Specification of the cross nested logit model with sampling of alternatives for route choice models

Xinjun Lai Michel Bierlaire

Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne

June 19, 2014

Outline

Motivation

Route choice model

- Given an origin and a destination
- what is the preferred itinerary of a given traveler?

Difficulties

- Data
- Very large choice set
- Structural correlation among alternatives

Data

Revealed preferences

- Usually GPS data
- Unavailability of socio-economic variables

Stated preferences

- Hypothetical paths
- Simplified paths

In this paper...

GPS data

Very large choice set

Issue

Number of paths grows exponentially with the number of nodes

Literature

- link elimination Azevedo et al. (1993)
- link penalty de la Barra et al. (1993)
- labeled paths Ben-Akiva et al. (1984)
- SP on random costs Ramming (2002), Bovy and Fiorenzo-Catalano (2006)
- Sampling Frejinger et al. (2009)

Structural correlation

Issue

Significant physical overlap

Literature

- C-logit Cascetta et al. (1996)
- Path-size Ben-Akiva and Bierlaire (1999)
- Link-based cross-nested logit Prashker and Bekhor (1999)
- Error components Ramming (2002); Frejinger and Bierlaire (2007)

In this paper...

Methodology

- Cross Nested logit
- Sampling of alternatives

Builds on...

- McFadden (1978)
- Vovsha and Bekhor (1998)
- Bierlaire et al. (2008)
- Frejinger et al. (2009)
- Guevara and Ben-Akiva (2013)
- Flötteröd and Bierlaire (2013)

Outline

Logit model

$$P(i|\mathcal{C}) = \frac{e^{V_i}}{\sum_{j\in\mathcal{C}} e^{V_j}}$$

McFadden (1978)

Sampling protocol

- Sample subset $\mathcal{D} \subseteq \mathcal{C}$
- Sampling probability $q(\mathcal{D}|j)$
- Positive conditioning property

$$q(\mathcal{D}|i) > 0 \implies q(\mathcal{D}|j) > 0 \ \forall j \in \mathcal{D}.$$

Logit model

$$P(i|\mathcal{C}) \approx P(i|\mathcal{D}) = \frac{e^{V_i + \ln q(\mathcal{D}|i)}}{\sum_{j \in \mathcal{D}} e^{V_j + \ln q(\mathcal{D}|j)}}$$

Simple random sampling

- $q(\mathcal{D}|i) = q(\mathcal{D}|j)\ \forall i, j \in \mathcal{C}$
- Correction terms cancel out
- Irrelevant, circuitous paths
- How to draw?

Importance sampling

- $\ln q(\mathcal{D}|i)$ are confounded with ASC
- How to draw?

How to draw?

Shortest path-based procedures

- link elimination: deterministic
- link penalty: deterministic
- labeled paths: deterministic
- SP on random costs:
  - some paths have 0 probability to be drawn
  - how to compute the sampling probability?

Metropolis-Hastings algorithm

Flötteröd and Bierlaire (2013)

Features

- Designed to draw from complex distributions
- Does not require the exact pmf/pdf
- Only a quantity proportional to it.
- For instance, to draw a path p with probability

$$\frac{b_p}{\sum_{q \in \mathcal{C}} b_q}$$

only b_p are needed.

Metropolis-Hastings algorithm

Methodology

- Design a Markov chain Q visiting the states/paths
- Accept/reject method
- Accept probability depends on
  - target (unnormalized) probabilities
  - transition probabilities of the Markov chain:

$$P(\text{accept}) = \min\left(\frac{b_q Q_{qp}}{b_p Q_{pq}}, 1\right)$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Example

$$b = (20, 8, 3, 1) \quad \pi = (\frac{5}{8}, \frac{1}{4}, \frac{3}{32}, \frac{1}{32})$$

$$Q = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$$

Run MH for 10000 iterations. Collect statistics after 1000

- Accept: [2488, 1532, 801, 283]
- Reject: [0, 952, 1705, 2239]
- Simulated: [0.627, 0.250, 0.095, 0.028]
- Target: [0.625, 0.250, 0.09375, 0.03125]

Sampling of paths

Difficulties

Design Q such that

- Every path can be generated with nonzero probability
- Both Q_{pq} and Q_{qp} are known

Flötteröd and Bierlaire (2013)

- Proof of concept on synthetic data
- Application to Tel Aviv (17K links, 8K nodes)

Outline

MEV models

Generic model

$$P(i|\mathcal{C}) = \frac{\exp(V_i + \ln G_i(\mathcal{C}))}{\sum_{j\in\mathcal{C}} \exp(V_j + \ln G_j(\mathcal{C}))}$$

where $G_i(\mathcal{C}) = G_i(e^{V_1}, \ldots, e^{V_J})$ is the derivative of the CPGF wrt e^{V_i}.

Cross nested logit

$$G_i(\mathcal{C}) = \sum_{m=1}^{M} \left[\mu \alpha_{im} e^{V_i(\mu_m - 1)} \left(\sum_{j\in\mathcal{C}} \alpha_{jm} e^{\mu_m V_j} \right)^{\frac{\mu - \mu_m}{\mu_m}} \right],$$

MEV models

Generic model

$$P(i|\mathcal{C}) = \frac{\exp(V_i + \ln G_i(\mathcal{C}))}{\sum_{j\in\mathcal{C}} \exp(V_j + \ln G_j(\mathcal{C}))}$$

where $G_i(\mathcal{C}) = G_i(e^{V_{1n}}, \ldots, e^{V_j})$ is the derivative of the CPGF wrt e^{V_i}.

Cross nested logit

$$G_i(\mathcal{C}) = \sum_{m=1}^{M} \left[\mu\alpha_{im} e^{V_i(\mu_m - 1)} \left(\sum_{j\in\mathcal{C}} \alpha_{jm} e^{\mu_m V_j} \right)^{\frac{\mu-\mu_m}{\mu_m}} \right],$$

Sampling and MEV

$$P(i|\mathcal{C}) = \frac{\exp(V_i + \ln G_i(\mathcal{C}))}{\sum_{j\in\mathcal{C}} \exp(V_j + \ln G_j(\mathcal{C}))}$$

Sampling correction

Bierlaire et al. (2008)

- If $\ln G_j(\mathcal{C})$ is known, same idea as for logit

$$\Pr(i|\mathcal{D}) = \frac{\exp(V_i + \ln G_i(\mathcal{C}) + \ln \Pr(\mathcal{D}|i))}{\sum_{j\in\mathcal{D}} \exp(V_j + \ln G_j(\mathcal{C}) + \ln \Pr(\mathcal{D}|j))}.$$

- Not counfounded with the constants anymore.

Sampling and MEV

Correction term

$$\Pr(\mathcal{D}|p) \propto \frac{k_p}{q(p)}$$

where

- k_p is the number of times path p has been generated
- $q(p)$ is the sampling probability of path p
- $q(p) \propto b_p$

Model I

$$\Pr(i|\mathcal{D}) = \frac{\exp(V_i + \ln G_i(\mathcal{C}) + \ln \frac{k_i}{b_i})}{\sum_{j\in\mathcal{D}} \exp(V_j + \ln G_j(\mathcal{C}) + \ln \frac{k_j}{b_j})},$$

Approcximation of ln $G_i(\mathcal{C})$

Guevara and Ben-Akiva (2013)

$$G_i(\mathcal{C}) \approx \widehat{G}_i(D, w) = \sum_{m=1}^{M} \left[\mu \alpha_{im} e^{V_i(\mu_m - 1)} \left(\sum_{j \in \mathcal{D}} w_j \alpha_{jm} e^{\mu_m V_j} \right)^{\frac{\mu - \mu_m}{\mu_m}} \right]$$

where w_j expansion factor to be defined.

Expansion factors: Guevara and Ben-Akiva (2013)

Realized / expected

$$w_j^G = \frac{k_j}{\mathrm{E}[k_j]} = \frac{k_j}{q(j)R} = \frac{k_j B}{b(j)R}$$

where

- R is the number of draws used to generate $\mathcal{D}$
- $B = \sum_{j \in \mathcal{C}} b(j)$ [Requires enumeration of $\mathcal{C}$]

Approximate B

$$B = \sum_{j \in \mathcal{C}} b(j) = |\mathcal{C}| \frac{\sum_{i \in \mathcal{C}} b(i)}{|\mathcal{C}|} = |\mathcal{C}| \bar{b},$$

and

$$\bar{b} = \frac{\sum_{i \in \mathcal{C}} b(i)}{|\mathcal{C}|} \approx \frac{\sum_{i \in \mathcal{D}} b(i)}{|\mathcal{D}|}.$$

Expansion factors: Guevara and Ben-Akiva (2013)

Approximation

$$w_j^G = \frac{k_j}{b(j)R}\frac{|\mathcal{C}|}{|\mathcal{D}|}\sum_{i\in\mathcal{D}} b(i)$$

which require $|\mathcal{C}|$

Approximate $|\mathcal{C}|$

Roberts and Kroese (2007)

N random walks in the network

$$|\mathcal{C}| \approx \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\ell(i)}.$$

$\ell^{(i)}$: likelihood of the path generated by the algorithm during run i

Expansion factors: Frejinger et al. (2009)

Account for the upper bound

$$w_j^F = \begin{cases} 1 & \text{if } b(j)R > B, \\ \frac{B}{b(j)R} & \text{otherwise.} \end{cases}$$

Same approximation of B

$$B \approx \frac{|\mathcal{C}|}{|\mathcal{D}|} \sum_{i \in \mathcal{D}} b(i)$$

Again, requires $|\mathcal{C}|$

Expansion factors: Lai and Bierlaire (2014)

Avoiding $|\mathcal{C}|$

- Let s be the path which has been sampled the most in $\mathcal{D}$
- $k_s \geq k_p$, for each $p \in \mathcal{D}$.
- If sample is large enough, $k_s \approx q(s)R$

$$w_j^G = \frac{k_j}{q(j)R} \approx w_j^L = \frac{k_j}{q(j)R}\frac{q(s)R}{k_s} = \frac{k_j}{b(j)}\frac{b(s)}{k_s}$$

which does not require B or $|\mathcal{C}|$.

Expansion factors

- Guevara and Ben-Akiva (2013)

$$w_j^G = \frac{k_j}{b(j)R} B \text{ with } B \approx \frac{|\mathcal{C}|}{|\mathcal{D}|} \sum_{i \in \mathcal{D}} b(i)$$

- Frejinger et al. (2009)

$$w_j^F = \begin{cases} 1 & \text{if } b(j)R > B, \\ \frac{B}{b(j)R} & \text{otherwise.} \end{cases} \quad \text{with } B \approx \frac{|\mathcal{C}|}{|\mathcal{D}|} \sum_{i \in \mathcal{D}} b(i).$$

- Lai and Bierlaire (2014)

$$w_j^L = \frac{k_j}{b(j)} \frac{b(s)}{k_s}$$

Models to be compared

- Model I: true G_i (impossible in practice)

$$\Pr(i|\mathcal{D}) = \frac{\exp(V_i + \ln G_i(\mathcal{C}) + \ln \frac{k_i}{b(i)})}{\sum_{j\in\mathcal{D}} \exp(V_j + \ln G_j(\mathcal{C}) + \ln \frac{k_j}{b(j)})}$$

- Model II: the proposed model

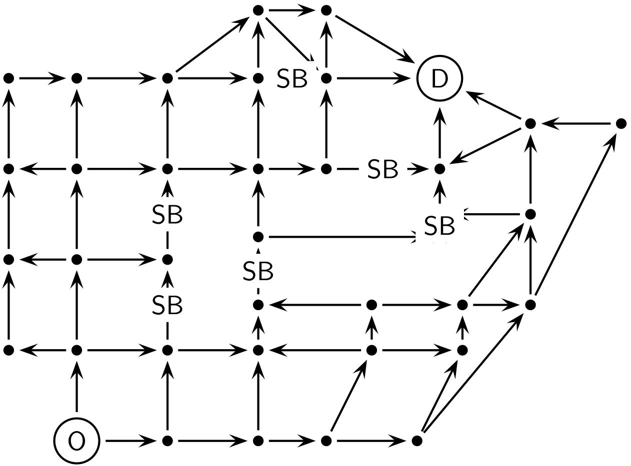
$$\Pr(i|\mathcal{D}, \mathcal{D}', w) = \frac{\exp(V_i + \ln G_i(\mathcal{D}', w)) + \ln \frac{k_i}{b(i)})}{\sum_{j\in\mathcal{D}} \exp(V_j + \ln G_j(\mathcal{D}', w) + \ln \frac{k_j}{b(j)})}.$$

- Model III: no expansion factor, no sampling correction (benchmark)

$$\Pr(i|\mathcal{D}, \mathcal{D}') = \frac{\exp(V_i + \ln G_i(\mathcal{D}', 1))}{\sum_{j\in\mathcal{D}} \exp(V_j + \ln G_j(\mathcal{D}', 1))},$$

Outline

The network: 170 paths (Frejinger (2008))



The true model: cross-nested logit

Utility

$$V_i = \beta_L L_i + \beta_{\mathrm{SB}} \mathrm{SB}_i,$$

"True" parameters

- $\beta_L = -0.5$ and $\beta_{\mathrm{SB}} = -0.1$
- $\mu_m = 1.5$ for each link m
- $\alpha_{im} = \ell_m / L_i$

Data

3000 synthetic choices

Re-estimate the parameters of the true model

Full choice set

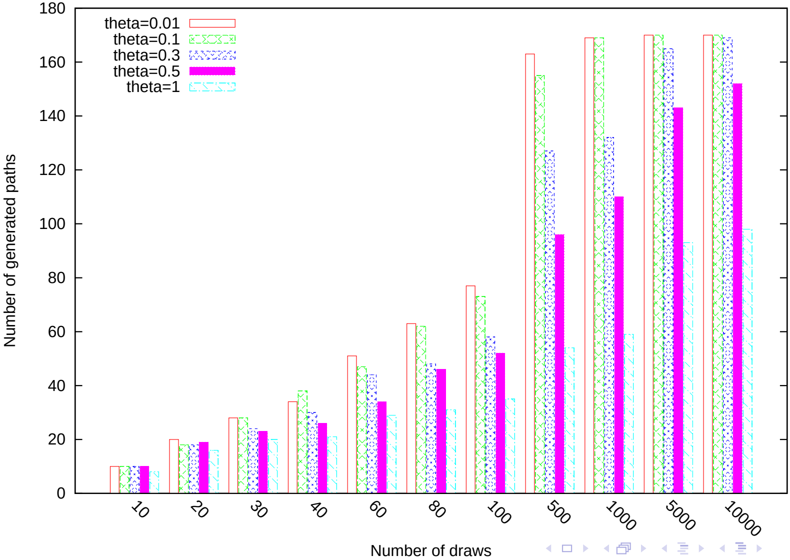
Parameters	Est.	Std err.	t-test (0)	t-test (true)
β_L	-0.501	0.0118	43.1	0.678
β_{SB}	-0.0910	0.0240	3.19	0.375
μ_m	1.49	0.0269	55.2	0.0535

Sampling paths

Metropolis-Hastings

$$b(i) = \exp(-\theta L_i), \quad \theta \geq 0$$

Number of generated paths



Model I: true G_i — MH $\theta = 0.5$

10 draws	Est.	Std err.	t-test(0)	t-test(true)
β_L (-0.5)	-0.443	0.0163	27.3	3.48
β_{SB} (-0.1)	-0.0647	0.0427	1.51	0.826
μ_m (1.5)	1.56	0.0340	45.8	1.72
Estimation time: 1362 seconds				
40 draws	**Est.**	**Std err.**	**t-test(0)**	**t-test(true)**
β_L (-0.5)	-0.479	0.0156	30.8	1.34
β_{SB} (-0.1)	-0.0720	0.0393	1.83	0.713
μ_m (1.5)	1.51	0.0322	47.0	0.367
Estimation time: 4648 seconds				

Model I: true G_i — MH $\theta = 0.01$

10 draws	Est.	Std err.	t-test(0)	t-test(true)
β_L (-0.5)	-0.535	0.0174	30.8	2.01
β_{SB} (-0.1)	-0.132	0.0545	2.42	0.580
μ_m (1.5)	1.41	0.0355	39.8	2.47
Estimation time: 1612 seconds				
40 draws	**Est.**	**Std err.**	**t-test(0)**	**t-test(true)**
β_L (-0.5)	-0.544	0.0160	33.9	2.76
β_{SB} (-0.1)	-0.130	0.0410	3.16	0.726
μ_m (1.5)	1.41	0.0322	43.8	2.85
Estimation time: 4914 seconds				

Model I: comments

- Trade-off between dispersion (low θ) and number of draws
- Lower value of θ requires more draws
- $\theta = 0.5$, 40 draws: parameters are correctly estimated
- First sampling scheme is validated
- No specific guideline for θ and R

Approximeting $\bar{b}$ and $|\mathcal{C}|$

Protocol

- For $\bar{b}$: generate $\mathcal{D}$ using MH with 100 draws and $\theta = 0.01$
- For $|\mathcal{C}|$: generate 10000 paths using random walk
- Repeat 100 times
- Compute the empirical mean and standard error

Results

	True	Mean	Std err	t-test(true)
$\bar{b}$	0.688	0.684	0.0023	1.62
$\lvert\mathcal{C}\rvert$	170	169.8	2.52	0.0722

Model II

Protocol

- Denominator: $\mathcal{D}$ generated with MH (40 draws, $\theta = 0.5$)
- Expansion factor: $\mathcal{D}'$ MH with various values

Model II: 100 draws (t-test vs true value)

Sampling protocol for $\mathcal{D}'$: $\theta = 0.5$					
	Mod. II				Mod. III
	w^G	w^F	w^L	$w = 1$	
β_L	2.48	4.34	**1.25**	3.59	19.4
β_{SB}	**0.910**	**0.867**	**0.722**	**0.179**	**0.221**
μ_m	2.02	3.09	**0.437**	2.98	**1.06**
Sampling protocol for $\mathcal{D}'$: $\theta = 0.01$					
	Mod. II				Mod. III
	w^G	w^F	w^L	$w = 1$	
β_L	4.61	4.23	4.48	4.30	18.9
β_{SB}	**0.303**	**0.297**	**0.254**	**0.467**	**0.634**
μ_m	4.70	4.71	5.38	4.55	3.63

Model II: 200 draws (t-test vs true value)

Sampling protocol for $\mathcal{D}'$: $\theta = 0.5$					
	Mod. II				Mod. III
	w^G	w^F	w^L	$w = 1$	
β_L	**0.578**	10.5	**0.0374**	3.38	18.9
β_{SB}	**0.513**	**0.194**	**0.440**	**0.259**	**0.269**
μ_m	**1.36**	5.02	**1.34**	3.07	**0.965**
Sampling protocol for $\mathcal{D}'$: $\theta = 0.01$					
	Mod. II				Mod. III
	w^G	w^F	w^L	$w = 1$	
β_L	3.51	3.84	2.86	4.37	18.5
β_{SB}	**0.173**	**0.119**	**0.298**	**0.409**	**0.571**
μ_m	9.11	8.65	7.19	5.41	3.72

Model II: 300 draws (t-test vs true value)

Sampling protocol for $\mathcal{D}'$: $\theta = 0.5$					
	Mod. II				Mod. III
	w^G	w^F	w^L	$w = 1$	
β_L	**0.981**	3.62	**0.703**	**0.981**	19.3
β_{SB}	**0.428**	**1.34**	**0.537**	**0.428**	**0.0052**
μ_m	2.28	3.12	**1.70**	2.28	**1.66**
Sampling protocol for $\mathcal{D}'$: $\theta = 0.01$					
	Mod. II				Mod. III
	w^G	w^F	w^L	$w = 1$	
β_L	**0.809**	**0.0271**	**1.02**	5.05	18.5
β_{SB}	**0.565**	**0.780**	**0.480**	**0.564**	**0.654**
μ_m	**1.66**	**0.650**	**1.84**	5.19	3.01

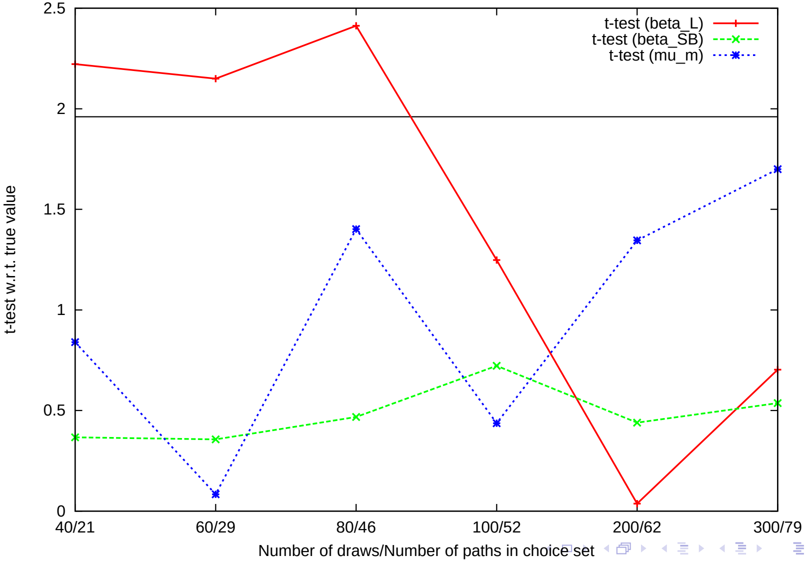
TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Comments

- $\theta = 0.5$ seems again the most appropriate
- Model II outperforms Model III (no correction, no expansion factor)
- New expansion factor is the most appropriate (already good resuts with 100 draws)
- μ_m seems to be the most sensitive parameters

t-tests with w^L and $\theta = 0.5$



Outline

Tianhe region (CBD) of Guangzhou (China)

Data

Network

- 208 nodes
- 662 links
- 24 major roads
- 34 arterial streets
- 32 minor streets
- 57 signalized intersections

GPS traces from taxis

- 7 ODs
- 740 trips

Model

Utility

$$V_i = \beta_L \text{Length}_i + \beta_{\text{ARR}} \text{ArteryRoadRatio}_i + \beta_S \text{Signal}_i.$$

Cross-nested logit

- Two nests: μ: non-artery roads, μ_{mA}: artery roads
- $\alpha_{im} = \ell_m / L_i$

MH sampling

θ	$\lvert\mathcal{D}\rvert$	θ	$\lvert\mathcal{D}\rvert$
0.005	29	0.0025	3813
0.004	54	0.0023	5624
0.003	201	0.002	7766
0.0028	2036	0.001	9836

Estimation results (with Matlab, Intel i5 with 4GB RAM, one processor)

$\theta = 0.003$			
	Model II		
	Est.	Std. err.	t-test (0)
β_L	-1.58	0.0566	27.9
β_{ARR}	8.09	0.636	12.7
β_S	-0.513	0.267	1.91
μ_m	3.90	0.117	33.3
μ_{mA}	2.22	0.257	8.62
Number of observations	740 trips from 7 OD		
Null log likelihood	-3.4078e+03		
Final log likelihood	-1.9206e+03		
Estimation time	22.32 hours		

Conclusion

Contributions

- Application of sampling of alternative for MEV and route choice
- New expansion factor
- Validity check: synthetic data
- Feasibility check: real data
- Heavy, but tractable

Future work

- Investigate other nesting structures
- Different ways to approximate G_i
- Estimation of α_{im} (?)

Bibliography I

Azevedo, J., Costa, M. S., Madeira, J. S., and Martins, E. V. (1993). An algorithm for the ranking of shortest paths. 69(1):97–106.

Ben-Akiva, M., Bergman, M., Daly, A., and Ramaswamy, R. (1984). Modeling inter urban route choice behaviour. In Vollmuller, J. and Hamerslag, R., editors, *Proceedings of the 9th International Symposium on Transportation and Traffic Theory*, pages 299–330. VNU Science Press, Utrecht, Netherlands.

Ben-Akiva, M. and Bierlaire, M. (1999). Discrete choice methods and their applications to short-term travel decisions. In Hall, R., editor, *Handbook of Transportation Science*, Operations Research and Management Science, pages 5–34. Kluwer. ISBN:0-7923-8587-X.

Bierlaire, M., Bolduc, D., and McFadden, D. (2008). The estimation of generalized extreme value models from choice-based samples. *Transportation Research Part B: Methodological*, 42(4):381–394.

Bibliography II

Bovy, P. H. L. and Fiorenzo-Catalano, S. (2006). Stochastic route choice set generation: behavioral and probabilistic foundations. In *Proceedings of the 11th International Conference on Travel Behaviour Research*, Kyoto, Japan.

Cascetta, E., Nuzzolo, A., Russo, F., and Vitetta, A. (1996). A modified logit route choice model overcoming path overlapping problems: Specification and some calibration results for interurban networks. In *Proceedings of the 13th International Symposium on Transportation and Traffic Theory*, pages 697–711. Pergamon.

de la Barra, T., Pérez, B., and Añez, J. (1993). Mutidimensional path search and assignment. In *Proceedings of the 21st PTRC Summer Meeting*, pages 307–319.

Flötteröd, G. and Bierlaire, M. (2013). Metropolis-hastings sampling of paths. *Transportation Research Part B: Methodological*, 48:53–66.

Bibliography III

Frejinger, E. (2008). *Route choice analysis - data, models, algorithms and applications*. PhD thesis, Ecole Polytechnique Fédérale de Lausanne.

Frejinger, E. and Bierlaire, M. (2007). Capturing correlation with subnetworks in route choice models. *Transportation Research Part B: Methodological*, 41(3):363–378.

Frejinger, E., Bierlaire, M., and Ben-Akiva, M. (2009). Sampling of alternatives for route choice modeling. *Transportation Research Part B: Methodological*, 43(10):984–994.

Guevara, C. A. and Ben-Akiva, M. E. (2013). Sampling of alternatives in multivariate extreme value (mev) models. *Transportation Research Part B: Methodological*, 48(0):31–52.

Bibliography IV

Lai, X. and Bierlaire, M. (2014). Specification of the cross nested logit model with sampling of alternatives for route choice models. Technical Report 140602, Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne.

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

Prashker, J. and Bekhor, S. (1999). Stochastic user-equilibrium formulations for extended-logit assignment models. *Transportation Research Record: Journal of the Transportation Research Board*, 1676(-1):145–152.

Ramming, M. S. (2002). *Network Knowledge and Route Choice*. PhD thesis, Massachusetts Institute of Technology.

Bibliography V

Roberts, B. and Kroese, D. P. (2007). Estimating the number of st paths in a graph. *J. Graph Algorithms Appl.*, 11(1):195–214.

Vovsha, P. and Bekhor, S. (1998). Link-nested logit model of route choice: Overcoming route overlapping problem. *Transportation Research Record: Journal of the Transportation Research Board*, 1645(-1):133–142.