Preferred citation style

Schuessler, N. (2007) Similarities in air transport connection choice, Third Workshop on Applications of Discrete Choice Models, EPF Lausanne, Lausanne, August 2007.

Similarities in air transport connection choice

Nadine Schuessler

IVT
ETH
Zurich

September 2007

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Problem and objectives

Two main objectives

- Modelling air transport connection choice

- Investigate similarity measures for public transport connection choice

Accounting for the IIA property of the MNL

In general, there are three different ways:

- Probit models and Mixed Multinomial Logit models
  (e.g. Yai et al. 1997, Ben-Akiva and Bolduc 1996, Walker 2002, Guo and Bhat 2005, Hess et al. 2005, Frejinger and Bierlaire 2007)

- Nested Approaches
  (Voshva and Bekhor 1998, Wen and Koppelman 2001, Bierlaire 2002)

- Similarity measures

Similarity measures

Based on the Implicit Availability and Perception (IAP) model by Cascatta et al. (1996)

The utility of an alternative is extended by a term q_{Cin} representing its degree of membership in the individual choice set C_n:

$$U_{in} = V_{in}^{'} + q_{Cin} + \varepsilon_{in}$$

The similarity of an alternative with other alternatives

- decreases its probability to be perceived as separate alternative
- decreases its probability to be included in the individual choice set

=> Thus decreases its probability to be chosen

Existing formulations for similarity measures

Private transport route choice

- C-Logit	Cascetta et al. (1996)
- Path Size Logit	Bierlaire and Ben-Akiva (1999), Ramming (2002)

Multimodal route choice

- Path Size Logit	Hoogendoorn-Lanser and Bovy (2007)

Public transport Connection Choice

- Autonomy of a connection	Friedrich et al. (2000)

Spatial Correlations

- Competing Destinations	Fotheringham (1988)
- Spatial Dependency Parameter	Mohammadian et al. (2005)
- Field Effect Variable	Dugundji and Walker (2005)

Autonomy of a connection

Choice probability depending on the independence of a connection

$$P_{in} = \frac{e^{V_{in}} \cdot IND(i)}{\sum_{j \in C} e^{V_{in}} \cdot IND(j)} = \frac{e^{V_{in} + \ln(IND(i))}}{\sum_{j \in C} e^{V_{in} + kn(IND(j))}}$$

Independence of a connection as reciprocal sum of similarities:

$$IND(c) = \frac{1}{\sum_{c' \in C} f_c(c')} = \frac{1}{1 + \sum_{c' \in C; c' \neq c} f_c(c')}$$

Similarity takes into account similarities in departure and arrival times $x_c(c')$, perceived journey time $y_c(c')$ and fare $z_c(c')$

$$f_c(c') = \left(1 - \frac{x_c(c')}{s_x}\right)^+ \cdot \left(1 - \gamma \cdot \min\left(1, \frac{|y_c(c')|}{s_y} + \frac{|z_c(c')|}{s_z}\right)\right)$$

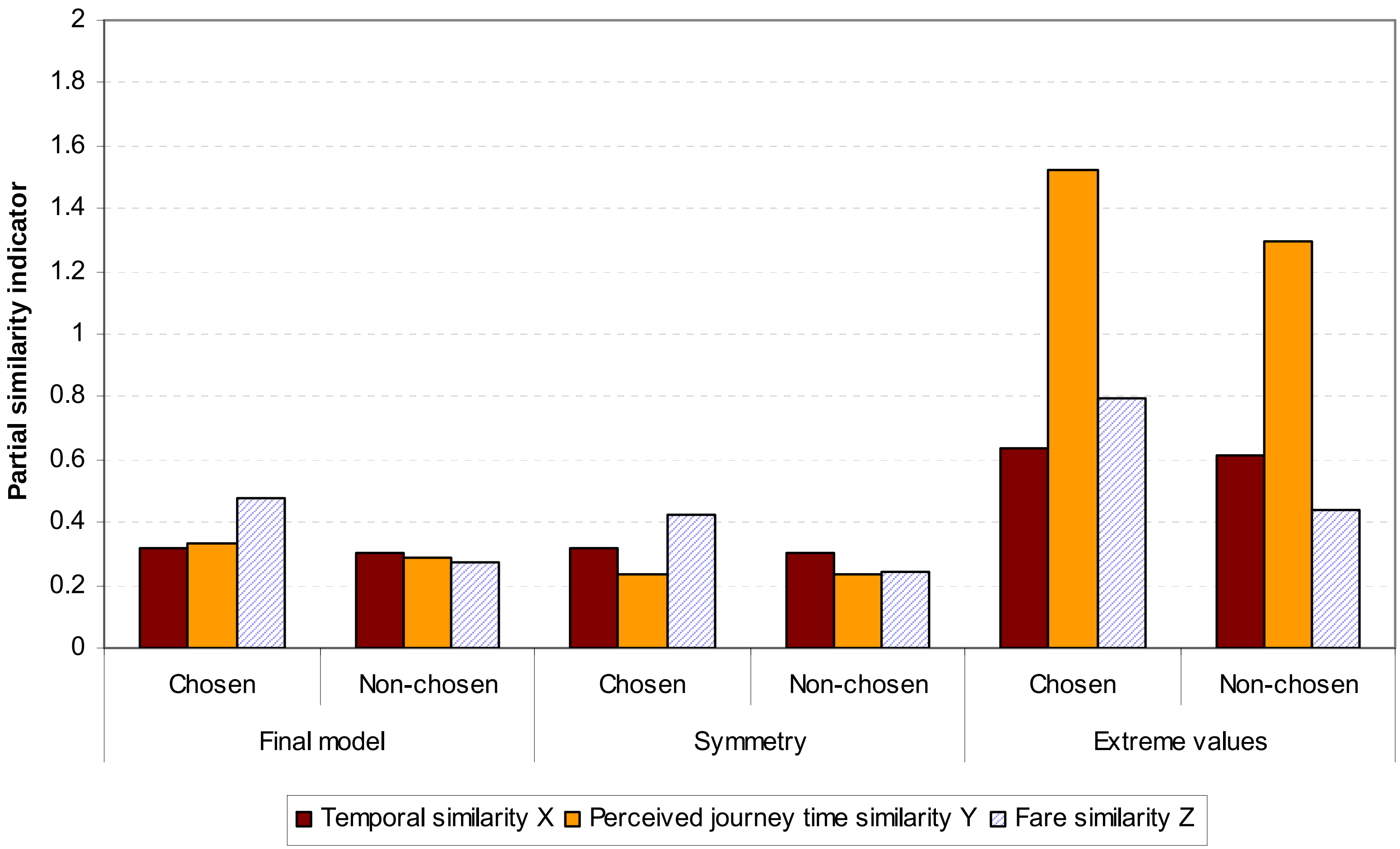
s_x, s_y and s_z set the range of influence of $x_c(c')$, $y_c(c')$ and $z_c(c')$

γ weights the right part of the formula

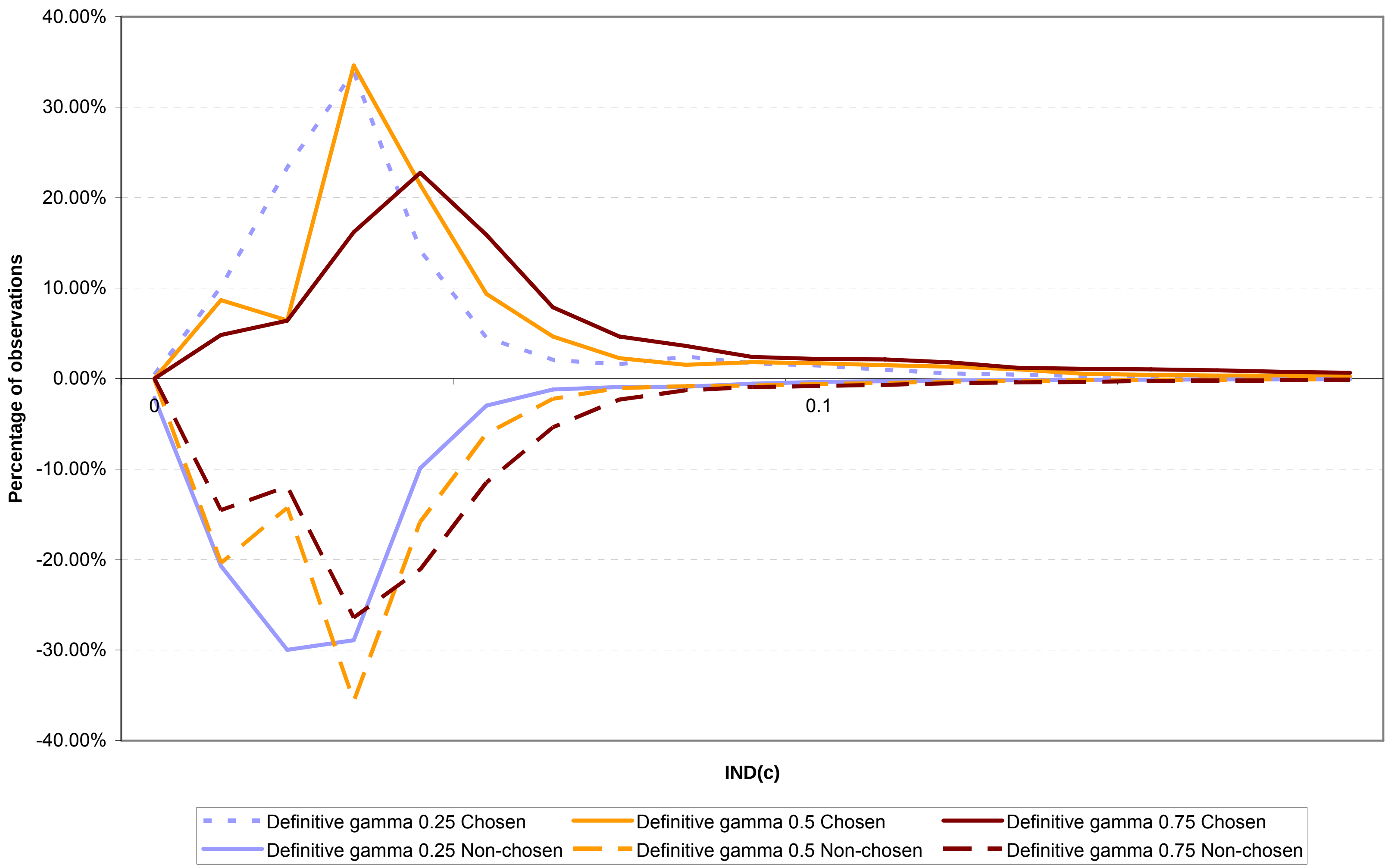
Parameter settings for the independence measure

Parameter	s_x, s_y, s_z tests		γ tests		Final model
	Symmetric setting	Extreme values	$\gamma = 0.25$	$\gamma = 0.75$	$\gamma = 0.5$
s_x	720	360	720	720	720
s_y^+	780	180	780	780	780
s_y^-	780	120	540	540	540
s_z^+	1.7	0.9	1.5	1.5	1.5
s_z^-	1.7	0.9	1.5	1.5	1.5
γ	0.5	0.5	0.25	0.75	0.5

Effect of varying s_x, s_y and s_z



Effect of varying γ



Available Data – Data Sources

Marketing Data Information Transfer (MIDT) Dataset – booked tickets in November 2006

~ 200,000 bookings

Expedia Dataset observed prices on 70 origin-destination pairs between September and November 2006

~ 12,000,000 observations

Official Airline Guide (OAG) November – information on code share, type of aircraft, detailed waiting times can be extracted

Matching these datasets on several criteria leads to almost 19,000 observed choices

Variables for connection choice model

Alternatives are return flights with several characteristics:

- Carrier
- Code share
- Aircraft type (mainline jet, regional aircraft, propeller aircraft)

- Total travel time
- Number of transfers

- Fare without tax

- Combination of time of day of out- and inbound flight and return period

- Independence of the connection

Modelling results

Model	MNL		MNL with IND(c)	
	Parameter	Robust t-test	Parameter	Robust t-test
Carrier const.	not presented			
Code share	-0.7465	-12.81	-0.7622	-12.29
Travel time	-0.0006	-5.86	0.0002	1.25
N.o. Transfers	-6.2404	-40.42	-6.6154	-37.75
Fare	-0.0057	-73.71	-0.0063	-66.68
Similarity	0.0000	0.00	-1.0362	-12.97
Time of day	not presented			
Aircraft type				
Regional	-0.2170	-8.55	-0.2087	-8.65
Propeller	-1.6427	-19.04	-1.6711	-18.40

Modelling results for different parameter settings

Model	Parameter for IND(c)	Average for IND(c)	Final Log-Likelihood	Adjusted r-square	Difference r-square*
MNL	--	--	-48816	0.2948	--
s_x=720,γ=0.25	-1.1966	0.0386	-48712	0.2963	0.5%
s_x=720,γ=0.5	-1.0362	0.0473	-48684	0.2967	0.6%
s_x=720,γ=0.75	-0.7210	0.0621	-48752	0.2957	0.3%
s_x=120,γ=0.5	-0.1835	0.1650	-48798	0.2950	0.1%
s_x=240,γ=0.5	-0.6776	0.0984	-48622	0.2976	0.9%
s_x=360,γ=0.5	-0.7666	0.0734	-48640	0.2973	0.8%

* Compared to MNL

Conclusion

Estimated parameters for the independence measure are highly significant and the model performance increases

Traveller's perception of similarity between air transport connections is multi-dimensional

The inclusion of multiple dimensions can lead to more insight in preferences without penalising other alternatives too strong

Setting of the parameters defining the range of influence has to be determined for specific data set

The sign of the parameter indicates that tra perceive similar alternatives as positive

Outlook

For the air connection choice data set:

- Testing the influence of the individual similarity components
- Comparison with other model structures, such as (cross)-nested model structures
- Closer look to the formation of the choice

Investigation of similarity measures based on a GPS data set:

- Investigation of other choice situations
- Testing different similarity measures
- Comparison to complex model structures

The GPS data set

- On-person GPS data for travellers living in Zurich, Winterthur and Geneva

- Complete multi-modal urban and inter-urban trip chains

- GPS records of 4878 persons comprising on average 6.64 days per persons

- No additional information about socio-demographics, trip details etc.

Modelling the GPS data

Motorised private transport route choice

Public transport connection choice

Combined route and mode choice

Destination choice

Combined route, mode and destination choice

Post-processing the GPS records

Adaptation of the method by Tsui and Shalaby (2006) for

- Trip and activity detection
- Identification of single-mode segments
- Mode detection

Map-matching with method by Marchal, Hackney and Axhausen (2006)

Activity type and trip purpose classification

Validation with Swiss Microcensus 2005

Choice set generation

References (1)

Ben-Akiva, M. E. and D. Bolduc (1996) Multinomial probit with a logit kernel and a general parametric specification of the covariance structure, in ICS (ed.) the 3rd Invitational Choice Symposium.

Bierlaire, M. (2002) The Network GEV model, in STRC (ed.) the 2nd Swiss Transport Research Conference (STRC), Ascona, Mar. 2002, http://www.strc.ch/2002.html.

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

Dugundji, E. R. and J. L. Walker (2005) Discrete choice with social and spatial network interdepencies: An empirical example using mixed gev models with field and "panel" effects, in TRB (ed.) the 84th Annual Meeting of the Transportation Research Board, Washington, D.C., Jan. 2005.

Fotheringham, A. S. (1988) Consumer store choice and choice set definition, Marketing Science, 7 (3) 299–310.

References (2)

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

Friedrich, M., I. Hofsäss and S. Wekeck (2001) Timetable-based transit assignment using branch and bound techniques, Transportation Research Record, 1752, 100–107.

Guo, J. Y. and C. R. Bhat (2005) Operationalizing the concept of neighborhood: Application to residential location choice analysis, in TRB (ed.) the 84th Annual Meeting of the Transportation Research Board, Washington, D.C., Jan. 2005.

Hess, S., M. Bierlaire and J. W. Polak (2005) Capturing taste heterogeneity and correlation structures with Mixed GEV models, in TRB (ed.) the 84th Annual Meeting of the Transportation Research Board, Washington, D.C., Jan. 2005.

Hoogendoorn-Lanser, S. and P. H. L. Bovy (2007) Modeling overlap in multi-modal route choice by inclusion of trip part specific path size factors, in TRB (ed.) the 86th Annual Meeting of the Transportation Research Board, Washington, D.C., Jan. 2007.

Reference (3)

Marchal, F., J.K. Hackney and K.W. Axhausen (2006) Efficient map-matching of large GPS data sets - Tests on a speed monitoring experiment in Zurich, Transportation Research Record, 1935, 93-100 .

Mohammadian, A. K., M. Haider and P. S. Kanaroglou (2005) Incorporating spatial dependencies in random parameter discrete choice models, in TRB (ed.) the 84th Annual Meeting of the Transportation Research Board, Washington, D.C., Jan. 2005.

Ramming, M. S. (2002) Network knowledge and route choice, Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge.

Tsui S.Y.A. and A.S. Shalaby (2006) An Enhanced System for Link and Mode Identification for GPS-based Personal Travel Surveys, in TRB (ed.) the 85th Annual Meeting of the Transportation Research Board, Washington, D.C., Jan. 2006.

van Eggermond, M., N. Schüssler and K.W. Axhausen (2007) Accounting for similarities in air transport route choice, Working Paper, XXX, IVT, ETH Zurich, Zurich.

Reference (4)

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

Walker, J. L. (2001) Extended discrete choice models: Integrated framework, flexible error structures, and latent variables, Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge.

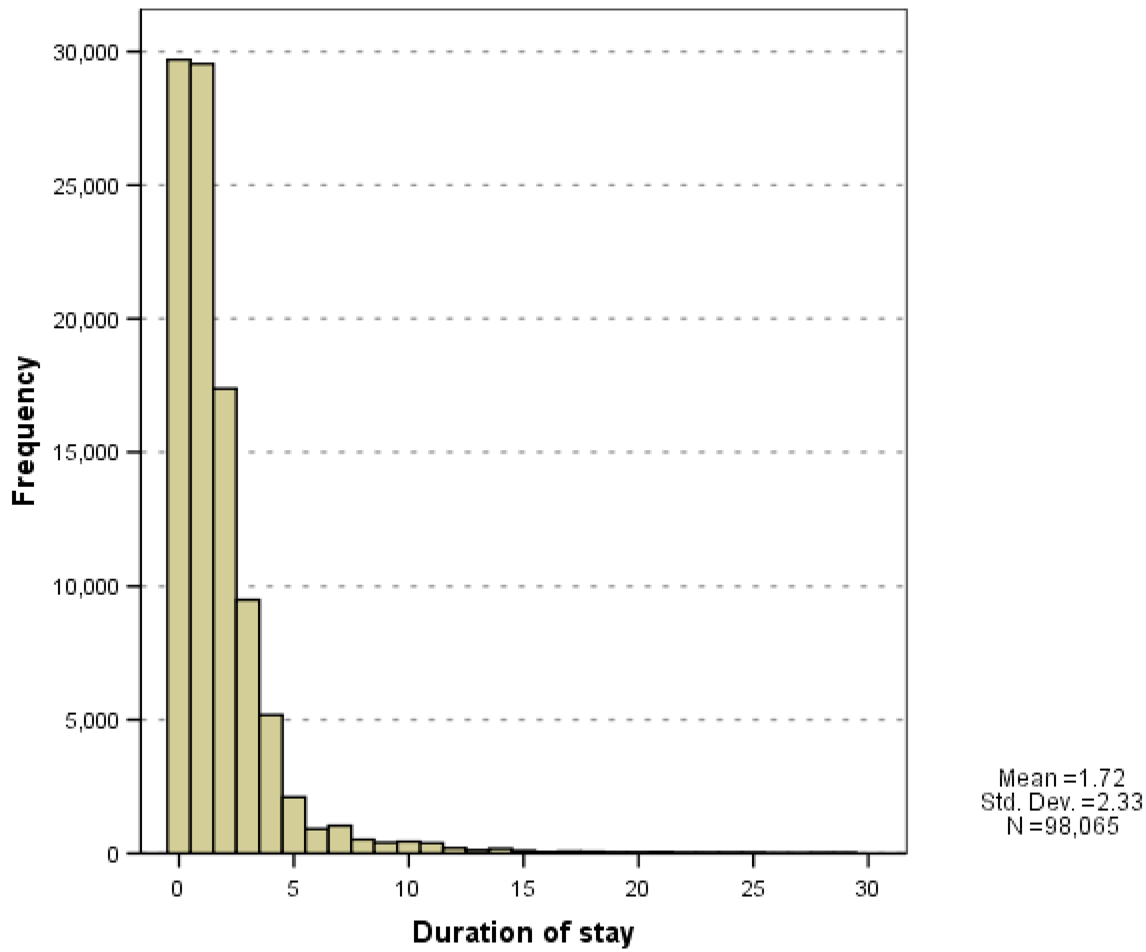
Weis, C. (2006) Routenwahl im ÖV, Master Thesis, IVT, ETH Zurich, Zurich.

Wen, C.-H. and F. S. Koppelman (2001) The generalized nested logit model, Transportation Research Part B: Methodological, 35 (7) 627–641.

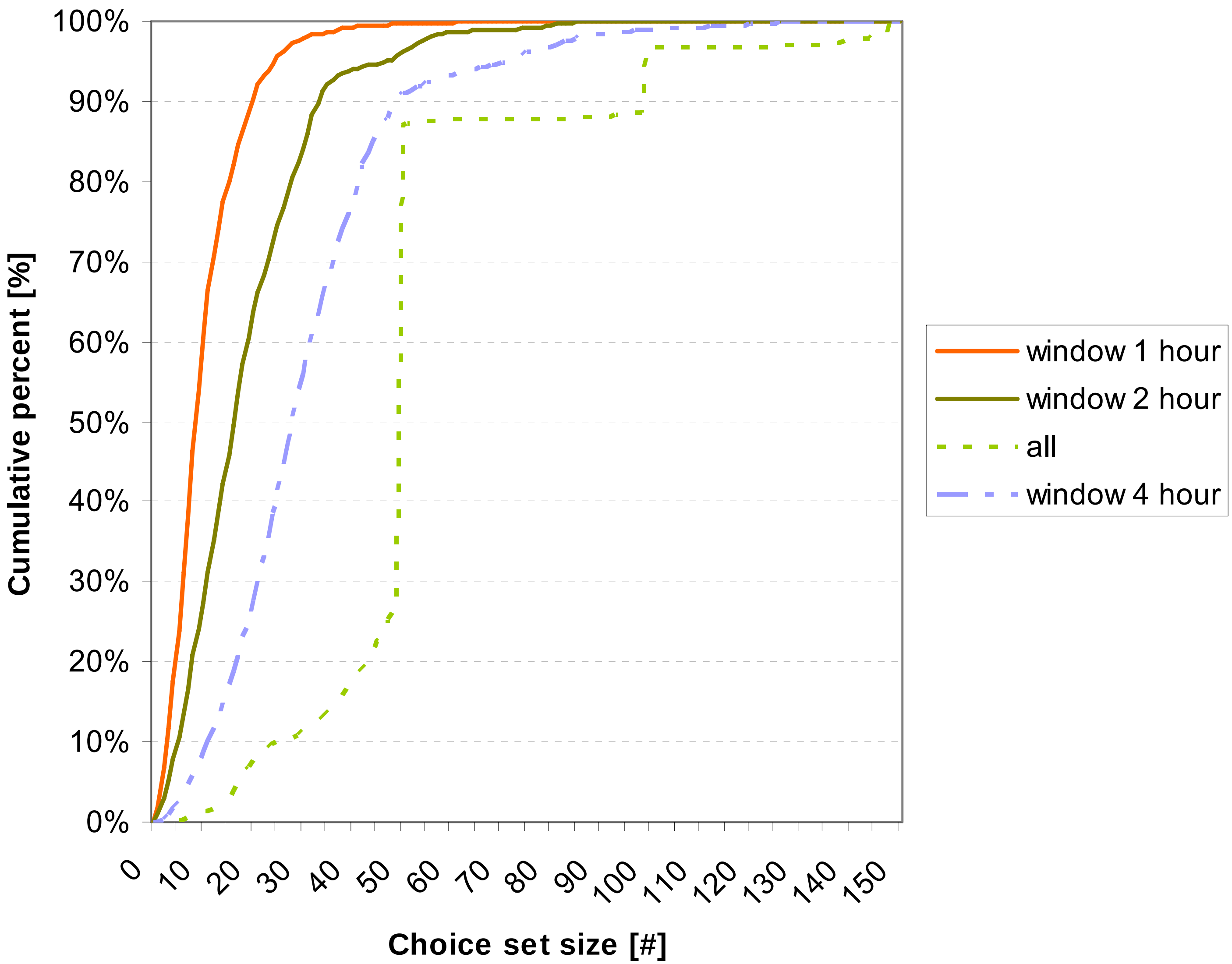
Yai, T., S. Iwakura and S. Morichi (1997) Multinomial probit with structured covariance for route choice behaviour, Transportation Research Part B: Methodological, 31 (3) 195–208.

Appendices

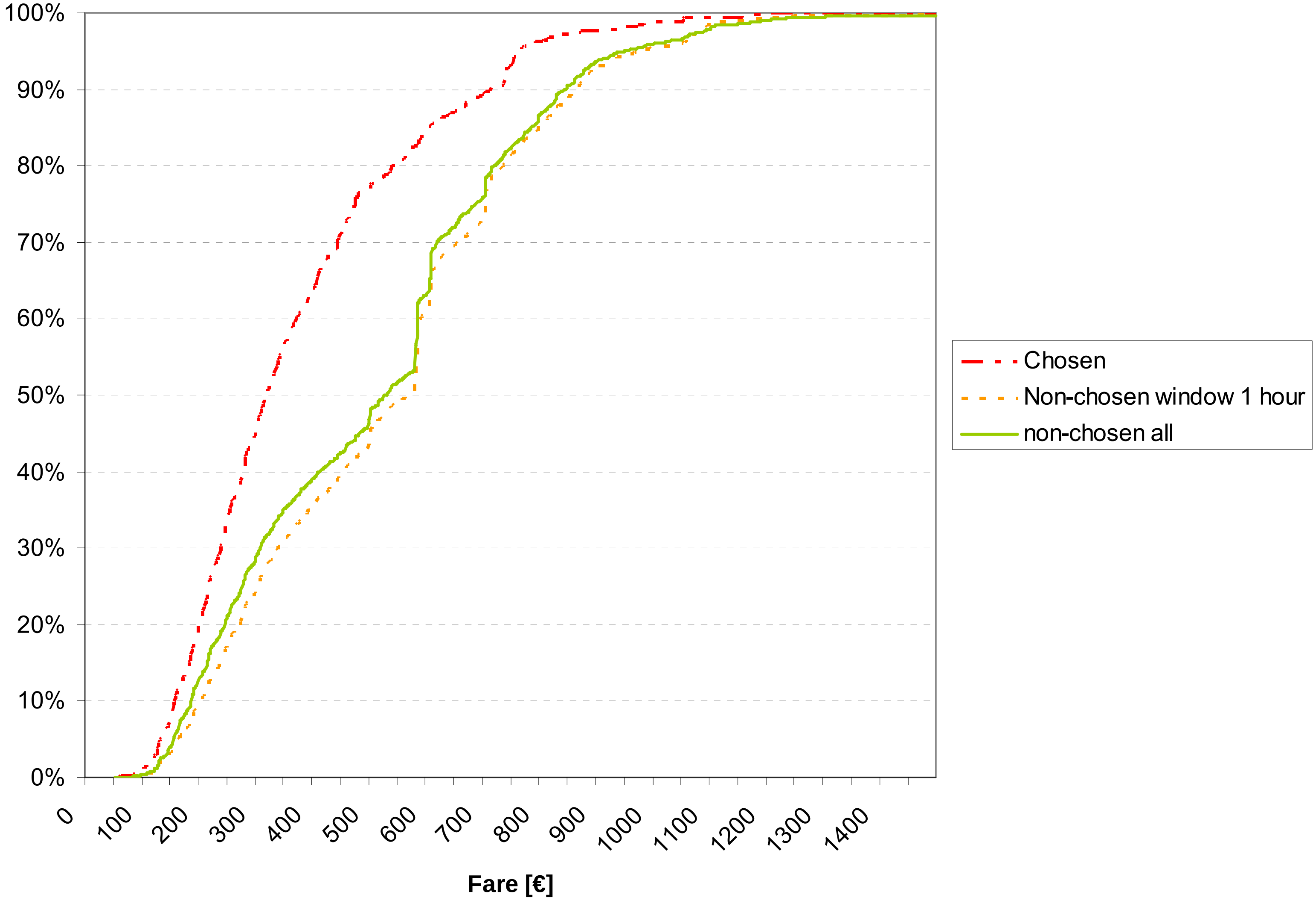
Available Data – Duration of Stay



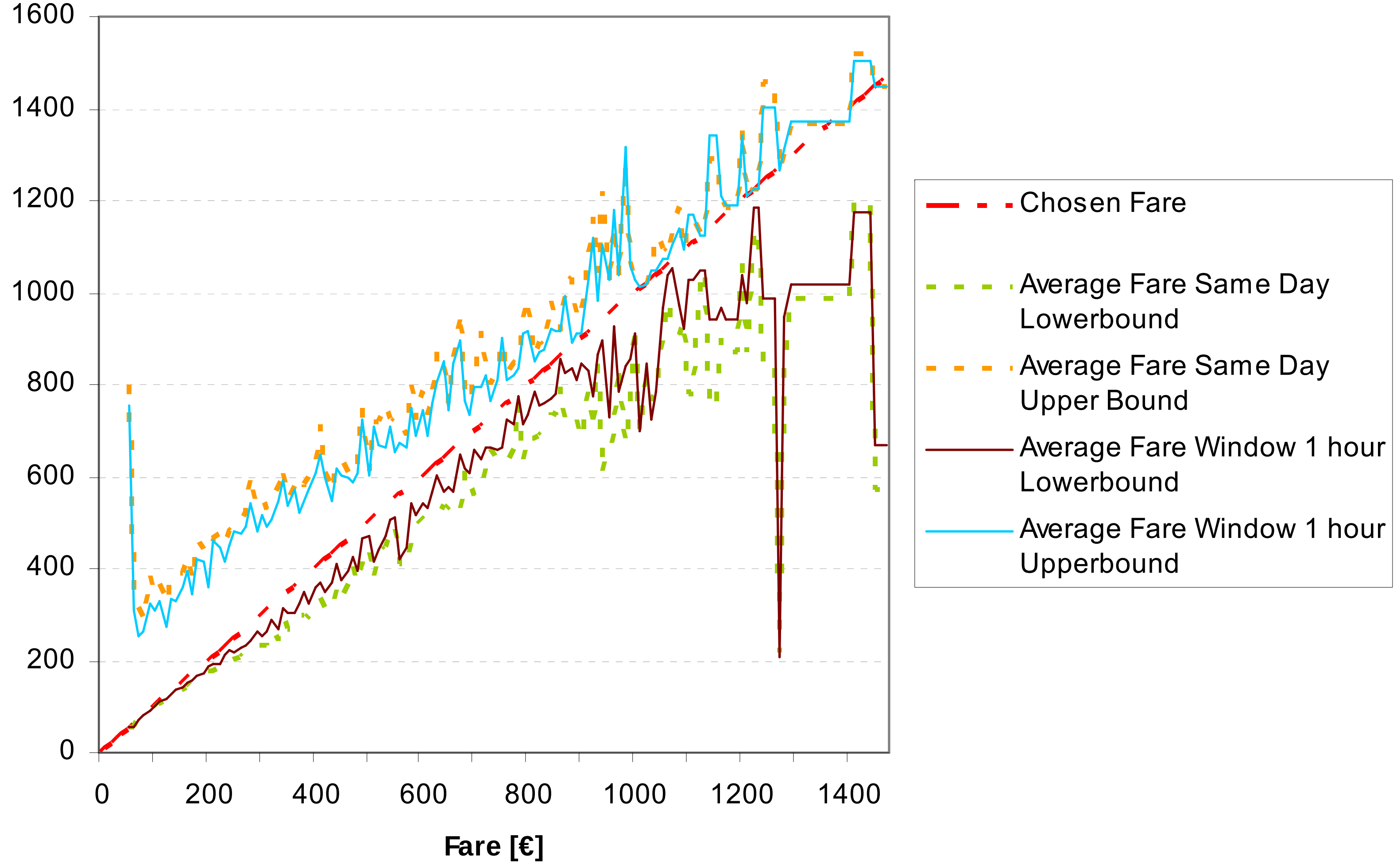
Available Data – Choice Set Size



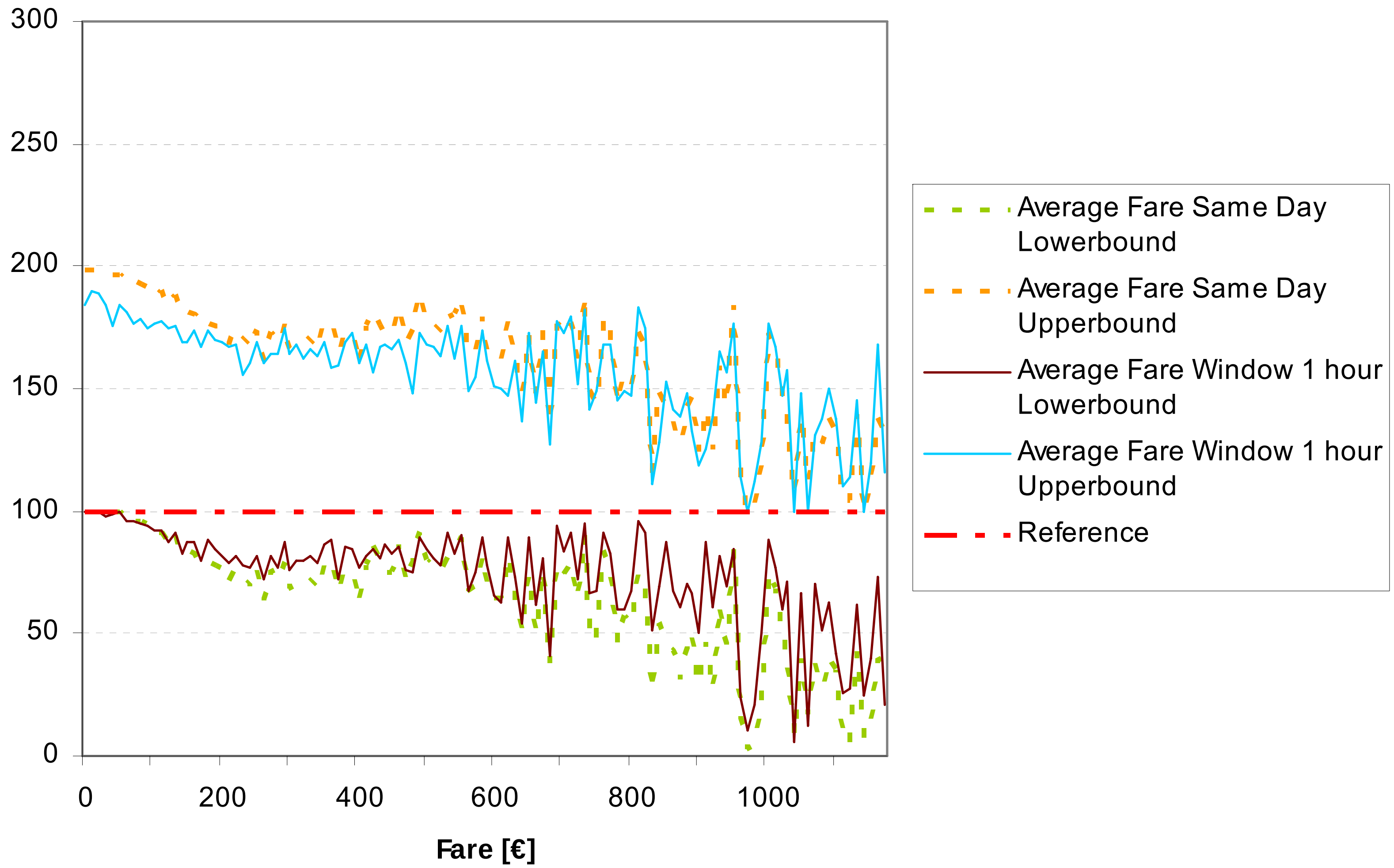
Available Data – Comparison Choice Set Fare



Available Data – Comparison Choice Set Fare



Available data – Comparison Choice Set Fare



Modelling results (2)

	MNL		MNL with IND(c)	
	Parameter	Robust t-test	Parameter	Robust t-test
Same day return				
Afternoon Uut	-1.6057	-63.39	-1.5867	-61.10
Evening Out	-2.4960	-18.59	-2.087	-14.97
Overnight stay				
Afternoon Out	-0.7807	-24.92	-1.0462	-26.70
Evening Out	-0.8573	-26.15	-2.0964	-14.97
Fortnight stay				
Afternoon Out	-0.2829	-2.75	-0.5711	-4.85
Evening Out	not significant			
Same day return				
Afternoon In	not significant			
Evening In	0.1907	2.04	0.1958	1.88
Overnight stay				
Afternoon In	1.1967	12.90	1.2171	12.56
Evening In	1.5915	17.28	1.6055	16.63
Fortnight stay In	not significant			