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

Mogens Fosgerau and Michel Bierlaire

Technical University of Denmark

Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne





Introduction

• Random utility models:

$$P(i|\mathcal{C}) = \Pr(U_i \ge U_j \; \forall j \in \mathcal{C})$$

=
$$\Pr(\mu V_i + \varepsilon_i \ge \mu V_j + \varepsilon_j \; \forall j \in \mathcal{C})$$

- ε_i i.i.d. across individuals, so the scale is normalized.
- As a consequence, the scale is confounded with the parameters of V_i .
- The scale is directly linked with the variance of U_i





Introduction

- The scale may vary from one individual to the next
- The scale may vary from one choice context to the next
 SP/RP data
- Linear-in-parameter: $V_i = \mu \beta' x_i$
- Even if β is fixed, $\mu\beta$ is distributed





Introduction

Proposed solutions:

- Deterministically identify groups and estimate different scale parameters (introduces non linearities)
- Assume 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)





Multiplicative error

Our proposal:

• RUM with multiplicative error

$$U_i = \mu V_i \varepsilon_i.$$

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 .





Multiplicative error

- ε_i are i.i.d. across individuals
- Potential heteroscedasticity is captured by the individual specific scale μ .
- Sign restriction on V_i : natural if, for instance, generalized cost





The scale disappears

$$P(i|\mathcal{C}) = \Pr(U_i \ge U_j, j \in \mathcal{C})$$

=
$$\Pr(\mu V_i \varepsilon_i \ge \mu V_j \varepsilon_j, j \in \mathcal{C})$$

=
$$\Pr(V_i \varepsilon_i \ge V_j \varepsilon_j, j \in \mathcal{C}),$$

Taking logs

$$P(i|\mathcal{C}) = \Pr(V_i \varepsilon_i \ge V_j \varepsilon_j, j \in \mathcal{C})$$

= $\Pr(-V_i \varepsilon_i \le -V_j \varepsilon_j, j \in \mathcal{C})$
= $\Pr(\ln(-V_i) + \ln(\varepsilon_i) \le \ln(-V_j) + \ln(\varepsilon_j), j \in \mathcal{C})$
= $\Pr(-\ln(-V_i) - \ln(\varepsilon_i) \ge -\ln(-V_j) - \ln(\varepsilon_j), j \in \mathcal{C}).$





We define

$$-\ln(\varepsilon_i) = (c_i + \xi_i)/\lambda,$$

where

- c_i is the intercept,
- λ is the scale, constant across the population, as a consequence of the i.i.d. assumption on ε_i
- ξ_i are random variables with a fixed mean and scale





•
$$P(i|\mathcal{C}) =$$

$$\Pr(-\lambda \ln(-V_i) + c_i + \xi_i \ge -\lambda \ln(-V_j) + c_j + \xi_j, j \in \mathcal{C}),$$

which is now a classical RUM with additive error.

- Important: contrarily to μ , the scale λ is constant across the population
- V_i must be normalized for the model to be identified. Indeed, for any $\alpha > 0$,

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





- When V_i is linear-in-parameters, it is sufficient to fix one parameter to either 1 or -1.
- e.g. normalize the cost coefficient to 1. Others become willingness-to-pay indicators.





Choice probability: MNL

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

where

• e^{c_j} are constants to be estimated





If ξ_i is extreme value distributed, the CDF of ε_i is a generalization of an exponential distribution

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





Properties: elasticities

Define

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

Then

$$e_{i} = \frac{\partial P(i)}{\partial \bar{V}_{i}} \frac{\partial \bar{V}_{i}}{\partial V_{i}} \frac{\partial V_{i}}{\partial x_{ik}} \frac{x_{ik}}{P(i)} = -\frac{\lambda}{V_{i}} \frac{\partial P(i)}{\partial \bar{V}_{i}} \frac{\partial V_{i}}{\partial x_{ik}} \frac{x_{ik}}{P(i)}$$

where $\partial P(i)/\partial \bar{V}_i$ is derived from the corresponding additive model. For MNL:

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

and

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





In the paper (see transp-or.epfl.ch)

- Trade-offs: the same
- Expected Maximum Utility: derivation for MEV models
- Compensating variation when $-V_i$ is a generalized cost

$$-\int_{a}^{b}P(i)dV_{i}$$

- not as simple as the logsum
- integral with no closed form





Discussion

- Fairly general specification
- Free to make assumptions about ξ_i
- Parameters inside V_i can be random
- We may obtain MNL, GEV and mixtures of GEV models.
- 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 (examples in the paper).





Discussion

- If random parameters are involved, one must ensure that $P(V_i \ge 0) = 0$.
- How? The sign of a parameter can be restricted using, e.g., an exponential.
- For deterministic parameters: bounds constraints
- Maximum likelihood estimation is complicated in the general case.
- Taking logs provides an equivalent specification with additive independent error terms





Discussion

- Classical softwares can be used
- However, even when the Vs are linear in the parameters, the equivalent additive specification is nonlinear.
- OK with Biogeme





Case study: value of time in Denmark

- Danish value-of-time study
- SP data
- involves several attributes in addition to travel time and cost





Model 1: Additive specification

$$V_i = \lambda(- \cos t + \beta_1 ae + \beta_2 changes + \beta_3 headway + \beta_4 inVehTime + \beta_5 waiting),$$

Model 1: Multiplicative specification

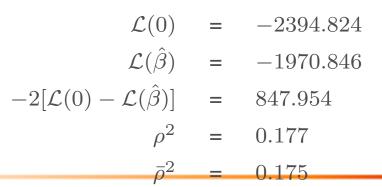
 $V_i = -\lambda \log(\cos t -\beta_1 \operatorname{ae} -\beta_2 \operatorname{changes} -\beta_3 \operatorname{headway} -\beta_4 \operatorname{inVehTime} -\beta_5 \operatorname{waiting})$





				Robust		
	Variable		Coeff.	Asympt.		
	number	Description	estimate	std. error	t-stat	<i>p</i> -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

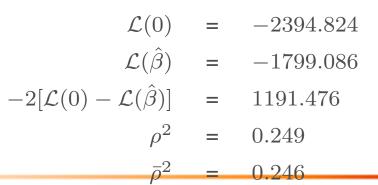




ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

			Robust		
Variable		Coeff.	Asympt.		
number	Description	estimate	std. error	t-stat	<i>p</i> -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







Model 1: result

- Same number of parameters
- Significant improvement of the fit: 171.76, from -1970.846 to -1799.086





Model 2: taste heterogeneity

• Additive specification:

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

where

• $Y_i =$

inVehTime $+e^{\beta_1}$ ae $+e^{\beta_2}$ changes $+e^{\beta_3}$ headway $+e^{\beta_4}$ waiting

- $\xi \sim N(0,1)$
- Multiplicative specification

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





				Robust		
Variab	ole		Coeff.	Asympt.		
numb	er	Description	estimate	std. error	t-stat	<i>p</i> -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
			Num	ber of obser	vations =	= 3455
			Num	ber of individ	duals = 5	23
			Num	ber of draws	for SML	E = 1000
			$\mathcal{L}(0)$) = -23	394.824	
	TRANSP-OR		$\mathcal{L}(\hat{eta})$) = -19	925.467	
TRAILS	I KANSI - UK			2 = 0.19	3 Worksho	p on Discrete Choice



				Robust		
	Variable		Coeff.	Asympt.		
	number	Description	estimate	std. error	t-stat	<i>p</i> -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
-			Nu	mber of obse	ervations =	- 3455
			Nu	mber of indiv	riduals = 5	23
			Nu	mber of draw	s for SML	E = 1000
			$\mathcal{L}(0)$	(0) = -2	2394.824	
5	TRANSP-0	IR	$\mathcal{L}(\mu$	$\hat{\beta}) = -1$	1700.060	
	FRANSI-L		Ā	$\bar{p}^2 = 0.2$	287 Workshop c	on Discrete Choice N



Model 2: result

- Same number of parameters
- Significant improvement of the fit: 225.764, from -1925.824 to -1700.060





Observed and unobs. heterogeneity

• Additive specification

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

where

• *Y_i* is defined as before

•
$$W_i =$$

 $\beta_5 \text{ highInc} + \beta_6 \log(\text{inc}) + \beta_7 \log(\text{Inc}) + \beta_8 \min(\beta_9 + \beta_{10}\xi)$

•
$$\xi \sim N(0,1)$$
.





Observed and unobs. heterogeneity

• Multiplicative specification:

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

Results:

- Again large improvement of the fit with the same number of parameters
- Additive: -1914.180
- Multiplicative: -1675.412
- Difference: 238.777





Summary: train data set

	3455	
	523	
Model	Additive	Difference
1	-1970.85	171.76
2	-1925.824	225.764
3	-1914.12	239.45





Summary: bus data set

	7751	
	1148	
Model	Additive	Difference
1	-4255.55	297.2
2	-4134.56	317.07
3	-4124.21	319.31





Workshop on Discrete Choice Models - EPFL - August 2007 – p.30/35

	8589	
	1585	
Model	Additive	Difference
1	-5070.42	766.41
2	-4667.05	858.83
3	-4620.56	858.99





Swiss value of time (SP)

- No improvement with fixed parameters
- Small improvement for random parameters

	Additive	Multiplicative	Diff.
Fixed param.	-1668.070	-1676.032	-7.96
Random param.	-1595.092	-1568.607	26.49

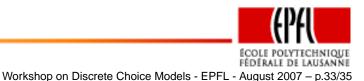




Swissmetro (SP)

- Nested logit
- 16 variants of the model
 - Alternative Specific Socio-economic Characteristics (ASSEC)
 - Error component (EC)
 - Segmented travel time coefficient (STTC)
 - Random coefficient (RC): the coefficients for travel time and headway are distributed, with a lognormal distribution.



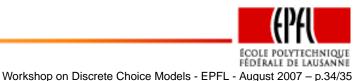


	RC	EC	STTC	ASSEC	Additive	Multiplicative	Difference
1	0	0	0	0	-5188.6	-4988.6	200.0
2	0	0	0	1	-4839.5	-4796.6	42.9
3	0	0	1	0	-4761.8	-4745.8	16.0
4	0	1	0	0	-3851.6	-3599.8	251.8
5	1	0	0	0	-3627.2	-3614.4	12.8
6	0	0	1	1	-4700.1	-4715.5	-15.4
7	0	1	0	1	-3688.5	-3532.6	155.9
8	0	1	1	0	-3574.8	-3872.1	-297.3
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

Concluding remarks

- Error term does not have to be additive
- With multiplicative errors, an equivalent additive formulation can be derived by taking logs
- Multiplicative is not systematically superior
- In our experiments, it outperforms additive spec. in the majority of the cases
- In quite a few cases, the improvement is very large, sometimes even larger than the improvement gained from allowing for unobserved heterogeneity.





Concluding remarks

• Model with multiplicative error terms should be part of the toolbox of discrete choice analysts

Thank you!



