Preferred citation style

Schüssler N. (2007) The role of similarities for air connection choice, presentation at the IVT-Seminar ETHZ, Zürich, Dezember 2007.

The role of similarities for air connection choice

Nadine Schüssler

IVT
ETH
Zürich

Dezember 2007

Institut für Verkehrsplanung und Transportsysteme
Institute for Transport Planning and Systems

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Agenda

Background and Modelling Framework

Available Data

Specification

Results

Outlook

Problem and objectives

Two main objectives

- Modelling air transport connection choice

- Investigating a similarity measure for public transport connection choice

The Multinomial Logit (MNL) model (Mc Fadden, 1974)

Basic assumptions:

Decision-maker *n* chooses from choice set C_n alternative i with the highest utility U_{in}

U_{in} can be represented by a function

$$U_{in} = V_{in} + \varepsilon_{in}$$

The MNL model estimates for each alternative the probability of being chosen:

$$P(i|C_n) = P[U_{in} \geq U_{jn}, \forall j \in C_n]$$

$$P(i|C_n) = \frac{e^{\mu V_{in}}}{\sum_j e^{\mu V_{jn}}}$$

Independence of Irrelevant Alternatives (IIA) property

The relative ratio of the choice probabilities of two alternatives is not effected by the presence or the characteristics of any other alternatives.

$$\frac{P(i|C_n)}{P(k|C_n)} = \frac{\dfrac{e^{\mu V_{in}}}{\sum_j e^{\mu V_{jn}}}}{\dfrac{e^{\mu V_{jn}}}{\sum_j e^{\mu V_{jn}}}} = e^{\mu(V_{in} - V_{kn})}$$

Similarity between alternatives are not accounted for, even though:

- Travelers are likely not able to distinguish between similar alternatives
- Error terms are not independently distributed when alternatives are similar

Overcoming the IIA property

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 – Basic idea

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

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

The similarity of an alternative with other alternatives

- influences its perception as separate alternative
- influences its inclusion in the individual choice set

=> Thus influences its probability to be chosen

$$P(i|C_n) = \frac{e^{\mu\left(V'_{in} + \ln \vartheta_{Cin}\right)}}{\sum_j e^{\mu\left(V'_{jn} + \ln \vartheta_{Cjn}\right)}}$$

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	
- Independence of a Connection	Friedrich et al. (2001)
Spatial Correlations	
- Competing Destinations	Fotheringham (1988)
- Spatial Dependency Parameter	Mohammadian et al. (2005)
- Field Effect Variable	Dugundji and Walker (2005)
- Concept of Dominance	Cascetta and Papola (2005)

Independence 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

Case study – Data sources

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

~ 200,000 bookings

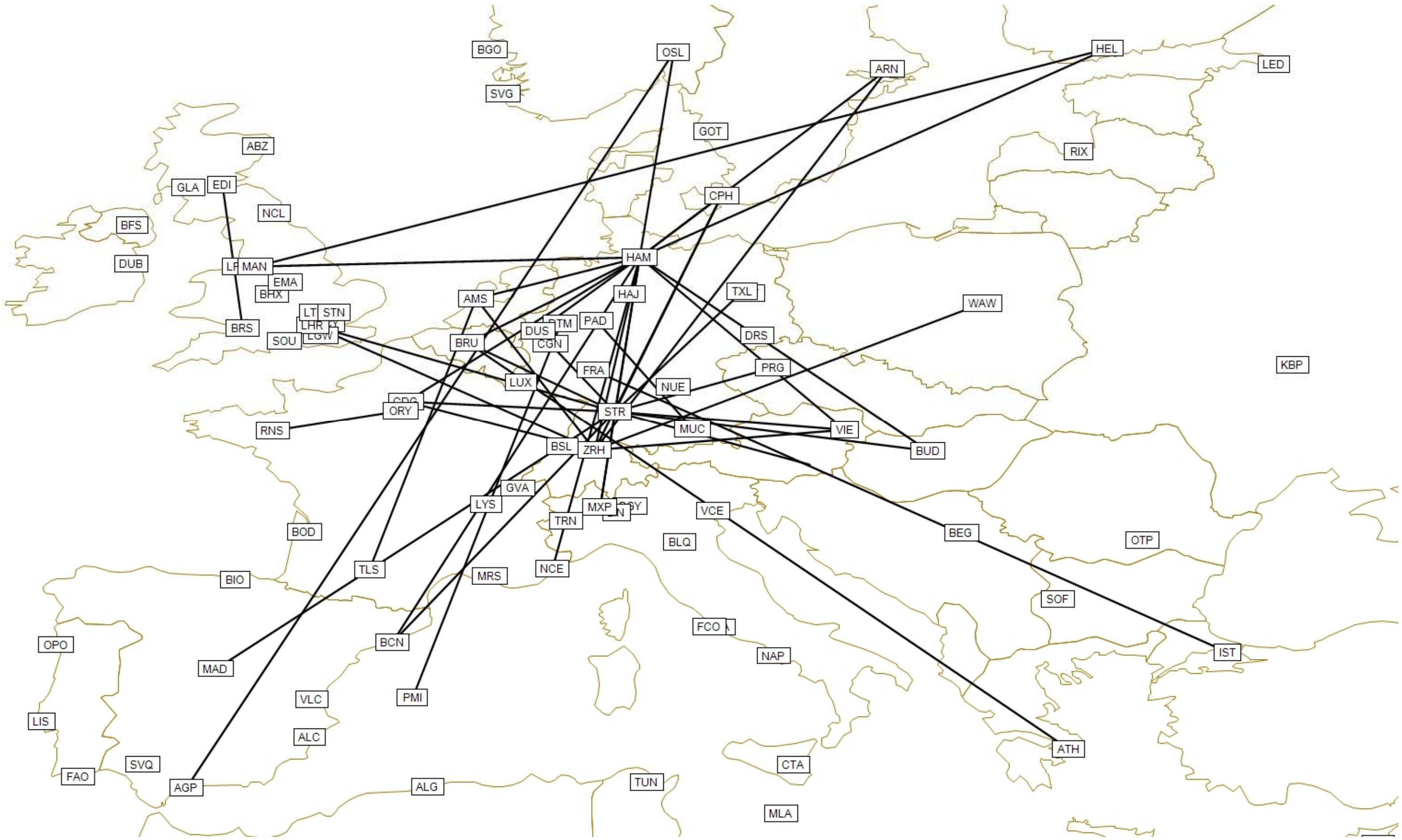
Expedia dataset observed prices on 70 origin-destination pairs between September and November 2006

~ 12,000,000 observations

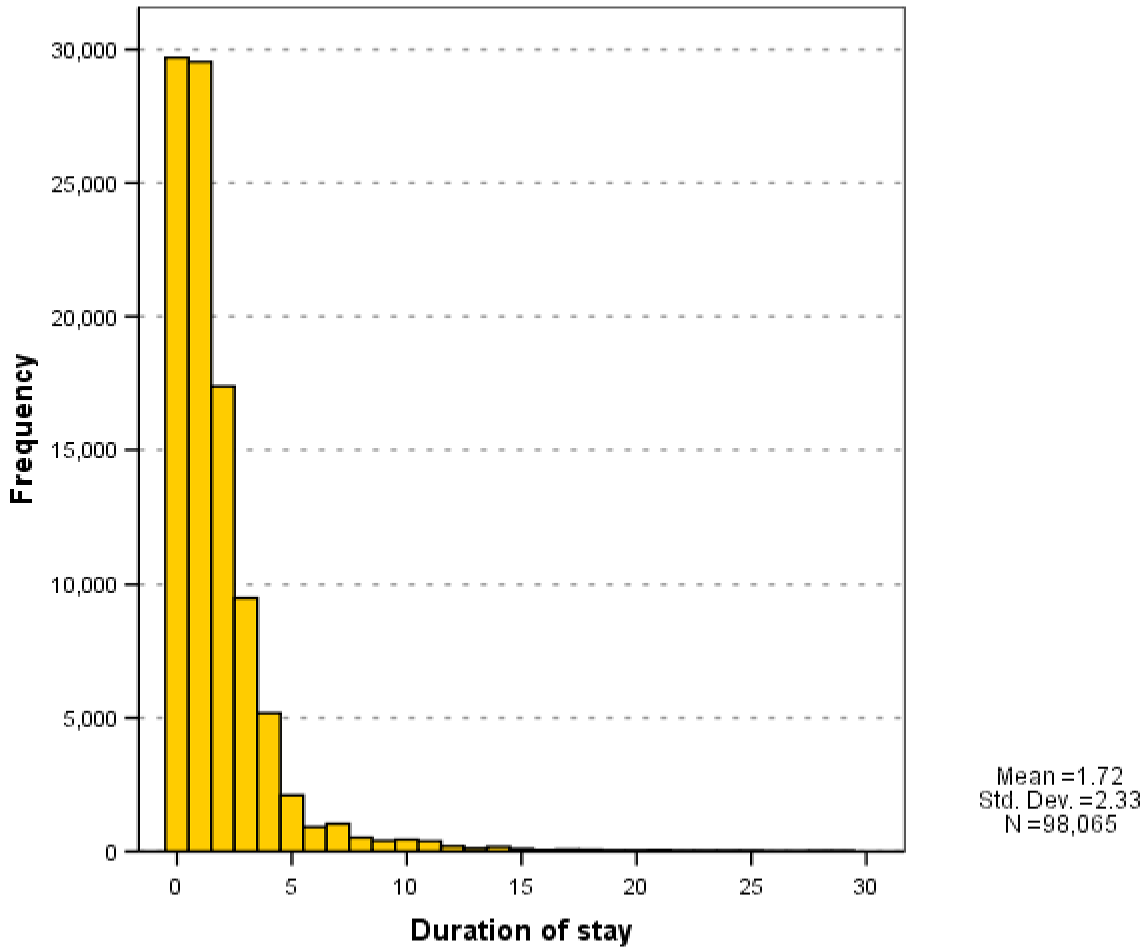
Official Airline Guide (OAG) November – information on code share, type of aircraft, detailed waiting times can be extracted

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

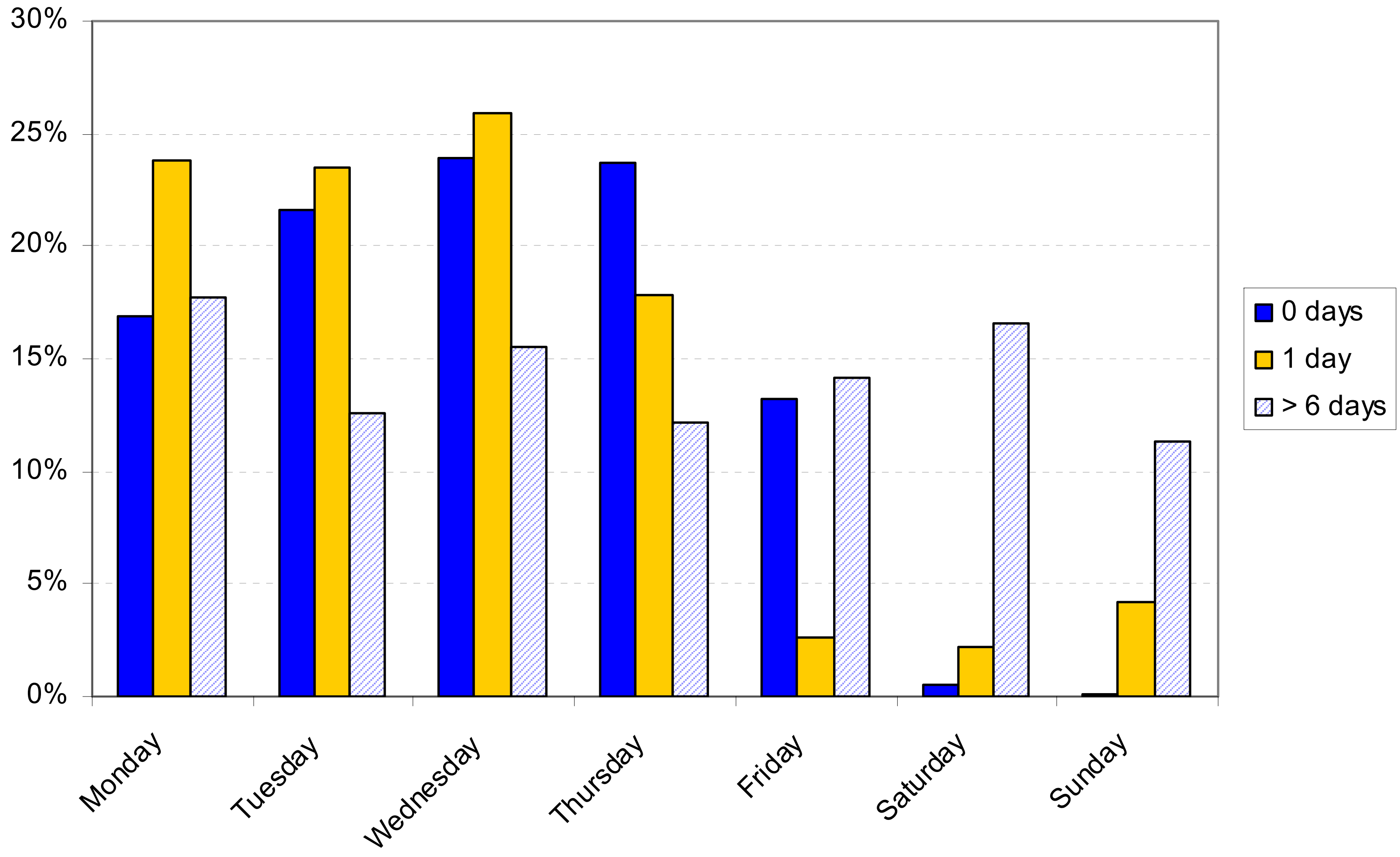
Observed markets



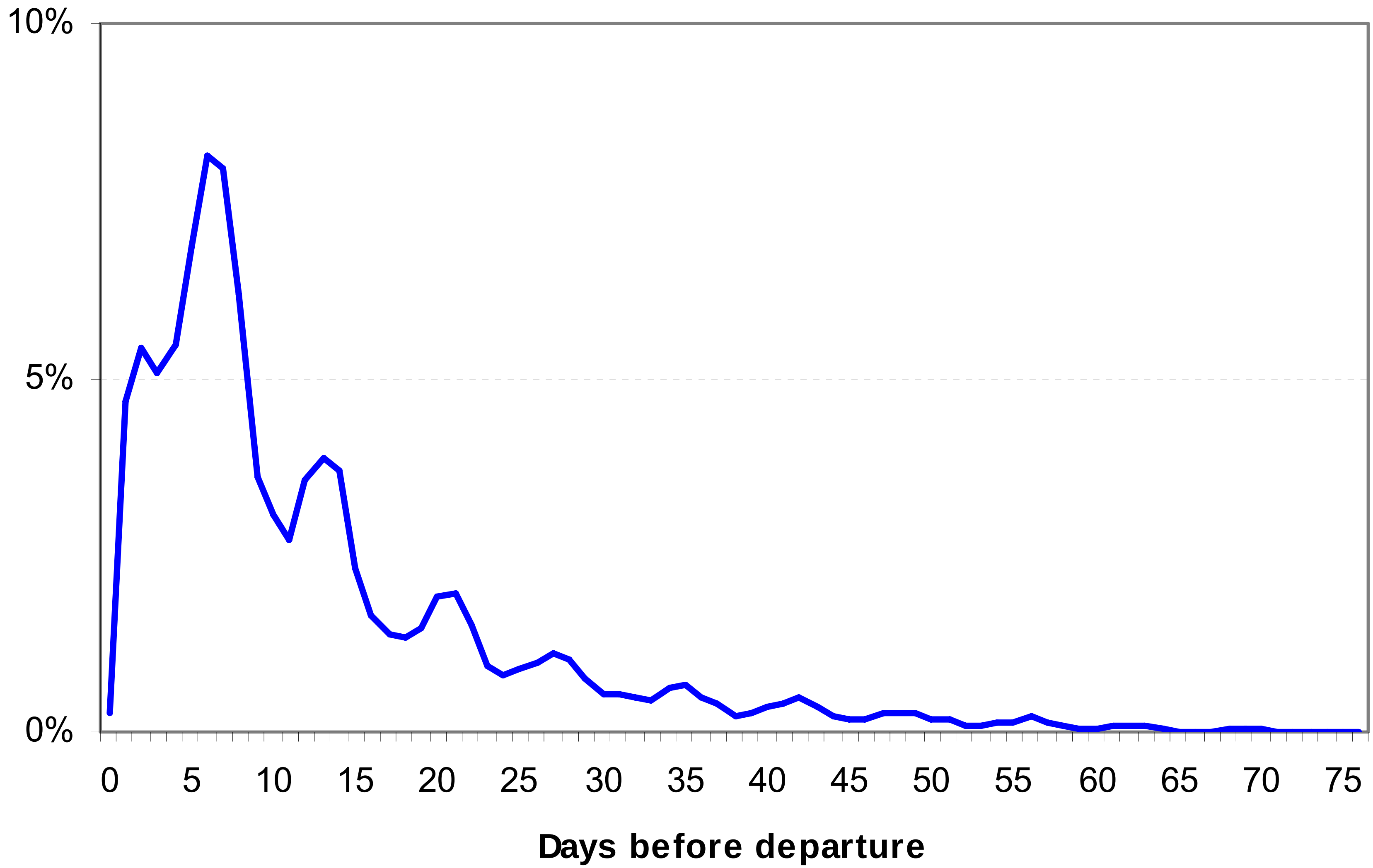
Duration of stay distribution



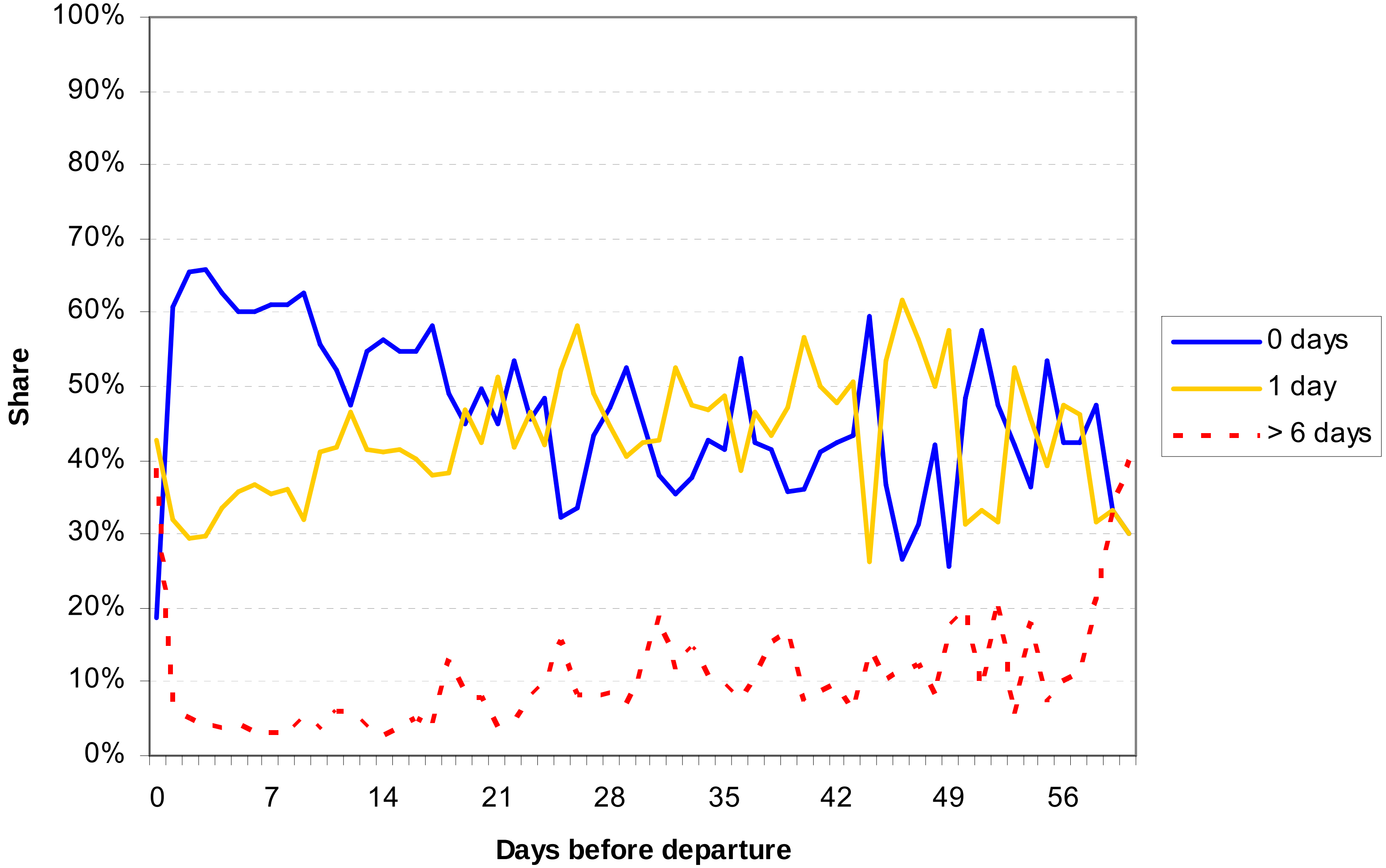
Duration of stay per weekday



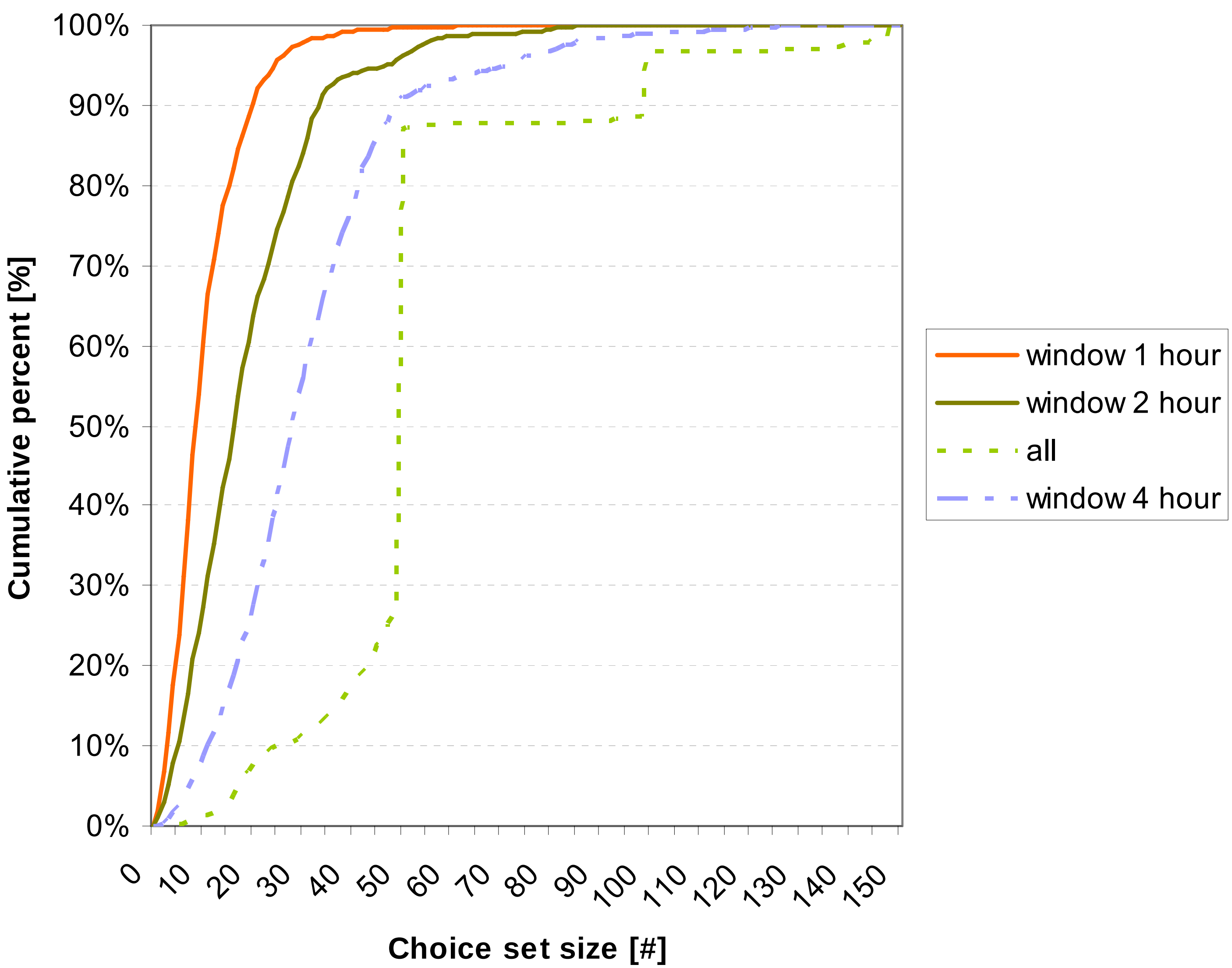
Share of bookings



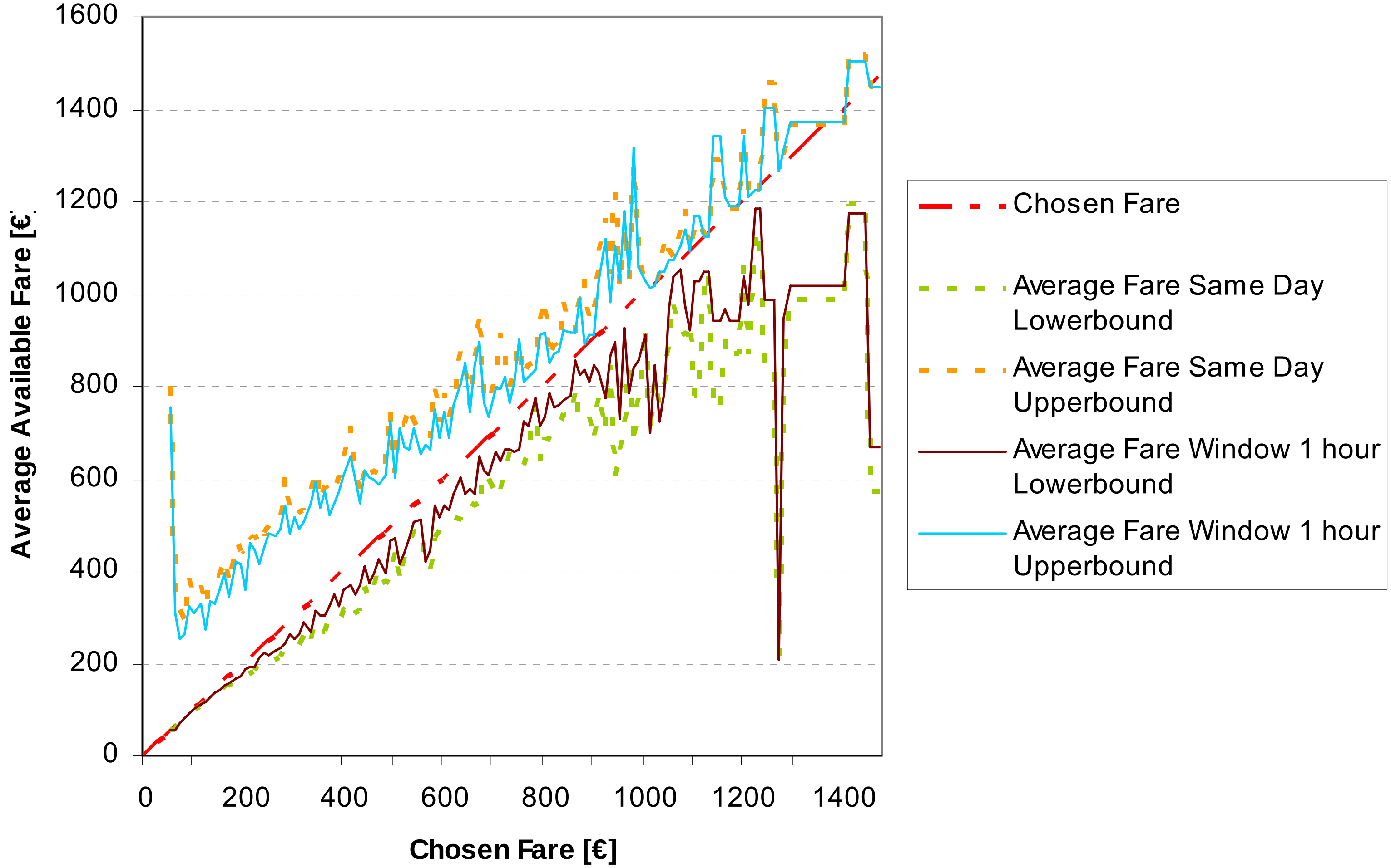
Share of bookings per days before departure



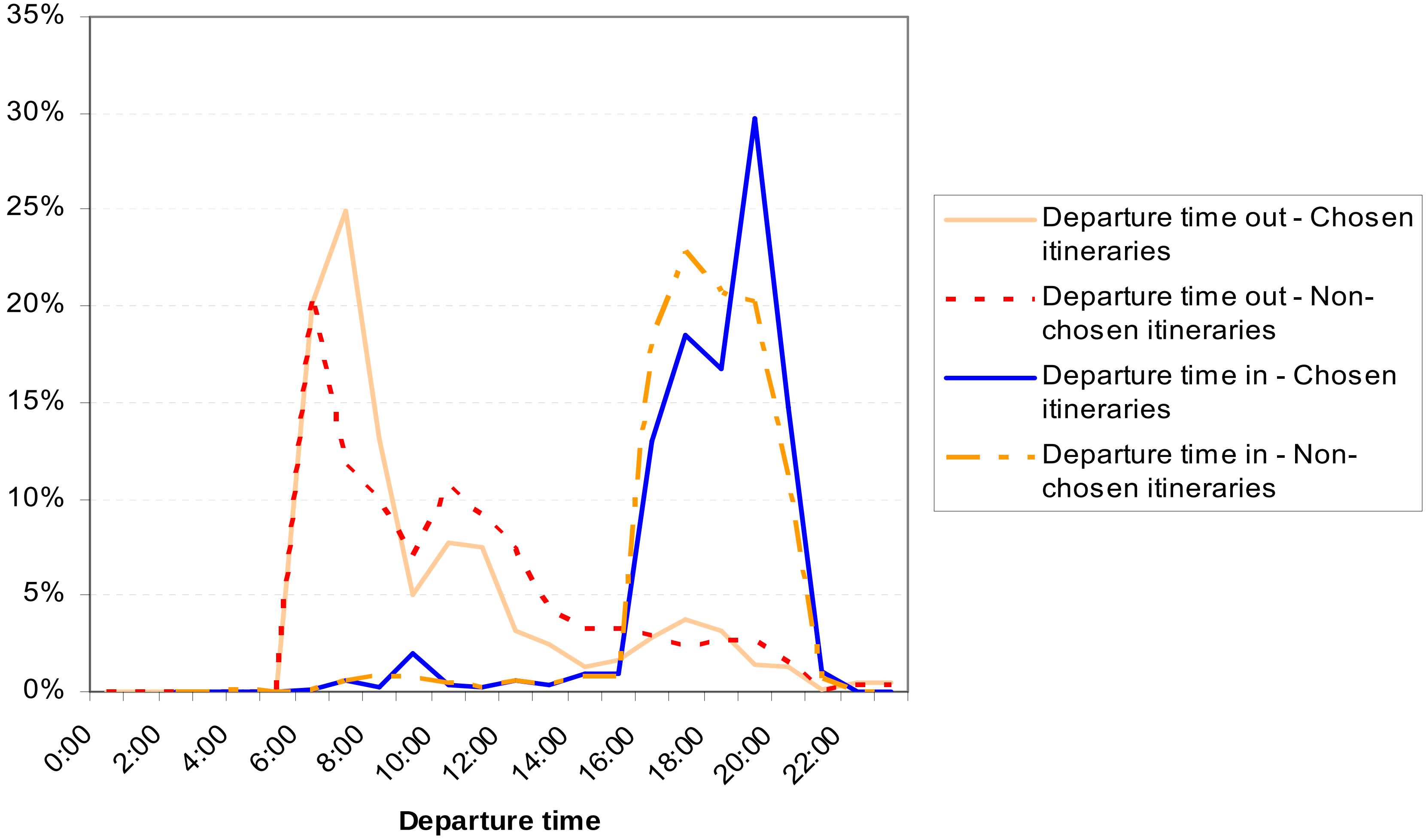
Choice set size



Comparison choice set fare



Departure time distribution – same day return



Specification of the independency measure

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 $z_c(c')$

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

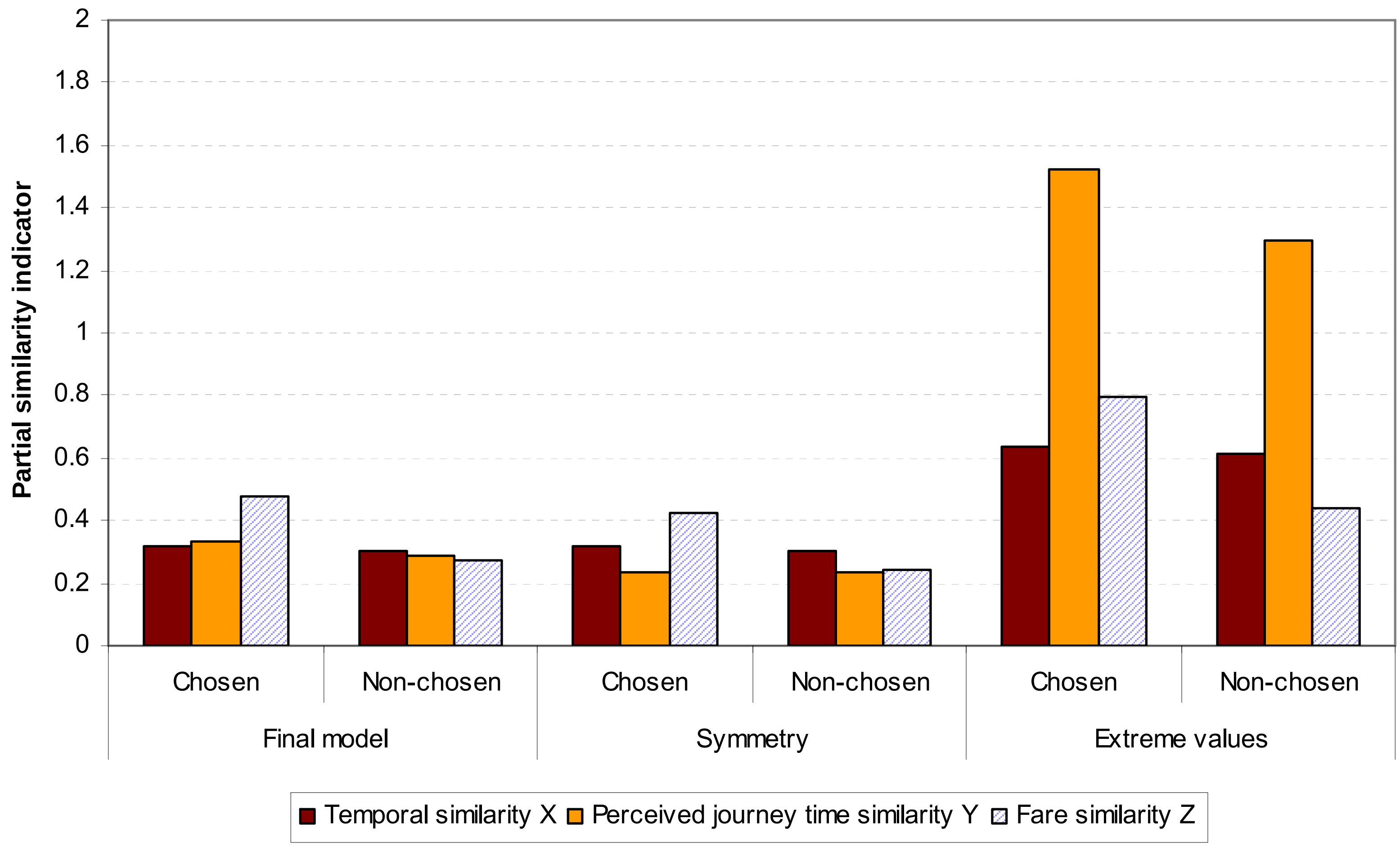
s_x, s_y and s_z set the range of influence of $x_c(c')$, $y_c(c')$ and $z_c(c')$

γ weights the right part of the formula

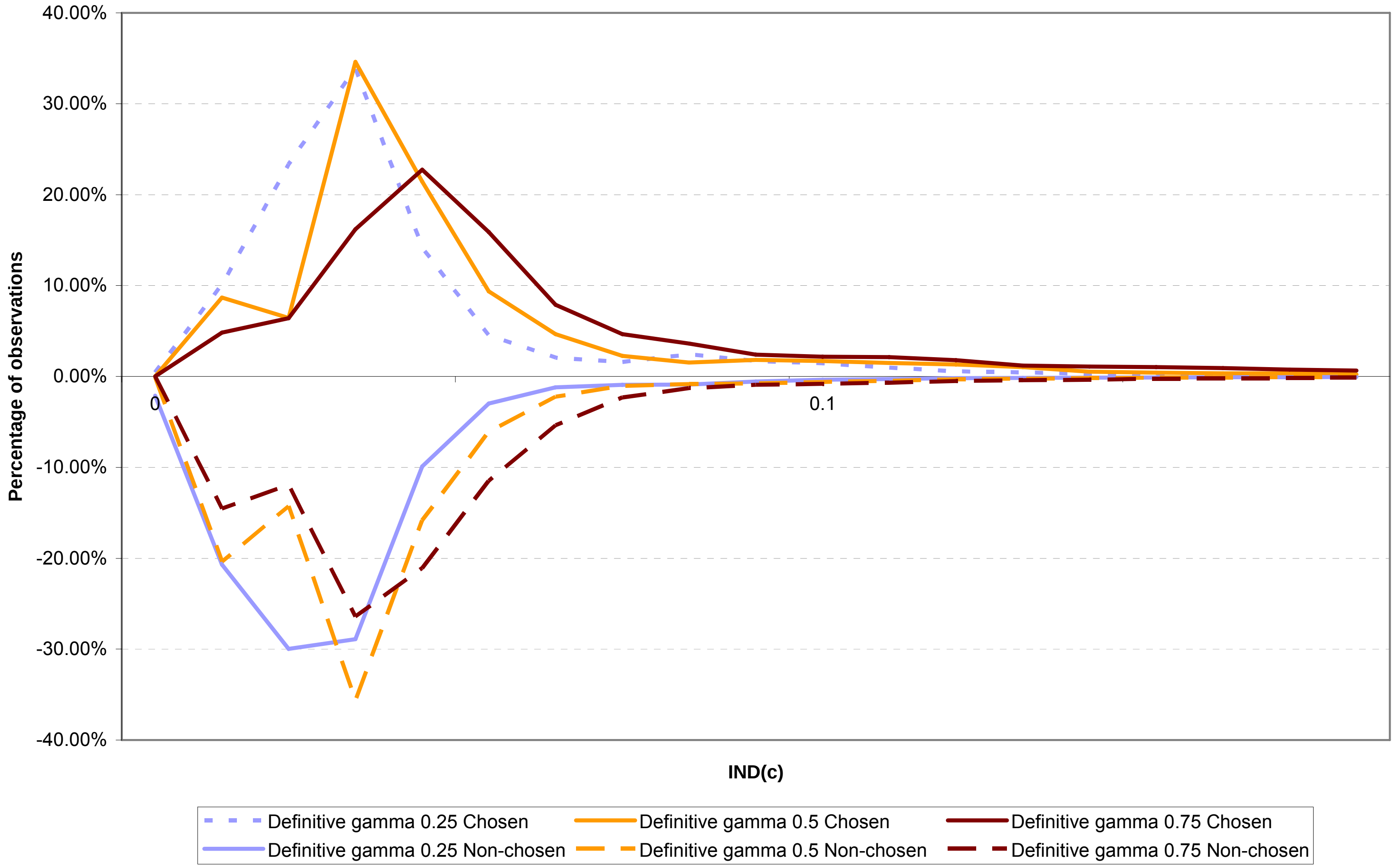
Parameter settings for the independence measure

Parameter	s_x, s_y, s_z tests		γ tests		Final model
	Symmetric setting	Extreme values	γ = 0.25	γ = 0.75	γ = 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

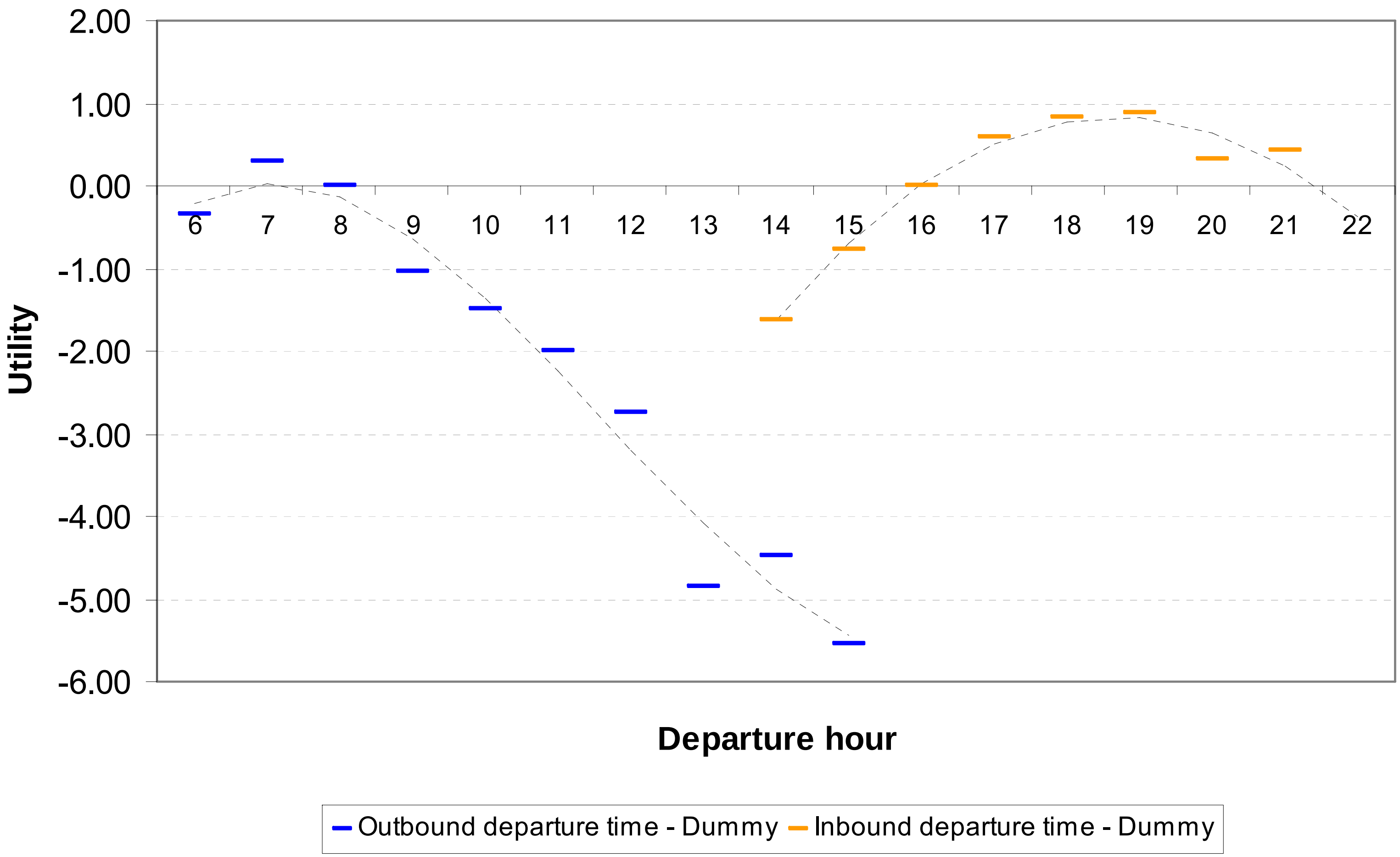


Valuation of service attributes

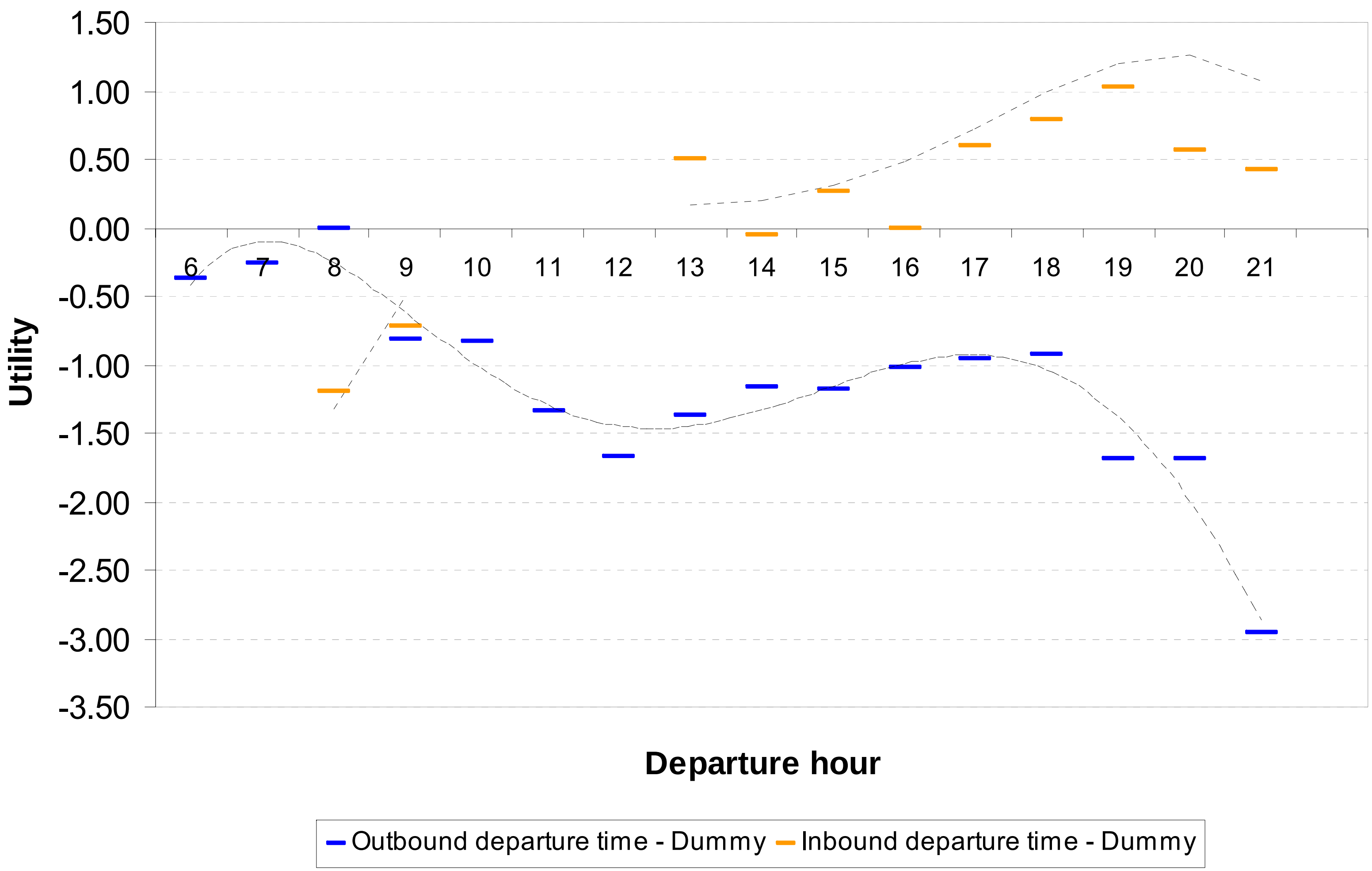
Valuation	Same day	Next day	> 6 days
Code share [€]	-119.39	-165.03	-110.84
Regional aircraft* [€]	-19.82	-27.40	-18.40
Propeller aircraft* [€]	-201.04	-277.90	-186.64
Travel time [€/h]	-90.50	-125.09	-84.02
Transfer [€]	-602.52	-832.92	-559.41
Transfer [min]	400.96	400.96	400.96

(*) base category: Mainline jet

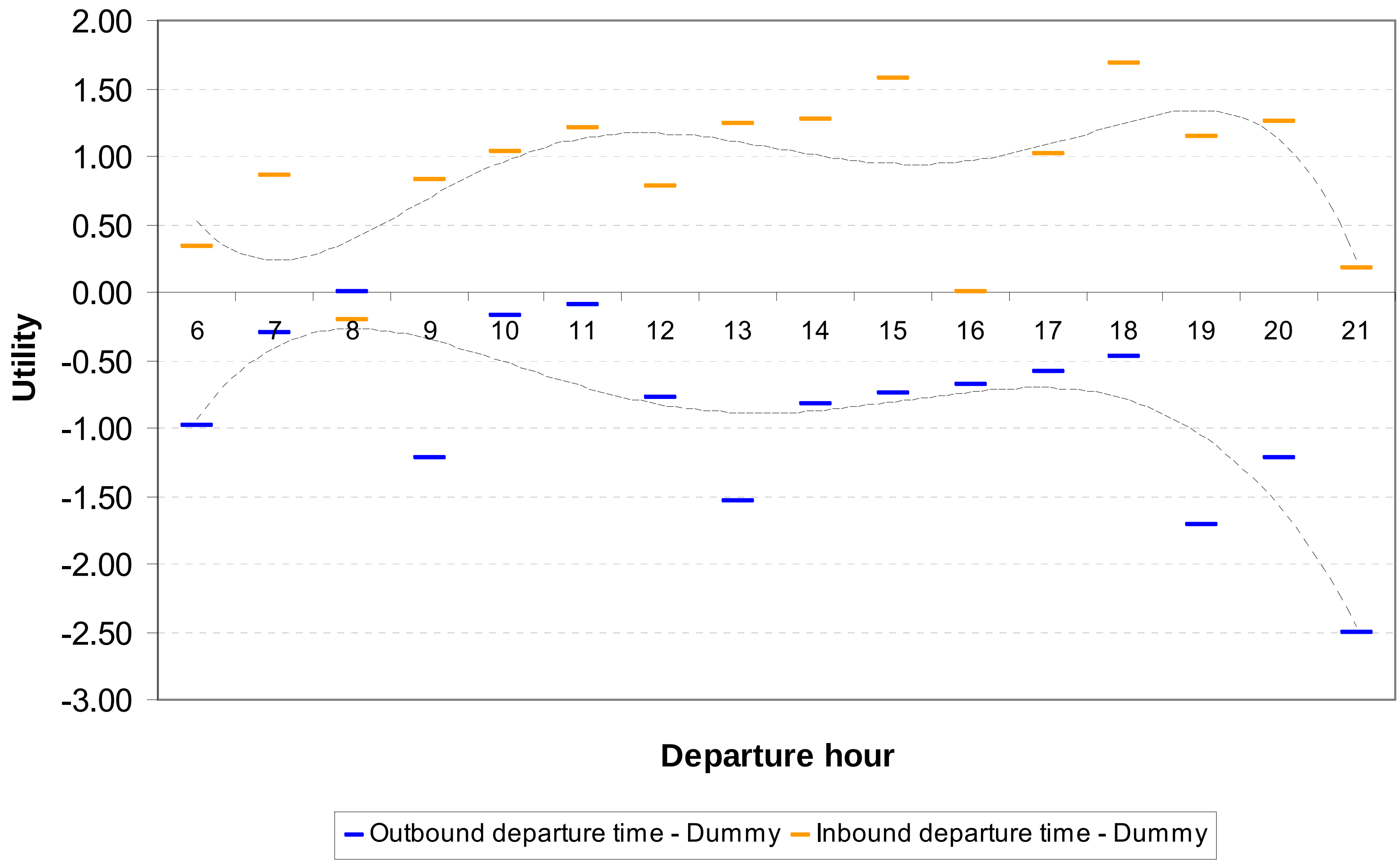
Departure hour preferences – Same day return



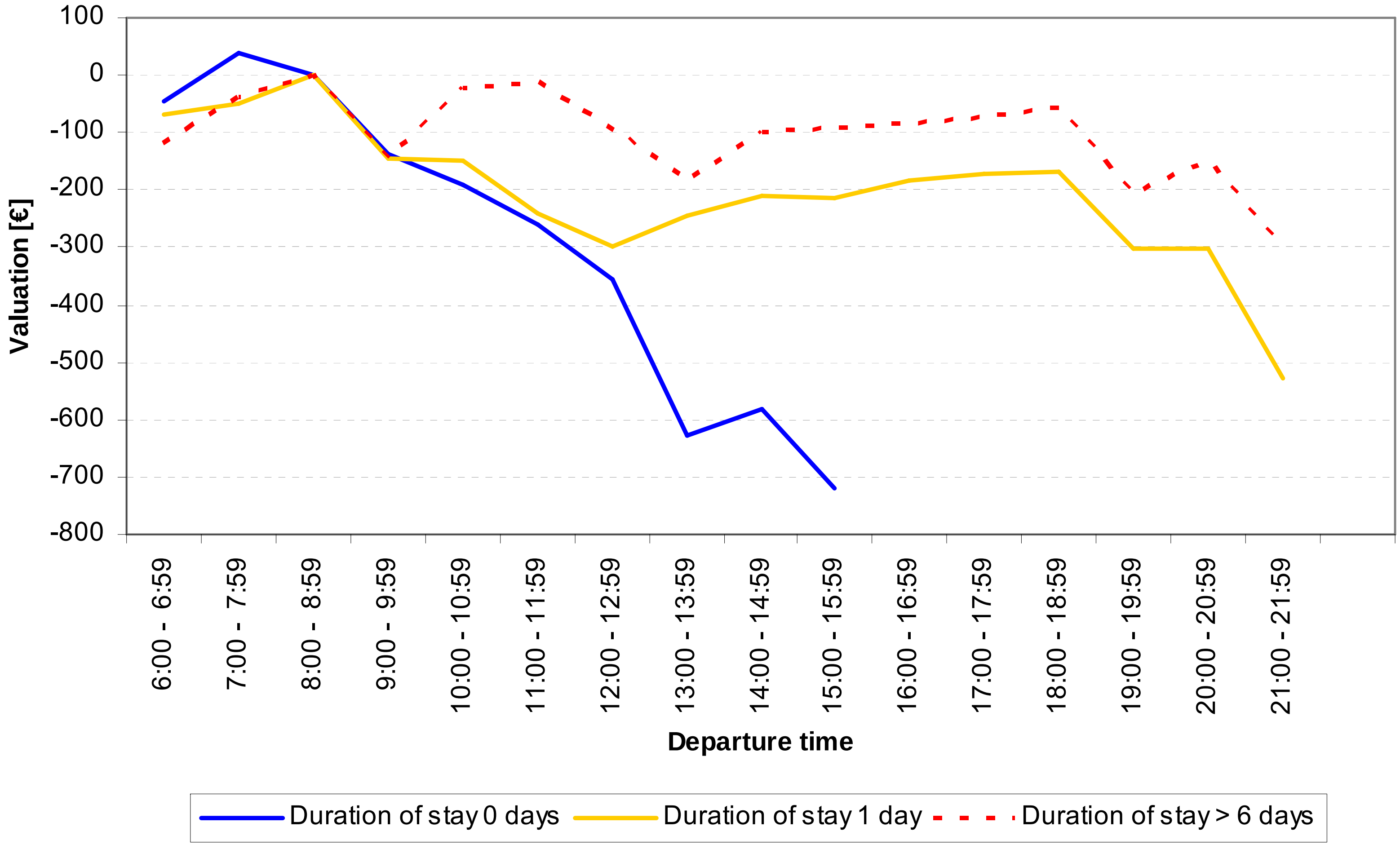
Departure hour preferences – Next day return



Returning after 6 days



Valuation of outbound departure hour



Parameter estimates with similarity measure

Variable	MNL model		MNL with IND(c)	
	Parameter	(t-test)	Parameter	(t-test)
Code share	-0.9247	(-12.84)	-0.9674	(-13.54)
Regional aircraft*	-0.1328	(-4.68)	-0.1236	(-4.31)
Propeller aircraft*	1.5388	(-14.79)	-1.4719	(-12.19)
Total travel time out	-0.0114	(-5.47)	-0.0105	(-5.13)
Transfer	-4.6923	(-12.42)	-4.6494	(-12.19)
Fare	-0.0068	(-75.79)	-0.0066	(-70.67)
Independence Measure	--	--	0.5379	(13.59)

(*) base category: Mainline jet

Conclusions

- Similar alternatives are perceived as negative
- Alternatives departing in a window of one hour are perceived as most similar
- In direct itineraries, fare and departure time play a large role
- In non-direct itineraries, transfers play an important role
- Fare and departure time sensitivity varies per duration of stay

Outlook

- Testing the influence of the individual similarity components
- More advanced model structures to overcome similarity problem
- Investigate further choice situations

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.

Modelling performance for different parameter settings

Model	Parameter for IND(c)	Average for IND(c)	Final Log-Likelihood	Adjusted r-square	Difference r-square*
MNL	--	--	-46101.7	0.3308	--
s_x=720,γ=0.25	0.1608	0.0388	-46100.5	0.3308	0.00%
s_x=720,γ=0.5	0.2194	0.0458	-46099.1	0.3308	0.01%
s_x=720,γ=0.75	0.3992	0.0626	-46089.2	0.3309	0.03%
s_x=120,γ=0.5	0.5379	0.1660	-46004.2	0.3322	0.21%
s_x=240,γ=0.5	0.2210	0.0983	-46089.6	0.3309	0.03%
s_x=360,γ=0.5	0.1804	0.0732	-46095.2	0.3308	0.01%

* Compared to MNL