced in the interpretation of MMNL results are illustrated in the VTTS case-study presented in Section 3. Finally, Section 4 summarises the findings of the paper, and offers some guidance for good practice.

2 Choice of distribution: the state of practice

As indicated in Section 1, the choice of distribution for a randomly distributed coefficient plays a crucial role in the specification of a MMNL model. In practice, only the Normal (Gaussian) and Lognormal distributions have found widespread application in MMNL modelling. Several authors have also advocated the use of the Triangular distribution [e.g. 17], while recently, good results have also been obtained with Johnson’s SB distribution [c.f. 29].

While the use of the Normal distribution can be appropriate in the case of coefficients without a strict sign assumption, the fact that it is unbounded can cause severe problems with interpretation when used for coefficients where such an a priori assumption exists in principle (e.g. travel time coefficients). Indeed, as discussed by Hess, Bierlaire and Polak [19], it can be argued that specifying a given coefficient to follow a Normal distribution is equivalent to making an a priori assumption that both positive and negative values for this coefficient exist in the population. In the case of a normally distributed travel-time coefficient, it can thus be seen that a positive probability of a non-negative coefficient is postulated by the researcher. Additionally, problems can arise in the case of asymmetrical true distributions. While the use of the Lognormal distribution can at least partly address these issues, problems can be caused by the long tails [c.f. 20]. Additionally, computational problems and slow convergence limit the applicability of the Lognormal distribution. The issues of the long tail can be avoided with the use of the Triangular distribution, which has the ad-
vantage of being bounded to either side. However, in its standard form, the
distribution faces the same problem as the Normal in terms of a symmetrical
shape, where the asymmetrical variant is difficult to implement, given the
complications with estimating the location of the peak.

The main factor leading to the almost exclusive reliance in MMNL mod-
eling on the above listed distributions is the relatively limited repertoire of
distributions supported by existing MMNL estimation packages, where only
a small subset of modellers make use of their own, purpose-written code.
Within the set of available distributions, the choice is often influenced by
issues of numerical problems when using the advanced distributions, but it
must be said that the limited awareness of the potential benefits of using a
broader range of distributions also plays a role in the preeminent position
of the Normal distribution.

3 VTTS case-study

We now turn our attention to the VTTS case-study, which serves as an
indication of the effects of distributional assumptions in the use of MMNL
models.

3.1 Introduction

The computation of VTTS measures has been one of the main applications
of random utility models, with some recent discussions of the topic including
Algers et al. [1], Hensher [14, 15, 16], Lapparent and de Palma [24], Cirillo
and Axhausen [8] and Sillano and Ortúzar [27]. The VTTS is an important
willingness-to-pay indicator, used for example for cost-benefit analysis in the
context of planning new transport systems, or for pricing. In discrete choice
models, the computation of VTTS measures is relatively straightforward,
given by the ratio of the partial derivatives of the utility function with
respect to travel-time and travel-cost (i.e. the marginal rate of substitution
between travel-time and travel-cost, at constant utility). Although this is
an intuitively plausible approach, it is important to appreciate that the
justification for this approach to the valuation of travel-time savings rests
not on plausibility but rather on a substantial body of microeconomic theory
that addresses the issue of how individuals allocate time amongst alternative
activities, including travel. Indeed, the topic of time allocation and valuation
has been the subject of intense study from a variety of different perspectives
for several decades (see, among others, Becker 4, Oort 26, De Serpa 9, Evans
11, Truong and Hensher 30, Bates 3 and Jara-Diaz and Guevara 23). The
papers by Jara-Diaz [22] and Mackie et al. [25] provide excellent overviews of the development of this literature.

Under the (strong but necessary) assumption that all effects of these two attributes are captured in the observed part of utility ($V$), the VTTS measure is simply computed as:

$$\frac{\partial V / \partial T_T}{\partial V / \partial T_C},$$

with $TT$ and $TC$ representing the travel-time and travel-cost attributes respectively. In the case of fixed taste coefficients, and with the commonly used linear-in-variables utility function, this formula reduces to $\beta_{TT}/\beta_{TC}$, where $\beta_{TT}$ and $\beta_{TC}$ are the time and cost coefficients, giving the marginal utilities of increases by one unit in travel-time and travel-cost respectively. Estimates of these marginal utilities are produced by calibrating the model on the choice data used in the estimation. Even with the use of non-linear transforms, such as the natural logarithm, the computation remains relatively straightforward, although the actual values of the attributes now enter into the computation of the trade-offs. Here, it should also be noted that the coefficient values themselves are estimators which are asymptotically normally distributed. As such, the ratio of $\beta_{TT}$ and $\beta_{TC}$ is itself a random variable, as discussed by Armstrong et al. [2].

With the increased use of the MMNL model in the area of transportation, researchers have begun to increasingly exploit the power of this model to represent a random variation in the marginal utility of travel-time and travel-cost across respondents [e.g. 1, 8]. However, the extension of the theoretical foundations of the calculation of VTTS to the case where $\beta_{TT}$ and/or $\beta_{TC}$ are modelled as random variables is not straightforward. Indeed, in the MMNL analysis of the distribution of VTTS measures across a population of respondents, the distributional assumptions play a crucial role, and have a significant effect on model interpretation. In this discussion, we focus on the marginal utility of travel-time, but a similar principle applies in the case of the marginal utility of access-cost, or indeed in the case where the VTTS is modelled directly, as opposed to being based on the ratio of $\beta_{TT}$ and $\beta_{TC}$ [c.f. 13].

In models that are based on the use of fixed taste coefficients, researchers generally have an a priori expectation of obtaining a negative travel-time coefficient, and models producing positive values will normally be rejected on the grounds of model misspecification (or lack of explanatory power in the data). While the sign-issue is thus relatively straightforward in the case of fixed-coefficient models, it becomes more complicated in the case of
models allowing for random taste heterogeneity. Indeed, in such models, the use of an unbounded distribution can lead to a non-zero probability of positive as well as negative travel-time parameters. In this case however, it is not clear a priori whether such estimates do in fact indicate the presence of respondents with negative VTTS in the population, or whether they are simply an artifact of the model specification or the poor quality of the data used in model estimation.

As discussed by Hess, Bierlaire and Polak [19], one potential source of model misspecification can come in the form of an inappropriate choice of mixing distribution for the travel-time coefficient. Like for most other coefficients, the most common choice of distribution for the travel-time coefficient is the Normal distribution. Here, the unbounded nature of the Normal distribution can lead to major complications. Indeed, in the case where the true distribution yields strictly negative values, but has a mean close to zero with a long tail into the negative space of numbers, the symmetrical nature of the Normal distribution can, in approximation, lead to a significant share of positive values, even though such values are not actually revealed by the data. On the other hand, in the case where the source for such estimates are some problems with the data, the Normal distribution has the potential to produce such values, hence allowing the modeller to identify the problem and motivate an investigation into its causes. Although, without an in-depth investigation, it is desirable not to explain a significant probability of a positive travel-time coefficient by the notion that some agents have negative VTTS, it is similarly bad practice to simply constrain the model to purely negative values for $\beta_{TT}$, hence ignoring the impact of data or model imperfections.

The issue with the Normal distribution is thus the problem of deciding whether a non-zero probability of a positive coefficient is revealed by the data or is simply an artifact of the symmetrical nature of the distribution. Here, it can be seen that the Triangular distribution, in its symmetrical form, leads to similar problems, although by being bounded, it at least avoids the long tails. As such, the aim should be to use distributions that can signal the presence of such effects with a minimal risk of the effects actually being caused by the distribution itself. Here, the above arguments in relation to the Normal and the Triangular suggest that such a distribution should not make too strict an a priori shape assumption. Additionally, it should be clear that distributions with estimated bounds have an advantage in terms of not making an a priori assumption about the range of the distribution. Here, it should be noted that with any distribution bounded on one side, the estimation of an additional offset parameter eliminates the issue of fixed bounds alluded
As such, even distributions generally seen as being bounded at zero can be specified with a flexible bound and thus have the potential to signal the presence of counter-intuitively signed coefficient values. In this context, it is also worth noting that a sign change on the attribute can be used in the case of asymmetrical distributions with an a priori constraint on the sign of the skewness (e.g. Lognormal), to act as a mirror function for the distribution of the coefficient.

Given the problems caused by the long tails of some the distributions bounded on just one side (e.g. Lognormal), it can be argued that distributions bounded to either side, with estimated bounds, such as the Johnson $S_B$, have an advantage. In the case of flexible underlying distributions, the risk of values with a counter-intuitive sign being caused by the shape of the distribution, as with the Normal, largely disappears, although problems may still occur in the case of a significant mass at the endpoints, such as in the presence of individuals with zero VTTS [c.f. 8, 19].

### 3.2 Data and model specification

The study presented here makes use of Stated Preference (SP) data collected as part of a recent value of time study undertaken in Denmark [5]. The same dataset was also used in the non-parametric VTTS study of Fosgerau [13], but the results of the two studies are not directly comparable, mainly because of the use of a different subsample.

In this study, we make use of data describing a binomial choice process for car-travellers on shopping trips, with alternatives described only in terms of travel-cost and travel-time. Each respondent was presented with 9 choice-situations, including one with a dominating alternative. After eliminating the observations with a dominating alternative, as well as additional data cleaning (removing non-traders and respondents who did not choose the dominating alternative), a sample of 1,767 observations was obtained, for 230 respondents. With the sole aim of exploring the effects of different distributional assumptions, a very basic utility function was used, such that, in addition to an ASC associated with the first alternative, two coefficients were specified, associated with travel-cost and travel-time respectively, with both attributes entering the utility in linear form. The repeated choice nature of the dataset was accommodated under the assumption of tastes varying across respondents, but not across observations for the same respondent. Finally, no treatment of the correlation between the travel-cost and travel-time coefficients was used in the present analysis.

Aside from a Multinomial Logit (MNL) structure, estimated as the base
model, a high number of different MMNL models were estimated in the current analysis, making use of different continuous distributions for the representation of the variation in the cost and time sensitivity across respondents. Given the high number of possible combinations of distributions, the models were in each case specified with the same choice of distribution for the two coefficients. In total, eleven MMNL models, each with different distributional assumptions, were estimated, making the analysis more comprehensive than most other existing studies looking at the choice of distribution. Several additional distributions, most notably the Beta, could not be used in the present analysis, as it was not possible to estimate an appropriate model. For each type of distribution, the model was coded in Ox 3.40 [10], and estimated using 1,000 Halton draws per dimension and per individual.

For ease of presentation, a common notation was used across distributions, with $\alpha$ and $\gamma$ giving the main parameters of the distribution, where $a$ and $b$ were additionally used as offset and range parameter respectively, where appropriate. We will now look at the various distributions used in the analysis, where details on the actual functional form of the distributions are only given for the Johnson $S_B$ and $S_U$ distributions, with details for the remaining distributions available in the general literature [e.g. 12].

**Normal**: specified with mean $\alpha$ and standard deviation $\gamma$

**Lognormal**: specified with mean $\alpha$ and standard deviation $\gamma$ for the underlying Normal distribution. An additional offset parameter $a$ was estimated, and the attribute entered the utility function under a sign change.

**Johnson $S_B$**: specified with offset parameter $a$, range parameter $b$, and shape parameters $\alpha$ and $\gamma$, with pdf given by:

$$f(x) = \frac{\gamma b}{(x-a)(a+b-x)}\phi\left(\alpha + \gamma \ln \left(\frac{x-a}{a+b-x}\right)\right),$$  \hspace{1cm} (2)

where $a < x < a+b$, $\phi()$ is the standard Normal density function, and where $\alpha \in (-\infty, +\infty)$ and $\gamma > 0$. The shape parameter $\alpha$ has an effect on the skewness of the distribution, where negative values give a right-skewed distribution, zero gives a symmetrical distribution, and positive values give a left-skewed distribution. The second shape parameter $\gamma$ defines the actual shape of the distribution in terms of peak, with values greater than 1 leading to a single, progressively steeper peak, while values lower than 1 will eventually lead to two peaks/modes at
the extremes of the domain. With the notation in equation (2), a draw from the $S_B$ distribution is obtained as:

$$x = a + \frac{b}{1 + \exp\left(\frac{-(z-\alpha)}{\gamma}\right)},$$  \hspace{1cm} (3)

where $z$ is a draw from a standard Normal distribution.

**Symmetrical Johnson $S_B$:** specified as in equation (2), but with $\alpha = 0$.

**Johnson $S_U$:** specified with probability density function

$$f(x) = \frac{\gamma}{\sqrt{(x-a)^2 + b^2}} \phi\left(\alpha + \gamma \ln\left(\frac{x-a}{b} + \sqrt{\left(\frac{x-a}{b}\right)^2 + 1}\right)\right),$$  \hspace{1cm} (4)

where a draw can be produced as:

$$x = a + b \frac{\exp\left(\frac{z-\alpha}{\gamma}\right)^2 - 1}{2 \exp\left(\frac{z-\alpha}{\gamma}\right)}.$$  \hspace{1cm} (5)

The $S_U$ distribution is unbounded, but, unlike the Normal distribution, it can be asymmetrical, with the skewness depending on $\alpha$. With this distribution, the meaning of $a$ and $b$ is different from the offset and range meaning, given that the distribution is unbounded. Here, they take on the meaning of a location and scale parameter.

**Triangular:** specified to be symmetrical, with lower bound $a$, upper bound $a + b$, and mode at $\frac{a+b}{2}$.

**Gamma distribution:** specified with shape parameter $\alpha$, and scale parameter $\gamma$. An additional offset parameter $a$ was used, and the attribute entered the utility under a sign change.

**Exponential:** specified with scale parameter $\alpha$. An additional offset parameter $a$ was used, and the attribute entered the utility under a sign change.

**Logistic:** specified with location parameter $\alpha$ and scale parameter $\gamma$. 
Weibull: specified with shape parameter $\gamma$, offset parameter $a$, with the attribute entering the utility under a sign change. In the implementation in Ox, the pdf is defined such that the estimate for $\alpha$ is in fact not the standard scale parameter, say $\eta$, but represents $\eta^{-\gamma}$.

Uniform: specified with lower endpoint $a$, and range parameter $b$.

### 3.3 Estimation results

The results of the estimation are summarised in Table 1 for the MNL model and the first group of MMNL models (Normal, Lognormal, Johnson $S_B$ and Johnson $S_U$), and Table 2 for the second group of MMNL models (Gamma, Triangular, Exponential, Logistic, Weibull and Uniform). In each case, the estimation results are shown for the ASC, along with the four parameters ($a$, $b$, $\alpha$, and $\gamma$) defining the distribution of the travel-cost and travel-time coefficient. Additionally, on the basis of the estimated parameters for the distribution of the two coefficients, the probabilities of counter-intuitively signed (i.e. positive) coefficients were calculated in each of the models. Finally, the results also present the mean implied VTTS for each model, along with the standard deviation. In each case, these measures were produced by a simple simulation process, making use of 1,000,000 random draws for each of the two coefficients, based on the final model estimates for the parameters of the distribution. Here, special care was required in the models showing a non-zero probability of a positive travel-cost coefficient. The fact that the domain for the denominator of the VTTS ratio straddles zero in such cases leads to extreme values in the simulation, and an overestimation of the variance in the VTTS. For this reason, the upper and lower (to minimise the distortion to the mean) few percentile points were removed from the distribution of the cost coefficient in the models using the Normal distribution (2%), Triangular distribution (2%) and Logistic distribution (3%). A similar treatment was not used in the case of the travel-time coefficient; here, the removal of a sufficient number of percentile points (e.g. 15% for the Normal) would have led to a severely underestimated standard deviation. Additionally, by being included in the numerator, the presence of values close to zero causes fewer problems than is the case for the cost coefficient.

We will now look at the results in more detail. The first observation relates to the performance of the various approaches in terms of model fit. All eleven MMNL models lead to significant improvements in log-likelihood (LL) over the MNL model, ranging from 57.27 units in the model based on the Exponential distribution, to 66.04 units in the model based on the
distribution with both shape parameters estimated. The differences in performance between the various MMNL models are very small, and, when taking into account the cost in terms of parameters, some of the other models (e.g. Gamma) in fact score slightly higher in the adjusted $\rho^2$ measure than the model based on the $S_B$. Even though the differences in model fit are too small to lend significant weight to any comparisons across models, some interesting observations can be made. As such, all distributions, except for the Exponential, Logistic and Weibull, lead to better model fit than the Normal, from the point of view of the final LL, as well as the adjusted $\rho^2$ measure. While this should come as no surprise in the case of flexible distributions, such as the Johnson $S_B$ and $S_U$, and the Gamma, it is striking that the Uniform distribution obtains slightly better fit than the Normal.

A crucial part of the results looks at the implications in terms of the presence of individuals with counter-intuitively signed values for the travel-time and travel-cost coefficient. A positive value for a travel-time coefficient represents a situation where, all else (i.e. travel-cost) being equal, a respondent prefers the slower alternative. Similarly, a positive value for the travel-cost coefficient represents a situation where, with constant travel-time, a respondent chooses the more expensive alternative. No such observations were included in the estimation; as such, results showing a non-zero probability of a positive travel-time or travel-cost coefficient should be seen as an artifact of the distributional assumptions.

For the travel-cost coefficient, a non-zero probability of a positive coefficient was indicated by the models based on the Normal (1.87%), Triangular (1.12%) and Logistic (2.10%) distributions. For the travel-time coefficient, the situation is more severe, with high probabilities in the case of the Normal (14.53%), Triangular (14.60%) and Logistic (14.74%) distributions. These results could lead to seriously misleading conclusions in terms of the presence of individuals with negative VTTS. Lower probabilities of a wrongly signed travel-time coefficient are observed with the Johnson $S_U$ (0.87%) and the Uniform (4.68%).

Important differences arise between the MNL model and the various MMNL structures in terms of the implied VTTS. Here, the failure to account for the variation in the sensitivity to travel-time and travel-cost leads to a significant underestimation of the mean VTTS in the MNL model. These findings in terms of differences between VTTS estimates produced by MNL and MMNL models are consistent with a similar observation by Algers et al. [1]; however, in that research, the MNL model produced significantly higher VTTS than the MMNL models, while in the present work, the opposite is the case. This is an indication that the bias can act in either direction.
Some differences also exist between the eleven MMNL models in the estimated VTTS. In general, these differences are however relatively small especially when looking at the mean VTTS, which is relatively stable aside from a few outliers, notably the Logistic (underestimation) and Weibull (overestimation), and to a lesser extent the Johnson $S_U$. The differences are more significant when looking at the implied variation in VTTS across respondents. Here, the use of the Normal leads to overestimation, which is partly caused by the presence of negative as well as positive travel-time coefficients in the simulation, a factor that also plays a role in the case of the Triangular, Logistic and Uniform distributions. Finally, the long tails in the case of the Lognormal and Johnson $S_U$ also leads to higher variation in the trade-off.

We will now look at the implied distribution for $\beta_{TC}$ and $\beta_{TT}$ in each of the different MMNL models. For this, plots of the distributions are shown in Figure 1 for the Normal and Lognormal, Figure 2 for the Johnson $S_B$ (asymmetrical and symmetrical), Figure 3 for the Johnson $S_U$ and Gamma, Figure 4 for the Triangular and Exponential, and Figure 5 for the Logistic and Weibull. The distribution for the Uniform is not reproduced here.

The results for the Normal and Lognormal (Figure 1) indicate the presence of a mode relatively close to zero, with a long tail to the left, which is given excessive weight by the Lognormal. In the case of the Normal, the symmetrical nature of the distribution means that, in order to accommodate the strong variation to the left of the mode, the tail to the right extends into the positive part of the domain for both $\beta_{TC}$ and $\beta_{TT}$. Although, with the Lognormal, the additional offset parameter is negative for both $\beta_{TC}$ and $\beta_{TT}$ (and hence positive after a sign-change), the difference to zero is in each case utterly insignificant.

The results for the models based on the asymmetrical and symmetrical Johnson $S_B$ distribution (Figure 2) indicate the presence of two modes for $\beta_{TC}$ and $\beta_{TT}$, a phenomenon that none of the other distributions is able to pick up. Here, a posterior analysis would be of interest, allowing modellers to relate the distribution to socio-demographic information. The difference in model fit between the two approaches is very small, suggesting that the additional constraint on $\alpha$ is acceptable, especially in the case of $\beta_{TT}$, where $\alpha$ was not significantly different from zero in the unconstrained model. Here, it should be noted that, in both models, the additional shape parameter $\gamma$ obtains very low levels of significance, where similar problems with the $S_B$ were already observed by Hess, Bierlaire and Polak [19], caus-
ing some concern\(^1\). Finally, in both models, the implied range for \(\beta_{TC}\) and \(\beta_{TT}\) is exclusively negative, on the basis of the estimated offset and range parameters.

The results for the Johnson \(SU\) and the Gamma distribution (Figure 3) are comparable to those obtained with the Lognormal distribution, a conclusion that also applies in terms of the implied VTTS (mean and standard deviation). However, the \(SU\) seems to lead to an even longer tail than the Lognormal in the case of \(\beta_{TC}\), while, with the Gamma, the tails are more moderate. The results for the Gamma imply strictly negative values for \(\beta_{TC}\) and \(\beta_{TT}\) (after a sign-change), although the offset parameters are different from zero only at the 90% and 45% level respectively. Additionally, some issues with significance arise for \(\alpha\), significant at the 90% and 68% level for \(\beta_{TC}\) and \(\beta_{TT}\) respectively, while \(\gamma\) is significant only at the 86% level for \(\beta_{TT}\). For the Johnson \(SU\), the location parameter is not significantly different from zero at reasonable levels of confidence for either \(\beta_{TC}\) or \(\beta_{TT}\), where issues with significance also arise for the remaining parameters in the case of \(\beta_{TT}\), with confidence levels of 91%, 84% and 87% for \(b\), \(\alpha\) and \(\gamma\) respectively.

The first observation that can be made from Figure 4 is that the model based on the Exponential distribution fails to pick up the variation in \(\beta_{TT}\). The results in Table 2 support the impression that this distribution is inappropriate for \(\beta_{TT}\), with a very high standard error associated with the \(\alpha\) parameter. Aside from that, the offset parameters for both coefficients are significantly different from zero, and, after a sign change, indicate an exclusively negative domain for \(\beta_{TC}\) and \(\beta_{TT}\). The results for the Triangular distribution are very similar to those obtained with the Normal, in terms of model fit, VTTS estimates\(^2\), and crucially, the conclusions in terms of the presence of individuals with positive values for \(\beta_{TC}\) and \(\beta_{TT}\), where this is again an effect of the symmetrical nature of the distribution, in conjunction with a mean close to zero, along with high variation. Here, it should also be noted that the upper bound \(b\) is not different from zero in the case of \(\beta_{TC}\), while, for \(\beta_{TT}\), it is.

The plots for the Logistic and Weibull distribution (Figure 5) again show that, while model fits are comparable, there are important differences

\(^1\)The t-statistic is calculated with respect to 0, and not 1, where the difference would be statistically significant. A value of \(\gamma\) tending to 0 leads to a bi-modal distribution. As such, while suggesting problems in terms of the robustness in the estimation of \(\gamma\), the results can also be seen as a strong indication of the presence of multiple peaks.

\(^2\)Here, the standard deviation is lower in the model based on the Triangular, but is still the second highest across all models.
between symmetrical and asymmetrical distributions in the present application. As such, both distributions pick up significant variation in the sensitivity to $\beta_{TC}$ and $\beta_{TT}$, along with a mean close to zero. While the symmetrical shape of the Logistic distribution leads to probabilities of 2.10% and 14.74% for positive values of $\beta_{TC}$ and $\beta_{TT}$ respectively, the offset parameters ($a$) in the case of the Weibull are significantly different from zero, and, after a sign-change, show an exclusively negative domain for the two coefficients. As was the case with a number of other distributions, issues with parameter significance also arise with the Weibull, namely with $\alpha$ for both coefficients, and $\gamma$ in the case of $\beta_{TT}$.

### 3.4 Summary of findings

The analysis presented in this section has revealed the prevalence of significant levels of variation in the sensitivity to travel-cost and travel-time in the population of travellers used in model estimation, allowing the various MMNL models to offer significant improvements in model fit over the MNL model.

The analysis has also shown that the model fit obtained by the eleven MMNL models is remarkably similar. However, in what is a strong indication that model fit on its own is not a reliable measure when comparing mixture models based on different distributional assumptions, the substantial differences across models are quite important. This manifests itself partly in the estimates for the VTTS, where, although the mean values are comparable, there are some differences in the implied variation in VTTS across respondents. More significant differences arise in terms of the implied shape for the distribution of $\beta_{TC}$ and $\beta_{TT}$. These differences manifest themselves mainly in the form of differences in the weight in the tails, between symmetrical as well as asymmetrical distributions. The most startling result however arises in the case of the $S_B$ distribution, which indicates the presence of two modes in the distribution of $\beta_{TC}$ and $\beta_{TT}$. There is clearly a possibility of more than two modes; the $S_B$ distribution is however limited to two modes.

In the present context, the most interesting differences arise in terms of the bounds on the distribution of the VTTS. In the absence of a direct

---

3With the exception of the model based on the Exponential distribution.

4It can be argued that too little weight is given to the behaviour in the tails of the population. However, it is the tails that are often important in the context of policy analysis, such as in the case of the introduction of tolled facilities, where the upper tails are of interest, and the case of welfare impacts, where the low tails are of interest.
estimation of the VTTS, we are constrained to looking at the probability of positive estimates for $\beta_{TC}$ and $\beta_{TT}$. On the basis of the data used in the estimation, the marginal valuations of changes in travel-time and travel-cost should be exclusively negative, leading to exclusively positive VTTS. However, in the results, several of the models indicate significant probabilities for negative values for $\beta_{TT}$. From the results, it becomes clear that problems arise in the case of symmetrical unbounded distributions (Normal & Logistic), as well as in the case of bounded distributions with a strong shape assumption (symmetrical Triangular). In the present analysis, no problems are observed in the case of flexible distributions bounded either on both sides ($S_B$), or on the side where a strict truncation occurs in the data (Lognormal, Gamma, Exponential and Weibull). These findings are consistent with the theoretical claims made in Section 3.1. Finally, the Johnson $S_U$ distribution, which is unbounded, is flexible enough to accommodate the correct mean and standard deviation, without leading to overestimated weight in the tails, where the probability of 0.87% for a positive $\beta_{TT}$ is negligible.

The findings from this case-study are supported by those obtained by Hess, Bierlaire and Polak [19], who, on the basis of simulated data, show that the use of the Normal distribution can wrongly indicate the presence of individuals with positive values for $\beta_{TT}$ in the case where the true distribution is entirely negative, with a long tail to the left. Additionally, results by Cherchi and Polak [6] support the findings that model fit is not the best indicator when interested in substantive findings such as VTTS, and that results with counter-intuitive signs are often just an artifact of distributional assumptions.

In closing this discussion, it is worth briefly returning to the comparison between the Normal and the Uniform distribution. The results in terms of tail behaviour show far fewer problems for the Uniform distribution with regards to wrongly-signed coefficients than for the Normal (4.68% for the travel-time coefficient, compared to 14.53%). This, in conjunction with the results in terms of model fit, could suggest that the Uniform distribution might be a more appropriate choice of default distribution in the initial search for random taste heterogeneity, especially when also taking into account that models based on the Uniform distribution are generally easier to estimate than those based on the Normal.
4 Summary and Conclusions

This paper has discussed issues of specification and interpretation that need to be faced when using mixture models such as MMNL to represent random variations in tastes across respondents.

The main aim of this paper was to explore the potential of hitherto little used distributions in the estimation of random coefficient models. In this context, a VTTS case-study was conducted, making use of distributions such as the Johnson $S_U$, Gamma, Weibull and Logistic, in addition to more common choices, such as the Normal, Lognormal and Johnson $S_B$. The results from this analysis have shown that, while the different distributions lead to similar performance in terms of model fit, they lead to major differences in the substantive results. These relate partly to the mean and standard deviation of the implied distribution of the VTTS, but manifest themselves especially in terms of the findings with regards to the presence of individuals with negative VTTS. Here, the flexible distributions, such as the Johnson $S_B$, indicate a zero probability for such negative measures, which is consistent with the data. On the other hand, the commonly used Normal distribution indicates a probability of $14.53\%$ for positive valuations of increases in travel-time. Similar problems are observed with other symmetrical distributions.

The results presented in this paper clearly support the notion that the distributional assumptions made during model specification can have a significant impact on model results, a fact that needs to be borne in mind in the interpretation of the results. Furthermore, at least in the present study, the results validate the theoretical claims from Section 3.1, showing a lower risk of misspecification with more flexible distributions. This suggests that modellers should increasingly look into the use of alternatives to the Normal distribution for the representation of random taste heterogeneity. While, in some cases, the use of the Normal may be appropriate in studies interested solely in the mean and standard deviation, the presence of strong asymmetries in the true distribution can lead to bias even in these more overall measures, as shown in the case of the standard deviation in the application presented in this paper. In no case however should a distribution like the Normal be used to infer any conclusions in relation to the behaviour in the tails of the population. Here, it seems preferable to use distributions bounded on either side, such as the Johnson $S_B$, with bounds estimated from the data, to still allow for the effects of data problems or incomplete specifications to manifest themselves.

More work remains to be done, in terms of further tests with the dis-
tributions used in this paper, as well as the use of other distributions, or mixtures of distributions. Additionally, the high standard errors observed for some of the parameters of the more flexible distributions (e.g. Johnson SB, SU and Gamma) are a cause for concern, and it remains to be seen whether these findings are specific to the application at hand. Finally, to allow more widespread use of such distributions in mixture models, some effort needs to go into devising efficient and robust approaches for estimating models based on distributions more complex than the Normal.

References


Mixed Logit model for vehicle choice, Transportation Research Part B: Methodological, forthcoming.


Table 1: Estimation results for MNL model and MMNL models based on Normal, Lognormal, Johnson $S_B$ and Johnson $S_U$ distributions

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MNL</th>
<th>Normal</th>
<th>Lognormal</th>
<th>Johnson $S_B$</th>
<th>Johnson $S_B$ (sym.)</th>
<th>Johnson $S_U$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final LL</td>
<td>-1096.64</td>
<td>-1038.00</td>
<td>-1031.85</td>
<td>-1030.60</td>
<td>-1031.97</td>
<td>-1032.77</td>
</tr>
<tr>
<td>Parameters</td>
<td>3</td>
<td>5</td>
<td>7</td>
<td>9</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>adj. $\rho^2$</td>
<td>0.1022</td>
<td>0.1484</td>
<td>0.1518</td>
<td>0.1512</td>
<td>0.1517</td>
<td>0.1494</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>est. (t-stat.)</th>
<th>est. (t-stat.)</th>
<th>est. (t-stat.)</th>
<th>est. (t-stat.)</th>
<th>est. (t-stat.)</th>
<th>est. (t-stat.)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ASC</strong></td>
<td>0.302 (5.92)</td>
<td>0.367 (6.34)</td>
<td>0.369 (6.35)</td>
<td>0.363 (6.29)</td>
<td>0.361 (6.28)</td>
<td>0.366 (6.46)</td>
</tr>
<tr>
<td><strong>Cost (DKK)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a$</td>
<td>-</td>
<td>-</td>
<td>-0.37 (-0.04)</td>
<td>-75.727 (-7.63)</td>
<td>-68.348 (-7.99)</td>
<td>-6.951 (-1)</td>
</tr>
<tr>
<td>$b$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>56.686 (6.14)</td>
<td>51.372 (6.02)</td>
<td>0.095 (3.52)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>-17.295 (-10.44)</td>
<td>-36.146 (-9.88)</td>
<td>3.543 (10.82)</td>
<td>-0.147 (-4.82)</td>
<td>0</td>
<td>5.906 (4.34)</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>-</td>
<td>17.373 (6.53)</td>
<td>0.7 (3.39)</td>
<td>0.006 (0.29)</td>
<td>0.065 (0.64)</td>
<td>0.953 (4.74)</td>
</tr>
<tr>
<td>% positive</td>
<td>0.00%</td>
<td>1.87%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

| **Time (min)** | | | | | | |
| $a$       | -             | -             | -0.22 (-0.03) | -35.958 (-3.84) | -30.747 (-6)  | 7.439 (0.38) |
| $b$       | -             | -             | -            | 29.94 (2.57)    | 25.036 (3.27) | 0.284 (1.68) |
| $\alpha$  | -8.293 (-6.37) | -16.21 (-7.11) | 2.687 (4.24) | -0.258 (-0.78)  | 0             | 10.385 (1.39) |
| $\gamma$  | -             | 15.341 (6.46) | 0.746 (2.48) | 0.276 (0.84)    | 0.152 (0.58)  | 2.022 (1.5)  |
| % positive| 0.00%         | 14.53%        | 0.00%        | 0.00%          | 0.00%          | 0.87%         |

<table>
<thead>
<tr>
<th>VTTS ($\mu$)</th>
<th>DKK/hour</th>
<th>DKK/hour</th>
<th>DKK/hour</th>
<th>DKK/hour</th>
<th>DKK/hour</th>
<th>DKK/hour</th>
</tr>
</thead>
<tbody>
<tr>
<td>VTTS ($\sigma$)</td>
<td>-</td>
<td>108.11</td>
<td>59.56</td>
<td>34.76</td>
<td>35.52</td>
<td>50.62</td>
</tr>
<tr>
<td>Parameter</td>
<td>est. (t-stat.)</td>
<td>est. (t-stat.)</td>
<td>est. (t-stat.)</td>
<td>est. (t-stat.)</td>
<td>est. (t-stat.)</td>
<td>est. (t-stat.)</td>
</tr>
<tr>
<td>-----------</td>
<td>---------------</td>
<td>---------------</td>
<td>---------------</td>
<td>---------------</td>
<td>---------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Final LL</td>
<td>-1031.61</td>
<td>-1037.74</td>
<td>-1039.37</td>
<td>-1038.81</td>
<td>-1039.18</td>
<td>-1035.75</td>
</tr>
<tr>
<td>Parameters</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>adj. $\rho^2$</td>
<td>0.1520</td>
<td>0.1486</td>
<td>0.1473</td>
<td>0.1478</td>
<td>0.1458</td>
<td>0.1503</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>ASC</th>
<th>est. (t-stat.)</th>
<th>est. (t-stat.)</th>
<th>est. (t-stat.)</th>
<th>est. (t-stat.)</th>
<th>est. (t-stat.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost (DKK)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a$</td>
<td>9.636 (1.63)</td>
<td>-80.341 (-7.94)</td>
<td>6.436 (3.4)</td>
<td>-</td>
<td>6.631 (2.34)</td>
<td>-77.714 (-7.69)</td>
</tr>
<tr>
<td>$b$</td>
<td>6.499 (1.38)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>73.602 (6.29)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>1.012 (1.66)</td>
<td>0.026 (7.91)</td>
<td>-35.014 (-9.92)</td>
<td>0.028 (1.36)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.03 (2.01)</td>
<td>-</td>
<td>9.194 (5.95)</td>
<td>0.986 (5.67)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>% positive</td>
<td>0.00%</td>
<td>1.12%</td>
<td>0.00%</td>
<td>2.10%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Time (min)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a$</td>
<td>3.656 (0.6)</td>
<td>-52.673 (-7.87)</td>
<td>16.508 (7.71)</td>
<td>-</td>
<td>16.518 (8.65)</td>
<td>-38.99 (-7.58)</td>
</tr>
<tr>
<td>$b$</td>
<td>19.449 (3.1)</td>
<td>-</td>
<td>19.541 (0.05)</td>
<td>-15.612 (-7.02)</td>
<td>6.559 (0.83)</td>
<td>-40.811 (4.61)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>1.011 (0.99)</td>
<td>-</td>
<td>19.541 (0.05)</td>
<td>-15.612 (-7.02)</td>
<td>6.559 (0.83)</td>
<td>-</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.066 (1.47)</td>
<td>-</td>
<td>-</td>
<td>8.875 (6.31)</td>
<td>0.1 (0.33)</td>
<td>-</td>
</tr>
<tr>
<td>% positive</td>
<td>0.00%</td>
<td>14.60%</td>
<td>0.00%</td>
<td>14.74%</td>
<td>0.00%</td>
<td>4.68%</td>
</tr>
</tbody>
</table>

| DKK/hour | 42.46 | 39.18 | 42.01 | 34.49 | 49.88 | 44.45 |
| DKK/hour | 49.32 | 64.24 | 33.95 | 53.04 | 38.02 | 59.06 |

Table 2: Estimation results for MMNL models based on Gamma, symmetrical Triangular, Exponential, Logistic, Weibull and Uniform distributions
Figure 1: Implied distributions for $\beta_{TC}$ and $\beta_{TT}$: MMNL models based on Normal and Lognormal distributions
Figure 2: Implied distributions for $\beta_{TC}$ and $\beta_{TT}$: MMNL models based on Johnson $S_B$ (asymmetrical and symmetrical) distribution
Figure 3: Implied distributions for $\beta_{TC}$ and $\beta_{TT}$: MMNL models based on Johnson $SU$ and Gamma distributions.
Figure 4: Implied distributions for $\beta_{TC}$ and $\beta_{TT}$: MMNL models based on Triangular and Exponential distributions
Figure 5: Implied distributions for $\beta_{TC}$ and $\beta_{TT}$: MMNL models based on Logistic and Weibull distributions