Discrete choice models with multiplicative error terms

M. Fosgerau* M. Bierlaire†

August 31, 2006

Report TRANSP-OR 060831
Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne
transp-or.epfl.ch

*Danish Transport Research Institute. Email: mf@dtf.dk

†Ecole Polytechnique Fédérale de Lausanne, Inst. of Mathematics, CH-1015 Lausanne, Switzerland. Email: michel.bierlaire@epfl.ch

Abstract

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.

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 circumvent 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, generalized extreme value (GEV) models and mixtures of MNL and GEV 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 GEV models assumes that 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.

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 RUM. In, for example, McFadden (2000), it is described 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 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.

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 interpret 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.

In previous work on the Danish value-of-time survey (Fosgerau, 2005; Fosgerau, 2006), 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 3. We conclude the paper with some remarks in Section 4.

2 Methodology

Assume a general multiplicative utility function over a finite set $\mathcal{C}$ of 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, λ is the scale, and ξ_i are random variables with a fixed mean and scale, and we obtain

$$P(i|\mathcal{C}) = \Pr(-\lambda \ln(-V_i) + c_i + \xi_i \geq -\lambda \ln(-V_j) + c_j + \xi_j, j \in \mathcal{C}), \tag{4}$$

which is now a classical random utility model with additive error.

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, GEV and mixtures of GEV models. 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 3.

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 GEV 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 GEV models, with nonlinear utility functions.

We conclude this section by deriving a nice property of the multiplicative error term 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.

3 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.

3.1 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 for the additive specification is defined as

$$
\begin{aligned}
V_i = \lambda(\quad & - \quad \text{cost} \quad && +\beta_1 \text{ ae} && +\beta_2 \text{ changes} \\
& + \quad \beta_3 \text{ headway} && +\beta_4 \text{ inVehTime} && +\beta_5 \text{ waiting} \quad),
\end{aligned} \tag{5}
$$

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

$$
\begin{aligned}
V_i = -\lambda \ \log(\quad & \quad \text{cost} \quad && -\beta_1 \text{ ae} && -\beta_2 \text{ changes} \\
& - \quad \beta_3 \text{ headway} && -\beta_4 \text{ inVehTime} && -\beta_5 \text{ waiting}) \quad .
\end{aligned} \tag{6}
$$

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 utility for the additive model is

$$
V_i = \lambda(-\text{cost} - e^{\beta_5 + \beta_6 \xi} Y_i) \tag{7}
$$

where

$$
Y_i = \text{inVehTime} + e^{\beta_1} \text{ ae} + e^{\beta_2} \text{ changes} + e^{\beta_3} \text{ headway} + e^{\beta_4} \text{ waiting}, \tag{8}
$$

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

$$
V_i = -\lambda \log(\text{cost} + e^{\beta_5 + \beta_6 \xi} Y_i), \tag{9}
$$

where Y_i is defined by (8).

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).

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

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

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

where Y_i is defined by (8),

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

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 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:	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

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

3.2 Value of time in Switzerland

We have estimated the models without socio-economics, that is (5), (6), (7) and (9), 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

3.3 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:

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

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:

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

where i={TRAIN,SM,CAR}.

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

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.

4 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

	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

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. This depends both on the data and on the specification of the systematic utility. 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.

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

5 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.

References

Bhat, C. R. (1997). Covariance heterogeneity in nested logit models: econometric structure and application to intercity travel, *Transportation Research Part B* **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* **39**(7): 621–640.

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. (2005). Specification of a model to measure the value of travel time savings, *European Transport Conference*.

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

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* **39**(9): 825–853.

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

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

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.

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

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