Discrete choice models with multiplicative error terms

Fosgerau, Mogens and Bierlaire, Michel

Technical University of Denmark

2009

Discrete choice models with multiplicative error terms

M. Fosgerau* M. Bierlaire†

September 15, 2008

2nd revised version, submitted for possible publication in *Transportation Research Part B*.

*Technical University of Denmark. Email: mf@transport.dtu.dk

†Ecole Polytechnique Fédérale de Lausanne, Transport and Mobility Laboratory, Station 18, CH-1015 Lausanne, Switzerland. Email: michel.bierlaire@epfl.ch

Abstract

The conditional indirect utility of many random utility maximization (RUM) discrete choice models is specified as a sum of an index V depending on observables and an independent random term ε. In general, the universe of RUM consistent models is much larger, even fixing some specification of V due to theoretical and practical considerations. In this paper we explore an alternative RUM model where the summation of V and ε is replaced by multiplication. This is consistent with the notion that choice makers may sometimes evaluate relative differences in V between alternatives rather than absolute differences. We develop some properties of this type of model and show that in several cases the change from an additive to a multiplicative formulation, maintaining a specification of V, may lead to a large improvement in fit, sometimes large than that gained from introducing random coefficients in V.

1 Introduction

Discrete choice models are widely used. They have a firm theoretical foundation in utility theory and can be adapted to a wide range of circumstances. Various very general and flexible nonparametric discrete choice models exist, but they tend not to be used so often in applied research for various reasons. Instead a more limited range of models is employed based on a set of often applied assumptions. The objective of the present paper is to show how a small modification of typical applied models may sometimes lead to large improvements in fit without requiring additional parameters to be estimated. The paper shows how these modified models fit into the general framework of random utility maximization and derives some basic results about the modified models that parallel established results for the linear in parameters multinomial logit model and some of its generalizations. The results of this paper should thus be of interest for the applied researcher.

McFadden and Train (2000), e.g., derive the general random utility maximization (RUM) discrete choice model from first principles. This model specifies the conditional indirect utility (CIU) associated with an alternative j as a function of observed and unobserved attributes (z_j and ε_j) of the alternative, and of observed and unobserved individual characteristics (s and ν), that is

$$U^*(z_j, s, \varepsilon_j, \nu). \tag{1}$$

Further specification of this model is necessary before it can be applied to data. In particular, we specify a subutility for each alternative j by

$$V_j = V(z_j, s, \nu), \tag{2}$$

which does not depend on unobserved characteristics of the alternative. We proceed under the assumption that the researcher desires to specify V up to a number of parameters to be estimated. Now, the CIU becomes

$$U^*(V_j, \varepsilon_j), \tag{3}$$

which embodies the assumption that the variables z_j, s and ν may be summarized by V_j. This is often called an index assumption where V_j is the index. The corresponding choice model is

$$P(i|z,s) = \int P(i|z,s,\nu)f(\nu)d\nu, \tag{4}$$

where $f(\nu)$ represents the distribution of ν in the population, and

$$P(i|z,s,\nu) = \Pr[U^*(V_i,\varepsilon_i) > U^*(V_j,\varepsilon_j) \quad \forall j]. \tag{5}$$

Most applications of these models have used a specification with additive independent error terms, that is

$$U^*(V_j,\varepsilon_j) = V_j + \varepsilon_j, \tag{6}$$

where ε_j is independent of V_j. Some normalization is required for identification, since any strictly increasing transformation of utility will lead to identical observations of choice. It is hence necessary at least to fix the location and scale of utility.[1] This may be done by imposing constraints on the distribution of the error term and on the specification of the V_j.

Linear-in-parameter specifications of V_j ignoring unobserved individual heterogeneity are commonly used, that is

$$V_j = V(z_j,s,\nu) = \beta' x(z_j,s), \tag{7}$$

so that (4)–(5) simplifies to

$$P(i|z,s) = \Pr(\beta' x(z_i,s) + \varepsilon_i > \beta' x(z_j,s) + \varepsilon_j \quad \forall j), \tag{8}$$

and the mixing in (4) is avoided.

Operational models are based on specific assumptions about the distribution of ε_j. Assuming i.i.d. extreme value distributions leads to the multinomial logit (MNL) model, which has been very successful due to its computational and analytical tractability. Multivariate[2] extreme value

[1] CITE Honore & Lewbel, Fosgerau & Nielsen some cases this is all that is required, then something else is identified and may be estimated consistently.

[2] These models are called *Generalized Extreme Value* models by McFadden (1978). However, the name GEV is also used for a family of univariate extreme value distributions (see Jenkinson, 1955).

(MEV) models (McFadden, 1978) relax the assumption of mutual independence. Mixtures of these models are derived to account for unobserved heterogeneity, based on (4)–(5). These models have gained popularity due to their flexibility (McFadden and Train, 2000), while retaining consistency with RUM.

The applied researcher may have theoretical and practical reasons for specifying V in certain ways. One concern is that the parameters of V should have interpretations in terms of elasticities or marginal rates of substitution such as willingness-to-pay. In particular, the linear-in-parameters specification (6)–(7) is very often used. In this paper we treat the specification of V as fixed and focus on the specification of the error structure. We do not require V to be linear-in-parameters.

Given some specification of V, the assumption of additive independent errors (6) is not innocuous. It has strict implications for the range of behavior that the model can describe. From (5), the additivity assumption implies that choice probabilities are invariant with respect to addition of a constant to all the Vs (Daly and Zachary, 1978). In contrast, multiplying the Vs by a positive constant does affect the choice probabilities.

This may or may not be an adequate description of observed behavior. It is quite conceivable that errors in (6) are heteroscedastic, violating the independence assumption. One way that can happen is if choice makers evaluate alternatives in terms of relative differences in V. Facing such issues, if they are indeed detected, one may experiment with the specification of V. Sometimes another quite straight-forward solution may sometimes apply, which is simply to replace the V_j's by logs. Then choice probabilities will no longer be invariant with respect to addition of a constant to all the V_j's, instead they will be invariant with respect to multiplication of all V_j's by a positive constant.

We shall show that under appropriate circumstances, this modified model is still a RUM model. It may be considered a RUM model where the assumption of additive independence of the error terms is replaced by an assumption of multiplicative independence of the error terms. We are thus exploring a second natural specification of (3)

A number of authors have relaxed the assumption of iid errors by explicitly specifying the variance of the additive error term as a function of observed and unobserved individual characteristics (Bhat, 1997; Swait and Adamowicz, 2001; De Shazo and Fermo, 2002; Caussade et al., 2005; Koppelman and Sethi, 2005; Train and Weeks, 2005). Our model modifies the assumption of iid errors in (6) by replacing the assumption by its multiplicative counterpart

$$U^*(V_j, \varepsilon_j) = V_j \varepsilon_j. \tag{9}$$

If we are able to assume that the signs of V_j and ε_j are known, then we are able to take logs without affecting choice probabilities, and the model becomes an additive model, where V_j is replaced by $\ln V_j$.

The realization that there are alternatives to the additive specification of utility is not new. There is a recent literature about nonparametric identification of econometric models, which includes discrete choice models with nonadditive unobservables. This literature is reviewed in Matzkin (2007). **(MICHEL INSERT REF: bibtex code included in this doc).**

The multiplicative formulation is set out in the next section, Section 3 derives some properties of the multiplicative formulation, while Section 4 provides illustrative examples and Section 5 concludes.

2 Model formulation

Assume a general multiplicative utility function over a finite set $\mathcal{C}$ of J alternatives given by (9) where $V_j < 0$ is the systematic part of the utility function, and $\varepsilon_j > 0$ is a random variable, independent of V_j.

We assume that the ε_j are i.i.d. across individuals. The sign restriction on V_j is a natural assumption in many applications, for example when it is defined as a generalized cost, that is, a linear combination of attributes with positive values such as travel time and cost and parameters that are a priori known to be negative.

The choice probabilities (5) under this model are given by

$$P(i|z, s, \nu) = \Pr(V_i \varepsilon_i \geq V_j \varepsilon_j, \quad \forall j). \tag{10}$$

The multiplicative specification is related to the classical specification with additive independent error terms, as can be seen from the following derivation. The logarithm is a strictly increasing function. Consequently,

$$\begin{aligned} P(i|z,s,\nu) &= \Pr(V_i\varepsilon_i \geq V_j\varepsilon_j, \ \ \forall j) \\ &= \Pr(-\ln(-V_i) - \ln(\varepsilon_i) \geq -\ln(-V_j) - \ln(\varepsilon_j), \ \ \forall j). \end{aligned}$$

We define

$$-\ln(\varepsilon_j) = \xi_j/\lambda, \tag{11}$$

where ξ_j are random variables, and $\lambda > 0$ is a scale parameter associated with ξ_j. We obtain

$$\begin{aligned} P(i|z,s,\nu) &= \Pr(\bar{V}_i + \xi_i \geq \bar{V}_j + \xi_j, j \in \mathcal{C}) \\ &= \Pr(-\lambda\ln(-V_i) + \xi_i \geq -\lambda\ln(-V_j) + \xi_j, j \in \mathcal{C}). \end{aligned} \tag{12}$$

Consequently, this model can also be written in the random utility framework with an additive specification, where V is replaced by a logarithmic form:

$$\bar{V}_i = -\lambda\ln(-V_i). \tag{13}$$

In the linear formulation $V_j = \beta' x_j$ with additive errors, identification requires that x_j does not contain a variable that is constant across alternatives. An equivalent normalization in the multiplicative case is to fix a parameter to a either 1 or -1, since multiplying V by a positive constant is equivalent to adding a constant to $\ln(V)$. A useful practice is to normalize the cost coefficient (if present) to 1 so that other coefficients can be readily interpreted as willingness-to-pay indicators.

This specification is fairly general and can be used for all the discrete choice models discussed in the introduction. We are free to make assumptions regarding the error terms ξ_i and the parameters inside V_i can be random. Thus we may obtain MNL, MEV and mixtures of MEV models. For instance, a MNL specification would be

$$P(i|z,s) = \frac{e^{-\lambda\ln(-V_i)}}{\sum_{j\in\mathcal{C}} e^{-\lambda\ln(-V_j)}} = \frac{-V_i^{-\lambda}}{\sum_{j\in\mathcal{C}} -V_j^{-\lambda}}. \tag{14}$$

If random parameters are involved, it is necessary to ensure that $P(V_i \geq 0) = 0$. The sign of a parameter can be restricted using, e.g., an exponential.

For instance, if β has a normal distribution then $\exp(\beta)$ is positive and log-normal. For deterministic parameters one may specify bounds as part of the estimation or transformations such as the exponential may be used to restrict the sign.

The use of (12) provides an equivalent specification with additive independent error terms, which fits into the classical modeling framework, involving MNL and MEV models, and mixtures of these. However, even when the V's are linear-in-parameters, the equivalent additive specification (12) is nonlinear. Therefore, estimation routines must be used, that are capable of handling this. The results presented in this paper have been generated using the software package Biogeme (`biogeme.epfl.ch`; Bierlaire, 2003; Bierlaire, 2005), which allows for the estimation of mixtures of MEV models, with nonlinear utility functions.

3 Model properties

We discuss now some basic properties of the model with multiplicative error terms. As we have noted, we may simply reinterpret the model to have CIU defined by $\bar{V}_i + \xi_i$, which is nonlinear when V_i is linear. This reformulation yields identical choice probabilities but has additive error terms, such that standard theory may be applied.

Distribution From (11), we derive the CDF of ε_i as

$$F_{\varepsilon_i}(x) = 1 - F_{\xi_i}(-\lambda \ln x).$$

In the case where ξ_i is extreme value distributed, the CDF of ξ_i is

$$F_{\xi_i}(x) = e^{-e^{-x}}$$

and, therefore,

$$F_{\varepsilon_i}(x) = 1 - e^{-x^\lambda}.$$

This is a generalization of an exponential distribution (obtained with $\lambda = 1$). We note that the exponential distribution is the maximum entropy distribution among continuous distributions on the positive half-axis of given mean, meaning that it embodies minimal information in addition to the mean (that is to V_i) and positivity. Thus, it is seems to be an appropriate choice for an unknown error term.

Elasticities The direct elasticity of alternative i with respect to an attribute of the ith alternative x_k is defined as

$$e_{ik} = \frac{\partial P(i)}{\partial x_k}\frac{x_k}{P(i)} = \frac{\partial P(i)}{\partial V_i}\frac{\partial V_i}{\partial x_k}\frac{x_k}{P(i)},$$

where $\partial V_i/\partial x_k = \beta_k$ if V_i is linear. We use (13) to obtain

$$e_{ik} = \frac{\partial P(i)}{\partial \bar{V}_i}\frac{\partial \bar{V}_i}{\partial V_i}\frac{\partial V_i}{\partial x_k}\frac{x_k}{P(i)} = -\frac{\lambda}{V_i}\frac{\partial P(i)}{\partial \bar{V}_i}\frac{\partial V_i}{\partial x_k}\frac{x_k}{P(i)}$$

where $\partial P(i)/\partial \bar{V}_i$ may be derived from the corresponding additive model. For instance, if the additive model is MNL, we have

$$\frac{\partial P(i)}{\partial \bar{V}_i} = P(i)(1 - P(i)),$$

and

$$e_{ik} = -\frac{\lambda}{V_i}(1 - P(i))\frac{\partial V_i}{\partial x_k}x_k.$$

Similarly, the cross-elasticity e_{ijk} of alternative i with respect to an attribute x_k of alternative j is given by

$$e_{ik} = -\frac{\lambda}{V_j}\frac{\partial P(i)}{\partial \bar{V}_j}\frac{\partial V_j}{\partial x_k}\frac{x_k}{P(i)}$$

where $\partial P(i)/\partial \bar{V}_j$ can be derived from the corresponding additive model. For instance, if the additive model is MNL, we have

$$\frac{\partial P(i)}{\partial \bar{V}_j} = -P(i)P(j),$$

and

$$e_{ik} = \frac{\lambda}{V_i}P(j)\frac{\partial V_j}{\partial x_k}x_k.$$

Trade-offs The trade-offs are computed in the exact same way as for an additive model, that is

$$\frac{\partial U_i/\partial x_{ik}}{\partial U_i/\partial x_{i\ell}} = \frac{\partial V_i/\partial x_{ik}}{\partial V_i/\partial x_{i\ell}},$$

as $\partial \varepsilon_i/\partial x_{ik} = \partial \varepsilon_i/\partial x_{i\ell} = 0$, because ε_i is independent of V_i.

Expected maximum utility The maximum utility under the definition of utility in (9) is

$$U^* = \max_{i\in\mathcal{C}} U_i = \max_{i\in\mathcal{C}} V_i\varepsilon_i = \max_{i\in\mathcal{C}} V_i e^{-\frac{\xi_i}{\lambda}}, \tag{15}$$

where ξ_i is defined by (11). We assume that $(\xi_1, \ldots, \xi_J)$ follows a MEV distribution, that is

$$F(\xi_1, \ldots, \xi_J) = e^{-G(e^{-\xi_1}, \ldots, e^{-\xi_J})}, \tag{16}$$

where G is a σ-homogeneous function with certain properties (see McFadden, 1978 and Daly and Bierlaire, 2006 for details). Then the expected maximum utility is given by (see derivation in Appendix A):

$$E[U^*] = -(G^*)^{-\frac{1}{\sigma\lambda}}\Gamma\left(1 + \frac{1}{\sigma\lambda}\right), \tag{17}$$

where

$$G^* = G((-V_1)^{-\lambda}, \ldots, (-V_J)^{-\lambda}), \tag{18}$$

and $\Gamma(\cdot)$ is the gamma function.

We can compare this to the expected maximum utility if utility is taken to be $-\lambda\ln(-V_i) + \xi_i$. Using the well-known result (McFadden, 1978), the expected maximum utility is then $\frac{1}{\sigma}(\ln G^* + \gamma)$.

It is thus apparent that for the same definition of V, the multiplicative and the additive specifications of the model lead to quite different expected utilities. But, essentially, the V_i enter the expected maximum utility through G^* in both expressions. Hence the marginal expected maximum utility of a change to some V_i divided by the marginal utility of income will be the same for either formulation.

Marshallian consumer surplus The Marshallian consumer surplus can be derived in the context where $-V_i$, the negative of the subutility of alternative i, is interpreted as a generalized cost. In this case, when a small perturbation dV_i is applied, the compensating variation is simply $-dV_i$ if alternative i is chosen, and 0 otherwise. Therefore, the compensating variation for a marginal change dV_i in V_i is

$$-P(i)dV_i, \tag{19}$$

and the compensating variation for changing V_i from a to b is given by

$$-\int_a^b P(i)dV_i. \tag{20}$$

When $P(i)$ is given by a classical MNL model, this integral leads to the well-known logsum formula[3]. When $P(i)$ is given by the model with multiplicative error (like (14)), the integral does not have a closed form in general and numerical integration must be performed[4]. We refer the reader to Dagsvik and Karlström (2005) for a discussion of compensating variation in the context of discrete choice.

Heterogeneity of the scale of utility Assume that the utility can be decomposed as

$$U^*(V_j, \varepsilon_j) = \tilde{V}(z_j, s)\mu(s, \nu)\varepsilon_j. \tag{21}$$

That is, individual observed and unobserved heterogeneity ν affects only the scale of the utility. Combining (5) and (6) under the additive specification gives

$$P(i|z, s, \nu) = \Pr(\tilde{V}(z_i, s)\mu(s, \nu) + \varepsilon_i > \tilde{V}(z_j, s)\mu(s, \nu) + \varepsilon_j \;\; \forall j), \tag{22}$$

while combining (5) and (9) under the multiplicative specification gives

$$P(i|z, s, \nu) = \Pr(\tilde{V}(z_i, s)\mu(s, \nu)\varepsilon_i > \tilde{V}(z_j, s)\mu(s, \nu)\varepsilon_j) \;\; \forall j), \tag{23}$$

[3] Anders Karlstrom has helped us find references for this result. The earliest reference we could find is Neuburger (1971)

[4] Complicated closed form expressions can be derived for (14) with integer values of λ. But λ is estimated and unlikely to be integer.

which simplifies to

$$P(i|z,s,\nu) = \Pr(\tilde{V}(z_i,s)\varepsilon_i > \tilde{V}(z_j,s)\varepsilon_j \;\; \forall j). \tag{24}$$

So the scale of utility is irrelevant for probabilities under the multiplicative formulation, also when the scale of utility is distributed in the population.

4 Empirical applications

We analyze three stated choice panel data sets. We start with two data sets for value of time estimation, from Denmark and Switzerland, where the choice model is binomial. The third data set, a trinomial mode choice in Switzerland, allows us to test the specification with a nested logit model.

4.1 Value of time in Denmark

We utilize data from the Danish value-of-time study. We have selected an experiment that involves several attributes in addition to travel time and cost. We report the analysis for the train segment in detail, and provide a summary for the bus and car driver segments. The experiment is a binary route choice with unlabeled alternatives.

The first model is a simple logit model with linear-in-parameters subutility functions. The attributes are the cost, in-vehicle time, number of changes, headway, waiting time and access-egress time (ae).

The subutility function is defined as

$$\begin{array}{rlll} V_i = \lambda(& - \;\; \text{cost} & +\beta_1 \, \text{ae} & +\beta_2 \, \text{changes} \\ & + \;\; \beta_3 \, \text{headway} & +\beta_4 \, \text{inVehTime} & +\beta_5 \, \text{waiting} \quad), \end{array} \tag{25}$$

where the cost coefficient is normalized to -1 and the parameter λ is estimated. The subutility function in log-form, used in the estimation software for the multiplicative specification, is defined as

$$\begin{array}{rlll} V_i = -\lambda \, \log(& \quad \text{cost} & -\beta_1 \, \text{ae} & -\beta_2 \, \text{changes} \\ & - \;\; \beta_3 \, \text{headway} & -\beta_4 \, \text{inVehTime} & -\beta_5 \, \text{waiting}) \quad . \end{array} \tag{26}$$

The estimation results are reported in Table 6 for the additive specification and in Table 7 for the multiplicative specification. We observe a significant improvement in the log-likelihood (171.76) for the multiplicative specification relative to the additive.

The second model captures unobserved taste heterogeneity. Its estimation accounts for the panel nature of the data. The specification of the subutility is

$$V_i = \lambda(-\text{cost} - e^{\beta_5+\beta_6\xi} Y_i) \tag{27}$$

where

$$Y_i = \text{inVehTime} + e^{\beta_1}\ \text{ae} + e^{\beta_2}\ \text{changes} + e^{\beta_3}\ \text{headway} + e^{\beta_4}\ \text{waiting}, \tag{28}$$

ξ is a random parameter distributed across individuals as $N(0,1)$, so that $e^{\beta_5+\beta_6\xi}$ is log-normally distributed. The exponentials guarantee the positivity of the parameters. The subutility function in log-form, used in the estimation software for the multiplicative specification, is defined as

$$V_i = -\lambda \log(\text{cost} + e^{\beta_5+\beta_6\xi} Y_i), \tag{29}$$

where Y_i is defined by (28).

The estimation results are reported in Table 8 for the additive specification and in Table 9 for the multiplicative specification. Again, the improvement of the goodness-of-fit for the multiplicative is remarkable (225.45).

Finally, we present a model capturing both observed and unobserved heterogeneity. The specification of the subutility is

$$V_i = \lambda(-\text{cost} - e^{W_i} Y_i)$$

where Y_i is defined by (28),

$$W_i = \beta_5\ \text{highInc} + \beta_6 \log(\text{inc}) + \beta_7\ \text{lowInc} + \beta_8\ \text{missingInc} + \beta_9 + \beta_{10}\xi$$

and ξ is a random parameter distributed across individuals as $N(0,1)$. The subutility function in log form is

$$V_i = -\lambda \log(\text{cost} + e^{W_i} Y_i).$$

Number of observations	3455		
Number of individuals	523		
Model	Additive	Multiplicative	Difference
1	-1970.85	-1799.09	171.76
2	-1924.39	-1698.94	225.45
3	-1914.12	-1674.67	239.45

Table 1: Log-likelihood of the models for the train data set

Number of observations:	7751		
Number of individuals:	1148		
Model	Additive	Multiplicative	Difference
1	-4255.55	-3958.35	297.2
2	-4134.56	-3817.49	317.07
3	-4124.21	-3804.9	319.31

Table 2: Log-likelihood of the models for the bus data set

The estimation results are reported in Table 10 for the additive specification and in Table 11 for the multiplicative specification. We again obtain a large improvement (239.45) of the goodness-of-fit for the multiplicative model.

The log-likelihood of these three models are summarized in Table 1. Similar models have been estimated on the bus and the car data set. The summarized results are reported in Tables 2 and 3.

The multiplicative specification significantly and systematically outperforms the additive specification in these examples. Actually, the multiplicative model where taste heterogeneity is not modeled (model 1) fits the data much better than the additive model where both observed and unobserved heterogeneity are modeled.

Number of observations:	8589		
Number of individuals:	1585		
Model	Additive	Multiplicative	Difference
1	-5070.42	-4304.01	766.41
2	-4667.05	-3808.22	858.83
3	-4620.56	-3761.57	858.99

Table 3: Log-Likelihood of the models for the car data set

4.2 Value of time in Switzerland

We have estimated the models without socio-economics, that is (25), (26), (27) and (29), on the Swiss value-of-time data set (Koenig et al., 2003). We have selected the data from the route choice experiment by rail for actual rail users. As a difference from the models with the Danish data set, we have omitted the attributes `ae` and `waiting`, not present in this data set. The log-likelihood of the four models are reported in Table 4, and the detailed results are reported in Tables 12–15.

The multiplicative specification does not outperform the additive one for the fixed parameters model. Introducing random parameters in a panel data specification improves the log-likelihood of both models, the fit of the multiplicative specification being now clearly the best, although the improvement is not as large as for the Danish data set.

	Additive	Multiplicative	Difference
Fixed parameters	-1668.070	-1676.032	-7.96
Random parameters	-1595.092	-1568.607	26.49

Table 4: Log-likelihood for the Swiss VOT data set

4.3 Swissmetro

We illustrate the model with a data set collected for the analysis of a future high speed train in Switzerland (Bierlaire et al., 2001). The alternatives are

1. Regular train (TRAIN),
2. Swissmetro (SM), the future high speed train,
3. Driving a car (CAR).

We specify a nested logit model with the following nesting structure.

	TRAIN	SM	CAR
NESTA	1	0	1
NESTB	0	1	0

In the base model, the subutilities V_i are defined as follows.

	Alternatives		
Param.	TRAIN	SM	CAR
B_TRAIN_TIME	travel time	0	0
B_SM_TIME	0	travel time	0
B_CAR_TIME	0	0	travel time
B_HEADWAY	frequency	frequency	0
B_COST	travel cost	travel cost	travel cost

We derive 16 variants of this model, each of them including or not the following features:

1. Alternative Specific Socio-economic Characteristics (ASSEC): we add the following terms to the subutility of alternatives SM and CAR:

$$\text{B\_GA\_i railwayPass} + \text{B\_MALE\_i male} + \text{B\_PURP\_i commuter}$$

   where i =SM,CAR;

2. Error component (EC): a normally distributed error component is added to each of the three alternatives, with an alternative specific standard error.

3. Segmented travel time coefficient (STTC): the coefficient of travel time varies with socio-economic characteristics:

$$\text{B\_SEGMENT\_TIME\_i} = \text{-exp(B\_i\_TIME + B\_GA\_i railwayPass +}$$
$$\texttt{B\_MALE\_i}\text{ male + }\texttt{B\_PURP\_i}\text{ commuter)}$$

where i={TRAIN,SM,CAR}.

4. Random coefficient (RC): the coefficients for travel time and headway are distributed, with a log-normal distribution.

For each variant, we have estimated both an additive and a multiplicative specification, using the panel dimension of the data when applicable. The results are reported in Table 5.

	RC	EC	STTC	ASSEC	Additive	Multiplicative	Difference
1	0	0	0	0	-5188.6	-4988.6	200.0
2	0	0	0	1	-4839.5	-4796.6	42.9
3	0	0	1	0	-4761.8	-4745.8	16.0
4	0	1	0	0	-3851.6	-3599.8	251.8
5	1	0	0	0	-3627.2	-3614.4	12.8
6	0	0	1	1	-4700.1	-4715.5	-15.4
7	0	1	0	1	-3688.5	-3532.6	155.9
8	0	1	1	0	-3574.8	-3872.1	-297.3
9	1	0	0	1	-3543.0	-3532.4	10.6
10	1	0	1	0	-3513.3	-3528.8	-15.5
11	1	1	0	0	-3617.4	-3590.0	27.3
12	0	1	1	1	-3545.4	-3508.1	37.2
13	1	0	1	1	-3497.2	-3519.6	-22.5
14	1	1	0	1	-3515.1	-3514.0	1.1
15	1	1	1	0	-3488.2	-3514.5	-26.2
16	1	1	1	1	-3465.9	-3497.2	-31.3

Table 5: Results for the 16 variants on the Swissmetro data

We observe that for simple models (1-5) the multiplicative specification outperforms the additive one. However, this is not necessarily true for more complex models. Overall, the multiplicative specification performs better on 10 variants out of 16. We learn from this example that the multiplicative (as expected) is not universally better, and should not be systematically preferred. However, it is definitely worth testing it, as it has a great potential for explaining the data better.

5 Concluding remarks

It seems to be a common perception that discrete choice models based on random utility maximization must have additive independent error terms. This is not the case, as we have discussed in this paper. It may happen that for some data and some specifications of the subutility, it is more appropriate to assume a multiplicative form. We have indicated how the multiplicative form may be estimated with existing software.

A priori, for a given specification of V, it is not possible to know whether the multiplicative formulation will provide a better fit than the additive formulation. However, in the majority of the cases we have looked at, we find that the multiplicative formulation fits the data better. In quite a few cases, the improvement is very large, sometimes even larger than the improvement gained from allowing for unobserved heterogeneity. We emphasize that we are reporting the complete list of results that we have obtained, whatever they turned out to be. The choice of applications was motivated only by data availability. As both formulations are equally well grounded in theory, we conclude that the choice between formulations is an empirical question and should be answered by the ability of models to fit data.

Of course, given some specification of subutility, the universe of possible models is still larger than we have considered here. We have focused on a multiplicative formulation as a clear cut alternative to the additive formulation with an equally clear cut invariance property. For the multiplicative specification, we are able to derive analogues to some theoretical properties of the additive specification. As an alternative one may redefine the subutility by $\bar{V}_j = -\lambda \ln(-V_j)$ and use existing theory.

6 Acknowledgments

The authors like to thank Katrine Hjort for very competent research assistance and Anders Karlstrom for comments on the paper. This work has been initiated during the First Workshop on Applications of Discrete Choice Models organized at Ecole Polytechnique Fédérale de Lausanne, Switzerland, in September 2005. Mogens Fosgerau acknowledges support from the Danish Social Science Research Council. Three anonymous reviewers provided us with very valuable comments on previous versions of this article.

References

Bhat, C. R. (1997). Covariance heterogeneity in nested logit models: econometric structure and application to intercity travel, *Transportation Research Part B: Methodological* **31**(1): 11–21.

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

Bierlaire, M. (2005). An introduction to BIOGEME version 1.4. biogeme.epfl.ch.

Bierlaire, M., Axhausen, K. and Abay, G. (2001). Acceptance of modal innovation: the case of the Swissmetro, *Proceedings of the 1st Swiss Transportation Research Conference*, Ascona, Switzerland. www.strc.ch.

Caussade, S., Ortúzar, J., Rizzi, L. I. and Hensher, D. A. (2005). Assessing the influence of design dimensions on stated choice experiment estimates, *Transportation Research Part B: Methodological* **39**(7): 621–640.

Dagsvik, J. K. and Karlström, A. (2005). Compensating variation and hicksian choice probabilities in random utility models that are nonlinear in income, *Review of Economic Studies* **72**(1): 57–76.

Daly, A. and Bierlaire, M. (2006). A general and operational representation of generalised extreme value models, *Transportation Research Part B: Methodological* **40**(4): 285–305.

Daly, A. J. and Zachary, S. (1978). Improved multiple choice, *in* D. A. Hensher and M. Q. Dalvi (eds), *Determinants of travel demand*, Saxon House, Sussex.

De Borger, B. and Fosgerau, M. (forthcoming). The trade-off between money and travel time: A test of the theory of reference-dependent preferences, *Journal of Urban Economics* .

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

Fosgerau, M. (2006). Investigating the distribution of the value of travel time savings, *Transportation Research Part B: Methodological* **40**(8): 688–707.

Fosgerau, M. (2007). Using nonparametrics to specify a model to measure the value of travel time, *Transportation Research Part A* **41**(9): 842–856.

Hanemann, W. M. (1984). Discrete/continuous models of consumer demand, *Econometrica* **52**(3): 541–561.

Jenkinson, A. F. (1955). Frequency distribution of the annual maximum (or minimum) values of meteorological elements, *Quarterly journal of the Royal Meteorological Society* **81**: 158–171.

Koenig, A., Abay, G. and Axhausen, K. (2003). Time is money: the valuation of travel time savings in switzerland, *Proceedings of the 3rd Swiss Transportation Research Conference*. http://www.strc.ch/Paper/Koenig.pdf.

Koppelman, F. and Sethi, V. (2005). Incorporating variance and covariance heterogeneity in the generalized nested logit model: an application to modeling long distance travel choice behavior, *Transportation Research Part B: Methodological* **39**(9): 825–853.

McFadden, D. and Train, K. (2000). Mixed MNL models for discrete response, *Journal of Applied Econometrics* **15**(5): 447–470.

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

Neuburger, H. (1971). User benefit in the evaluation of transport and land use plans, *Journal of Transport Economics and Policy* **5**: 52–75.

Swait, J. and Adamowicz, W. (2001). Choice environment, market complexity, and consumer behavior: a theoretical and empirical approach for incorporating decision complexity into models of consumer choice, *Organizational Behavior and Human Decision Processes* **86**(2): 141–167.

Train, K. and Weeks, M. (2005). Discrete choice models in preference space and willingness-to-pay space, *in* R. Scarpa and A. Alberini (eds), *Applications of simulation methods in environmental and resource economics*, The Economics of Non-Market Goods and Resources, Springer, pp. 1–16.

A Derivation of the expected maximum utility

From (15), the maximum utility is

$$U^* = \max_{i \in \mathcal{C}} V_i e^{-\frac{\xi_i}{\lambda}}, \tag{30}$$

where ξ_i is defined by (11). Note that $U^* \leq 0$. We assume that $(\xi_1, \ldots, \xi_J)$ follows a MEV distribution (16). The CDF of U^* is obtained as follows, for $t < 0$:

$$\begin{aligned}
F(t) &= \Pr(U^* \leq t) = \Pr(U_i \leq t, \ \ \forall i) \\
&= \Pr(\xi_i \leq -\lambda \ln(t V_i^{-1}), \ \ \forall i) \\
&= \exp(-G((tV_1^{-1})^{\lambda}, \ldots, (tV_J^{-1})^{\lambda}) \\
&= \exp(-(-t)^{\sigma\lambda} G((-V_1)^{-\lambda}, \ldots, (-V_J)^{-\lambda}) \\
&= \exp(-(-t)^{\sigma\lambda} G^*)
\end{aligned}$$

using the σ-homogeneity of G and the definition (18) of G^*. The CDF can be inverted as

$$F^{-1}(x) = -\left(-\frac{\ln x}{G^*}\right)^{\frac{1}{\sigma\lambda}} = -(G^*)^{-\frac{1}{\sigma\lambda}} \left(\ln\left(\frac{1}{x}\right)\right)^{\frac{1}{\sigma\lambda}}. \tag{31}$$

Denoting the pdf of U^* by $f(t) = F'(t)$, we have

$$E[U^*] = \int_{-\infty}^{0} t f(t) dt = \int_0^1 F^{-1}(x) dx = -(G^*)^{-\frac{1}{\sigma\lambda}} \int_0^1 \left(\ln\left(\frac{1}{x}\right)\right)^{\frac{1}{\sigma\lambda}} dx$$

which leads to (17).

B Parameter estimates for the Danish Value of Time data

Variable number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	ae	-2.00	0.211	-9.46	0.00
2	changes	-36.1	6.89	-5.23	0.00
3	headway	-0.656	0.0754	-8.71	0.00
4	in-veh. time	-1.55	0.159	-9.76	0.00
5	waiting time	-1.68	0.770	-2.18	0.03
6	λ	0.0141	0.00144	9.82	0.00

Number of observations = 3455

$$\mathcal{L}(0) = -2394.824$$

$$\mathcal{L}(\hat{\beta}) = -1970.846$$

$$-2[\mathcal{L}(0) - \mathcal{L}(\hat{\beta})] = 847.954$$

$$\rho^2 = 0.177$$

$$\bar{\rho}^2 = 0.175$$

Table 6: Model with fixed parameters and additive error terms

Variable number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	ae	-0.672	0.0605	-11.11	0.00
2	changes	-5.22	1.54	-3.40	0.00
3	headway	-0.224	0.0213	-10.53	0.00
4	in-veh. time	-0.782	0.0706	-11.07	0.00
5	waiting time	-1.06	0.206	-5.14	0.00
6	λ	5.37	0.236	22.74	0.00

Number of observations = 3455

$$\mathcal{L}(0) = -2394.824$$
$$\mathcal{L}(\hat{\beta}) = -1799.086$$
$$-2[\mathcal{L}(0) - \mathcal{L}(\hat{\beta})] = 1191.476$$
$$\rho^2 = 0.249$$
$$\bar{\rho}^2 = 0.246$$

Table 7: Model with fixed parameters and multiplicative error terms

Variable number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	ae	0.0639	0.357	0.18	0.86
2	changes	2.88	0.373	7.73	0.00
3	headway	-0.999	0.193	-5.17	0.00
4	waiting time	-0.274	0.433	-0.63	0.53
5	unobserved heterogeneity (mean)	0.331	0.178	1.86	0.06
6	unobserved heterogeneity (stderr)	0.934	0.130	7.19	0.00
7	λ	0.0187	0.00301	6.20	0.00

Number of observations = 3455
Number of individuals = 523
Number of draws for SMLE = 1000

$$\mathcal{L}(0) = -2394.824$$
$$\mathcal{L}(\hat{\beta}) = -1925.467$$
$$-2[\mathcal{L}(0) - \mathcal{L}(\hat{\beta})] = 938.713$$
$$\rho^2 = 0.196$$
$$\bar{\rho}^2 = 0.193$$

Table 8: Model unobserved heterogeneity — additive error terms

Variable number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	ae	0.0424	0.0946	0.45	0.65
2	changes	2.24	0.239	9.38	0.00
3	headway	-1.03	0.0983	-10.48	0.00
4	waiting time	0.355	0.207	1.72	0.09
5	unobserved heterogeneity (mean)	-0.252	0.106	-2.38	0.02
6	unobserved heterogeneity (stderr)	1.49	0.123	12.04	0.00
7	λ	7.04	0.370	19.02	0.00

Number of observations = 3455
Number of individuals = 523
Number of draws for SMLE = 1000

$$\mathcal{L}(0) = -2394.824$$
$$\mathcal{L}(\hat{\beta}) = -1700.060$$
$$-2[\mathcal{L}(0) - \mathcal{L}(\hat{\beta})] = 1389.528$$
$$\rho^2 = 0.290$$
$$\bar{\rho}^2 = 0.287$$

Table 9: Model with unobserved heterogeneity — multiplicative error terms

Variable number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	ae	0.0863	0.345	0.25	0.80
2	changes	2.91	0.387	7.51	0.00
3	headway	-0.955	0.190	-5.02	0.00
4	waiting time	-0.285	0.441	-0.65	0.52
5	high income	0.0744	0.321	0.23	0.82
6	log(income)	0.603	0.182	3.31	0.00
7	low income	0.420	0.321	1.31	0.19
8	missing income	-0.542	0.315	-1.72	0.09
9	unobserved heterogeneity (mean)	0.341	0.170	2.01	0.04
10	unobserved heterogeneity (stderr)	0.845	0.0680	12.42	0.00
11	λ	0.0193	0.00315	6.12	0.00

Number of observations = 3455
Number of individuals = 523
Number of draws for SMLE = 1000

$$\mathcal{L}(0) = -2394.824$$
$$\mathcal{L}(\hat{\beta}) = -1914.180$$
$$-2[\mathcal{L}(0) - \mathcal{L}(\hat{\beta})] = 961.286$$
$$\rho^2 = 0.201$$
$$\bar{\rho}^2 = 0.196$$

Table 10: Model with observed and unobserved heterogeneity — additive error terms

Variable number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	ae	0.0366	0.0925	0.40	0.69
2	changes	2.22	0.239	9.32	0.00
3	headway	-1.02	0.0962	-10.59	0.00
4	waiting time	0.366	0.199	1.84	0.07
5	high income	0.577	0.704	0.82	0.41
6	log(income)	1.21	0.272	4.47	0.00
7	low income	0.770	0.418	1.84	0.07
8	missing income	-0.798	0.371	-2.15	0.03
9	unobserved heterogeneity (mean)	-0.150	0.111	-1.34	0.18
10	unobserved heterogeneity (stderr)	1.28	0.108	11.87	0.00
11	λ	7.13	0.371	19.25	0.00

Number of observations = 3455
Number of individuals = 523
Number of draws for SMLE = 1000

$$\mathcal{L}(0) = -2394.824$$
$$\mathcal{L}(\hat{\beta}) = -1675.412$$
$$-2[\mathcal{L}(0) - \mathcal{L}(\hat{\beta})] = 1438.822$$
$$\rho^2 = 0.300$$
$$\bar{\rho}^2 = 0.296$$

Table 11: Model with observed and unobserved heterogeneity — multiplicative error terms

C Parameter estimates for the Swiss Value of Time data

Variable number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	travel time	-0.453	0.0383	-11.82	0.00
2	changes	-8.74	1.22	-7.17	0.00
3	headway	-0.284	0.0406	-7.01	0.00
4	λ	0.132	0.0188	7.02	0.00

Number of observations = 3501
Number of individuals = 389

$$\begin{aligned}
\mathcal{L}(0) &= -2426.708 \\
\mathcal{L}(\hat{\beta}) &= -1668.070 \\
-2[\mathcal{L}(0) - \mathcal{L}(\hat{\beta})] &= 1517.276 \\
\rho^2 &= 0.313 \\
\bar{\rho}^2 &= 0.311
\end{aligned}$$

Table 12: Model with fixed parameters and additive error terms

Variable number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	travel time	-0.339	0.0285	-11.89	0.00
2	changes	-3.91	0.789	-4.95	0.00
3	headway	-0.140	0.0287	-4.90	0.00
4	λ	8.55	0.907	9.42	0.00

Number of observations = 3501
Number of individuals = 389

$$\begin{aligned} \mathcal{L}(0) &= -2426.708 \\ \mathcal{L}(\hat{\beta}) &= -1676.032 \\ -2[\mathcal{L}(0) - \mathcal{L}(\hat{\beta})] &= 1501.353 \\ \rho^2 &= 0.309 \\ \bar{\rho}^2 &= 0.308 \end{aligned}$$

Table 13: Model with fixed parameters and multiplicative error terms

Variable number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	unobserved heterogeneity (mean)	-0.763	0.111	-6.86	0.00
2	unobserved heterogeneity (stderr)	0.668	0.0582	11.48	0.00
3	changes	2.67	0.108	24.78	0.00
4	headway	-0.798	0.126	-6.34	0.00
5	λ	0.202	0.0367	-5.51	0.00

Number of observations = 3501
Number of individuals = 389
Number of draws for SMLE = 1000

$$\begin{aligned} \mathcal{L}(0) &= -2426.708 \\ \mathcal{L}(\hat{\beta}) &= -1595.092 \\ -2[\mathcal{L}(0) - \mathcal{L}(\hat{\beta})] &= 1663.233 \\ \rho^2 &= 0.343 \\ \bar{\rho}^2 &= 0.341 \end{aligned}$$

Table 14: Model with unobserved heterogeneity — additive error terms

Variable number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	unobserved heterogeneity (mean)	-0.956	0.119	-8.04	0.00
2	unobserved heterogeneity (stderr)	-1.18	0.140	-8.39	0.00
3	changes	2.44	0.116	20.93	0.00
4	headway	-0.856	0.124	-6.90	0.00
5	λ	11.5	1.13	10.16	0.00

Number of observations = 3501
Number of individuals = 389
Number of draws for SMLE = 1000

$$\begin{aligned}
\mathcal{L}(0) &= -2426.708 \\
\mathcal{L}(\hat{\beta}) &= -1568.607 \\
-2[\mathcal{L}(0) - \mathcal{L}(\hat{\beta})] &= 1716.202 \\
\rho^2 &= 0.354 \\
\bar{\rho}^2 &= 0.352
\end{aligned}$$

Table 15: Model with unobserved heterogeneity — multiplicative error terms