MPRA
Munich Personal RePEc Archive

Circumventing the problem of the scale: discrete choice models with multiplicative error terms

Fosgerau, Mogens and Bierlaire, Michel

Technical University of Denmark

3 July 2007

Online at https://mpra.ub.uni-muenchen.de/3901/
MPRA Paper No. 3901, posted 08 Jul 2007 UTC

Circumventing the problem of the scale: discrete choice models with multiplicative error terms

M. Fosgerau* M. Bierlaire†

July 3, 2007

Report TRANSP-OR 070703
Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne
transp-or.epfl.ch

*Technical University of Denmark. Email: mf@dtf.dk
†Ecole Polytechnique Fédérale de Lausanne, Transport and Mobility Laboratory, Station 18, CH-1015 Lausanne, Switzerland. Email: michel.bierlaire@epfl.ch

Abstract

This paper is an updated version of the paper "Discrete choice models with multiplicative error terms" by Fosgerau and Bierlaire (2006).

We propose a multiplicative specification of a discrete choice model that renders choice probabilities independent of the scale of the utility. The scale can thus be random with unspecified distribution. The model mostly outperforms the classical additive formulation over a range of stated choice data sets. In some cases, the improvement in likelihood is greater than that obtained from adding observed and unobserved heterogeneity to the additive specification. The multiplicative specification makes it unnecessary to capture scale heterogeneity and, consequently, yields a significant potential for reducing model complexity in the presence of heteroscedasticity. Thus the proposed multiplicative formulation should be a useful supplement to the techniques available for the analysis of discrete choices. There is however a cost to be paid in terms of increased analytical complexity relative to the additive formulations.

1 Introduction

Discrete choice models have been a major part of the transport analyst's toolbox for decades. These models are able to accommodate diverse requirements and they have a firm theoretical foundation in utility theory. Random utility models with additive independent error terms pose the problem that the scale of the error terms is not identified. Earlier models assumed the problem away by requiring the scale to be constant. Later contributions have allowed the scale to vary across data sets and individuals. We propose instead a multiplicative specification of discrete choice models that circumvents the problem by making the scale irrelevant. It can thus be random and have any distribution. This specification is applicable in situations where we have a priori information about the sign of the systematic utility.

The multinomial logit (MNL) model has been very successful, due to its computational and analytical tractability. Later, multivariate extreme value (MEV) models (McFadden, 1978[1]) and mixtures of MNL and MEV models have gained popularity due to their flexibility and theoretical results relating these models to random utility maximization (McFadden and Train, 2000).

So far, most applications of these models have used a specification with additive independent error terms. It is computationally convenient, which may explain its systematic use. The basic formulation of MNL and MEV models assumes that the scale parameter μ is constant across the population, and can therefore be arbitrarily normalized. This assumption is strong, and a number of techniques to relax it have been developed in the literature, as detailed below.

Additive models are sensitive to the scale of the independent variables x. Multiplying the x by a positive number does affect the choice probabilities. We hypothesize that this may not always be a good description of behavior. Particularly in a stated choice context respondents may inter-

[1]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).

pret the presented numbers relatively to each other, performing an implicit scaling before making their choice. The multiplicative error specification is insensitive to such scaling, and would better describe this behavior.

The additive specification is however not required by utility theory. There are alternative formulations which cannot be ruled out a priori. In this paper we investigate a multiplicative specification, which is the natural alternative to the additive specification.

McFadden has formulated discrete choice theory based on Random Utility Maximization (RUM). For example, McFadden (2000) describes how the indirect conditional utility function is separated into a systematic part and a residual term summarizing all unobserved factors. It is clear that additivity and independence of the residual term are additional assumptions that are made for computational convenience. In this paper we look at an alternative to the specification of additive residuals while retaining the specification of the systematic part of the indirect conditional utility function.

With an additive specification, the scale of the error term is confounded with the parameters of V_i. Indeed, if $U_i = V_i + \mu\varepsilon_i$, normalizing the error terms across individuals amounts to estimating the utility function

$$\frac{1}{\mu}V_i + \varepsilon_i,$$

so that V_i/μ is actually estimated instead of V_i. This is problematic when the scale μ varies across the population. For instance, in the linear-in-parameters case where $V_i = \beta' x_i$, the distribution of β is confounded with the distribution of μ. Even if β is fixed, β/μ is distributed. Moreover, the distribution of μ introduces correlation across the β, which complicates the estimation.

These issues may be addressed by explicitly specifying a distribution for μ (Bhat, 1997; Swait and Adamowicz, 2001; De Shazo and Fermo, 2002; Caussade et al., 2005; Koppelman and Sethi, 2005; Train and Weeks, 2005). Our multiplicative specification avoids the problem altogether.

Train and Weeks (2005) compare a model in preference space to a model in willingness-to-pay space (WTP). The model in preference space assumes independent random coefficients for all alternative attributes and additive

errors, while the model in WTP space assumes a coefficient of one for the cost attribute and independent random coefficients for the remaining attributes as well as a random scale of the still additive error term. Random coefficients are assumed to be either normal or lognormal. They find that the model in preference space fits their data better while the model in WTP space produces more reasonable results for the distribution of WTP. For both models, they furthermore reject the maintained hypothesis that coefficients are independent.

In previous work on the Danish value-of-time survey (Fosgerau, 2006, Fosgerau, forthcoming), we have derived a model that circumvents the above-mentioned scaling effect. However, this model contains only travel time and cost, and is only applicable to very simple stated choice designs. The multiplicative specification proposed in this paper accommodates more general designs involving a higher number of factors.

Our multiplicative specification starts from the assumption that $U_i = \mu V_i \varepsilon_i$. If we are able to assume that the signs of μ, V_i and ε_i are known, then taking logs does not affect choice probabilities, and the model then becomes an additive model.

With this model it is the relative differences that matter. If V_i is linear in travel time then the effect on choice probabilities of a 10 minute difference in travel times depends on the length of the trip under the multiplicative specification. A 10 minute difference under the additive specification has constant effect on choice probabilities regardless of whether it relates to a very short or a very long journey. Thus using the multiplicative specification may reduce the need for segmentation and may hence be able to use data more efficiently.

This is similar to the common practice in econometrics of expressing most variables in regressions in logs. Applying logs in the regression context removes the scale from the data, such that the errors for small and large values of the independent variables have the same variance.

The methodology is set out in the next section, and illustrated in Section 4. We conclude the paper with some remarks in Section 5. Finally, appendices A and B provides the details of parameter estimates of various models, and appendix presents the derivation of the expected maximum

utility of the model with multiplicative error terms.

2 Methodology

Assume a general multiplicative utility function over a finite set $\mathcal{C}$ of J alternatives given by

$$U_i = \mu V_i \varepsilon_i, \tag{1}$$

where μ is an independent individual specific scale parameter, $V_i < 0$ is the systematic part of the utility function, and $\varepsilon_i > 0$ is a random variable, independent of V_i and μ.

We assume that the ε_i are i.i.d. across individuals, and potential heteroscedasticity is captured by the individual specific scale μ. The sign restriction on V_i 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 under this model are given by

$$\begin{aligned} P(i|\mathcal{C}) &= \Pr(U_i \geq U_j, j \in \mathcal{C}) \\ &= \Pr(\mu V_i \varepsilon_i \geq \mu V_j \varepsilon_j, j \in \mathcal{C}) \\ &= \Pr(V_i \varepsilon_i \geq V_j \varepsilon_j, j \in \mathcal{C}), \end{aligned} \tag{2}$$

such that the individual scale is irrelevant. The multiplicative specification (1) 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,

$$\begin{aligned} P(i|\mathcal{C}) &= \Pr(V_i \varepsilon_i \geq V_j \varepsilon_j, j \in \mathcal{C}) \\ &= \Pr(-V_i \varepsilon_i \leq -V_j \varepsilon_j, j \in \mathcal{C}) \\ &= \Pr(\ln(-V_i) + \ln(\varepsilon_i) \leq \ln(-V_j) + \ln(\varepsilon_j), j \in \mathcal{C}) \\ &= \Pr(-\ln(-V_i) - \ln(\varepsilon_i) \geq -\ln(-V_j) - \ln(\varepsilon_j), j \in \mathcal{C}). \end{aligned}$$

We define

$$-\ln(\varepsilon_i) = (c_i + \xi_i)/\lambda, \tag{3}$$

where c_i is the intercept, $\lambda > 0$ is the scale, and ξ_i are random variables with a fixed mean and scale, and we obtain

$$\begin{aligned} P(i|\mathcal{C}) &= \Pr(\bar{V}_i + \xi_i \geq \bar{V}_j + \xi_j, j \in \mathcal{C}) \\ &= \Pr(-\lambda \ln(-V_i) + c_i + \xi_i \geq -\lambda \ln(-V_j) + c_j + \xi_j, j \in \mathcal{C}), \end{aligned} \tag{4}$$

which is now a classical random utility model with additive error, where

$$\bar{V}_i = -\lambda \ln(-V_i) + c_i, \tag{5}$$

It is important to emphasize that, contrarily to μ in (1), the scale λ is constant across the population, as a consequence of the i.i.d. assumption on the ε_i. Note that V_i must be normalized for the model to be identified. Indeed, for any $\alpha > 0$,

$$-\lambda \ln(-\alpha V_i) + c_i = -\lambda \ln(-V_i) - \lambda \ln(\alpha) + c_i$$

meaning that changing the scale of V_i is equivalent to shifting the constant c_i. When V_i is linear-in-parameters, it is sufficient to fix one parameter to either 1 or -1. 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

$$P(i|\mathcal{C}) = \frac{e^{-\lambda \ln(-V_i)+c_i}}{\sum_{j\in\mathcal{C}} e^{-\lambda \ln(-V_j)+c_j}} = \frac{e^{c_i}(-V_i)^{-\lambda}}{\sum_{j\in\mathcal{C}} e^{c_j}(-V_j)^{-\lambda}}, \tag{6}$$

where e^{c_i}, $i \in \mathcal{C}$ are constants to be estimated. Furthermore, c_i may depend on covariates, such that it is also possible to incorporate both observed and unobserved heterogeneity both inside and outside the log. We illustrate some of these specifications in Section 4.

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 lognormal. 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.

Maximum likelihood estimation of the model can be complicated in the general case. The use of (4) 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 Vs are linear in the parameters, the equivalent additive specification (4) 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.

Distribution From (3), we derive the CDF of ε_i as

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

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,

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

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.

Elasticity The direct elasticity of alternative i with respect to an explanatory variable x_{ik} is defined as

$$e_i = \frac{\partial P(i)}{\partial x_{ik}} \frac{x_{ik}}{P(i)} = \frac{\partial P(i)}{\partial V_i} \frac{\partial V_i}{\partial x_{ik}} \frac{x_{ik}}{P(i)},$$

where $\partial V_i / \partial x_{ik} = \beta_k$ if V_i is linear-in-parameters. We use (5) to obtain

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

where $\partial P(i)/\partial \bar{V}_i$ is 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

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

Similarly, the cross-elasticity e_{ij} of alternative i with respect to an explanatory variable x_{jk} is given by

$$e_{ij} = -\frac{\lambda}{V_j} \frac{\partial P(i)}{\partial \bar{V}_j} \frac{\partial V_j}{\partial x_{jk}} \frac{x_{jk}}{P(i)}$$

where $\partial P(i)/\partial \bar{V}_j$ is 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

$$e_{ij} = \frac{\lambda}{V_i} P(j) \frac{\partial V_j}{\partial x_{jk}} x_{jk}.$$

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 is

$$U^* = \max_{i \in \mathcal{C}} U_i = \max_{i \in \mathcal{C}} V_i \varepsilon_i = \max_{i \in \mathcal{C}} V_i e^{-\frac{\xi_i + c_i}{\lambda}}, \tag{7}$$

where ξ_i is defined by (3). We assume that $(\xi_1, \ldots, \xi_J)$ follows a MEV distribution, that is

$$F(\xi_1, \ldots, \xi_J) = e^{-G(e^{\xi_1}, \ldots, e^{\xi_J})}, \tag{8}$$

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

$$E[U^*] = (G^*)^{-\frac{1}{\sigma\lambda}} \Gamma\left(1 + \frac{1}{\sigma\lambda}\right), \tag{9}$$

where

$$G^* = G(e^{c_1 - \lambda \ln V_1}, \ldots, e^{c_J - \lambda \ln V_J}), \tag{10}$$

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

Compensating variation The compensating variation can be derived in the context where $-V_i$, the negative of the utility 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

$$-P(i)dV_i, \tag{11}$$

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

$$-\int_a^b P(i)dV_i. \tag{12}$$

When $P(i)$ is given by a classical MNL model, this integral leads to the well-known logsum formula (see Small and Rosen, 1981). When $P(i)$ is given by the model with multiplicative error (like (6)), the integral

does not have a closed form in general[2], and numerical integration must be performed.

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 utility functions. The attributes are the cost, in-vehicle time, number of changes, headway, waiting time and access-egress time (ae).

The utility function is defined as

$$\begin{aligned} V_i = \lambda(\; & - \; \text{cost} & & +\beta_1 \text{ ae} & & +\beta_2 \text{ changes} \\ & + \; \beta_3 \text{ headway} & & +\beta_4 \text{ inVehTime} & & +\beta_5 \text{ waiting} \;), \end{aligned} \tag{13}$$

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

$$\begin{aligned} V_i = -\lambda \; \log(\; & \text{cost} & & -\beta_1 \text{ ae} & & -\beta_2 \text{ changes} \\ & - \; \beta_3 \text{ headway} & & -\beta_4 \text{ inVehTime} & & -\beta_5 \text{ waiting}) \; . \end{aligned} \tag{14}$$

The estimation results are reported in Table 6 for the additive specification and in Table 7 for the multiplicative specification. We observe a

[2]Complicated closed form expressions can be derived for (6) with integer values of λ. But λ is estimated and unlikely to be integer.

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 utility is

$$V_i = \lambda(-\text{cost} - e^{\beta_5+\beta_6\xi}Y_i) \tag{15}$$

where

$$Y_i = \text{inVehTime} + e^{\beta_1}\ \text{ae} + e^{\beta_2}\ \text{changes} + e^{\beta_3}\ \text{headway} + e^{\beta_4}\ \text{waiting}, \tag{16}$$

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

$$V_i = -\lambda \log(\text{cost} + e^{\beta_5+\beta_6\xi}Y_i), \tag{17}$$

where Y_i is defined by (16).

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 utility is

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

where Y_i is defined by (16),

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

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

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

The estimation results are reported in Table 10 for the additive specification and in Table 11 for the multiplicative specification. We again obtain

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

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.

4.2 Value of time in Switzerland

We have estimated the models without socio-economics, that is (13), (14), (15) and (17), 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

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

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 systematic parts V_i of the utilities 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 utility of alternatives SM and CAR:

   B_GA_i railwayPass + B_MALE_i male + 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:

   B_SEGMENT_TIME_i = -exp(B_i_TIME + B_GA_i railwayPass + B_MALE_i male + B_PURP_i commuter)

where i={TRAIN,SM,CAR}.

4. Random coefficient (RC): the coefficients for travel time and headway are distributed, with a lognormal 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 specification of the systematic utility, it is more appropriate to assume a multiplicative form. This is particularly relevant when it is desired to allow the scale of the error term to be random with unspecified distribution.

The strategy of taking logs is very natural in this situation. It allows us to derive an equivalent formulation with additive independent error terms. Although this transformation introduces non-linearity into the systematic part of the conditional indirect utility, this can be handled using available software.

A priori it is not possible to know for any given dataset whether the multiplicative formulation will provide a better fit, although we expect the multiplicative specification to perform better in the presence of scale heteroscedasticity. This may happen in particular when data has large variation (e.g. short and long trips in the same model). We have reported some cases where the additive specification is still best. However, in the majority of the cases that 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. Now, selecting the appropriate model depends also on other considerations, including the following.

Consistency with the theory As discussed in the introduction, both the additive and the multiplicative formulations are consistent with RUM.

Tractability In terms of computational and analytical tractability, the additive specification is clearly simpler.

Elasticities, trade-offs They are almost the same for both specification.

Welfare calculus The elegant logsum result of MNL, and the generalization to MEV (McFadden, 1978 and Ben-Akiva and Lerman, 1985), does no carry over to the multiplicative specification.

Our conclusion is that multiplicative model formulation should be part of the toolbox of discrete choice analysts, alongside the techniques that we have for representing observed and unobserved heterogeneity.

A natural extension of the multiplicative approach is to generalize the two specifications (additive and multiplicative) using a Box-Cox transform with parameter $\gamma > 0$, that is a model where

$$\bar{V}_i = -\lambda \frac{(-V_i)^\gamma - 1}{\gamma} + c_i$$

in (4). Indeed, $\gamma = 1$ provides the classical additive error model, $\gamma \to 0$ provides (5) that is, the multiplicative error model, while other values of γ provide new models. Not only do these new models bring an additional level of complexity, but they do also not address the issue of the scale. Consequently, we consider them out of the scope of this paper. Finally, we note that this general formulation can be used to test the nested hypotheses that the model is additive or multiplicative, although care must be exercized with the latter hypothesis as it involves testing against the value $\gamma = 0$ which lies on the boundary of the parameter space.

6 Acknowledgment

The authors like to thank Katrine Hjort for very competent research assistance. 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.

References

Ben-Akiva, M. E. and Lerman, S. R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, Cambridge, Ma.

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.

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.

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. (forthcoming). Using nonparametrics to specify a model to measure the value of travel time, *Transportation Research Part A* .

Fosgerau, M. and Bierlaire, M. (2006). Discrete choice models with multiplicative error terms, *Technical Report TRANSP-OR 060831*, Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne.

Jenkinson, A. F. (1955). Frequency distribution of the anjual 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. (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.

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

McFadden, D. L. (2000). Disaggregate behavioral travel demand's RUM side: a 30-year retrospective, *International Association for Travel Behaviour Conference*, Gold Coast, Queensland.

Small, K. A. and Rosen, H. S. (1981). Applied welfare economics with discrete choice models, *Econometrica* **49**(1): 105–130.

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 Annex: 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	scale (mean)	0.331	0.178	1.86	0.06
6	scale (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	scale (mean)	-0.252	0.106	-2.38	0.02
6	scale (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	scale (mean)	0.341	0.170	2.01	0.04
10	scale (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	scale (mean)	-0.150	0.111	-1.34	0.18
10	scale (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

B Annex: 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

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

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

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

Table 13: Model with fixed parameters and multiplicative error terms

Variable number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	scale (mean)	-0.763	0.111	-6.86	0.00
2	scale (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

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

Table 14: Model with unobserved heterogeneity — additive error terms

Variable number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	scale (mean)	-0.956	0.119	-8.04	0.00
2	scale (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

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

Table 15: Model with unobserved heterogeneity — multiplicative error terms

C Derivation of the expected maximum utility

From (7), the maximum utility is

$$U^* = \max_{i \in \mathcal{C}} V_i e^{-\frac{\xi_i + c_i}{\lambda}}, \tag{18}$$

where ξ_i is defined by (3). Note that $U^* \leq 0$. We assume that $(\xi_1, \ldots, \xi_J)$ follows a MEV distribution (8). The CDF of U^* is obtained as follows, for $t \leq 0$:

$$\begin{aligned}
F(t) &= \Pr(U^* \leq t) = \Pr(U_i \leq t, \ \ \forall i) \\
&= \Pr(\xi_i \leq -\lambda \ln(tV_i^{-1}) - c_i, \ \ \forall i) \\
&= \exp(-G((tV_1^{-1})^\lambda e^{c_1}, \ldots, (tV_J^{-1})^\lambda e^{c_J}) \\
&= \exp(-t^{\sigma\lambda} G(e^{c_1 - \lambda \ln V_1}, \ldots, e^{c_1 - \lambda \ln V_J}) \\
&= \exp(-t^{\sigma\lambda} G^*)
\end{aligned}$$

using the σ-homogeneity of G and the definition (10) of G^*. The CDF can be inverted as

$$F^{-1}(x) = \left(-\frac{\ln x}{G^*}\right)^{\frac{1}{\sigma\lambda}} = (G^*)^{-\frac{1}{\sigma\lambda}} \left(\ln\left(\frac{1}{x}\right)\right)^{\frac{1}{\sigma\lambda}}. \tag{19}$$

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 (9).