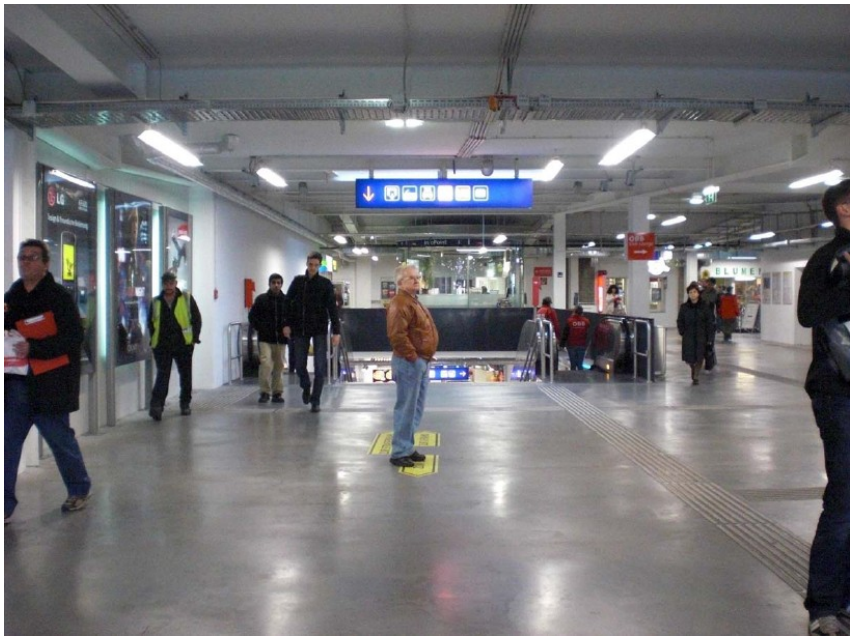
Modelling Random Taste Variations on Level Changes in Passenger Route Choice in a Public Transport Station

I. Zeiler, C. Rudloff and D. Bauer

Motivation

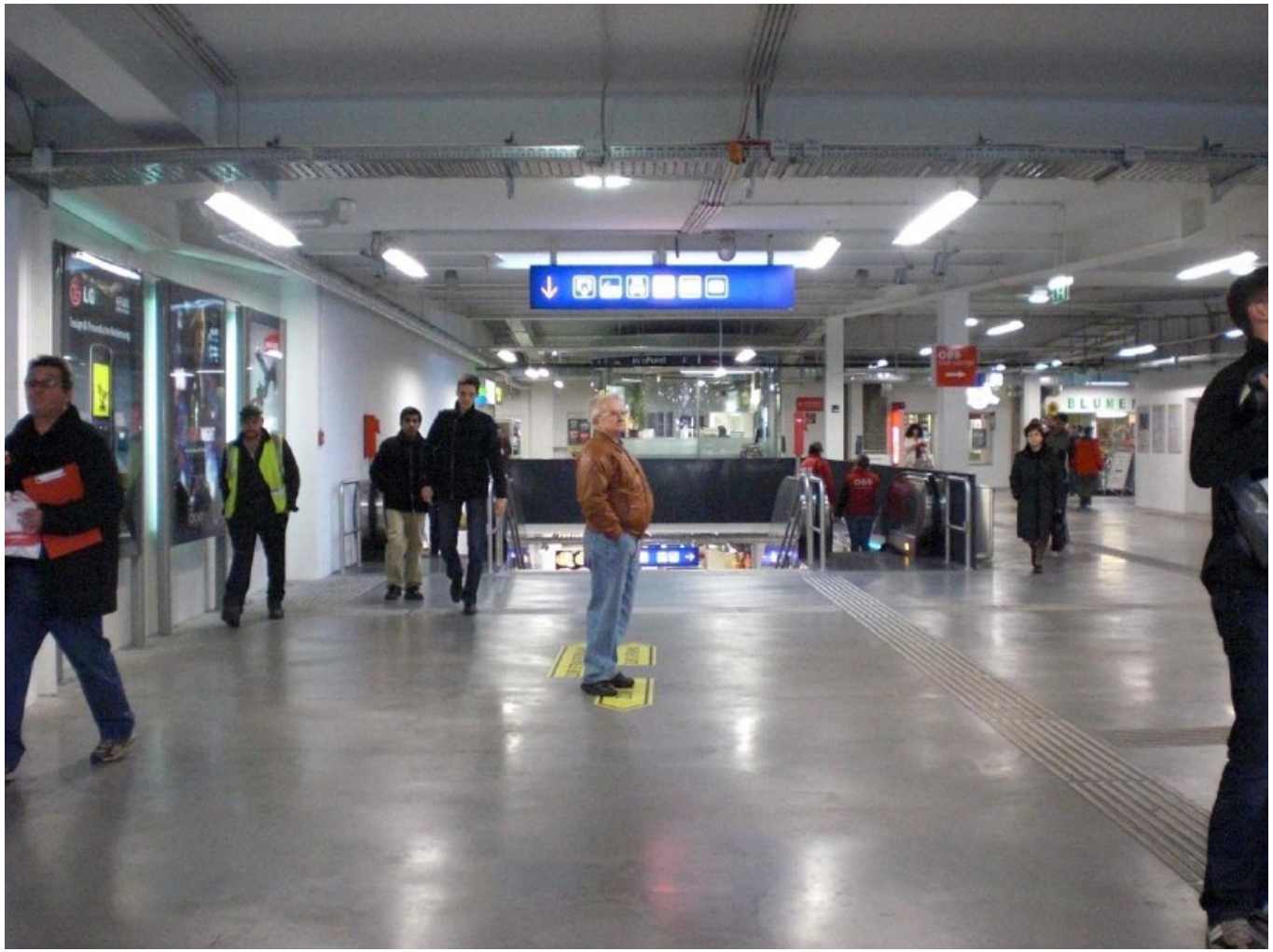
Adjacent Escalators and Stairs

- Facilities that bridge level changes are potential bottlenecks causing congestions and unnecessary delays
- *Actual* demand is of interest - which depends on distribution of passengers across different facilities



Choice Situation

Adjacent Escalators and Stairs



Questions Asked

In Substance and Methodology

- Which personal attributes influence the probability to take the stairs?
- What is the influence of ”dynamic” factors (crowding)?

- Are there random taste variations in route choice?
- Is there a value in SP data? Is there a value in combining RP and SP data?

Data Collection

Train Station ”Westbahnhof” in Vienna, Austria

- 200 passenger’s choices observed and recorded (*revealed preferences*)
- 6 hypothetical choice situations per participant (*stated preferences*)
- Resulting in a total of 1400 choice situations for estimating discrete choice models
- For modelling purposes (explanatory variables) we collected
  - 7 personal attributes
  - 2 variables to describe ”dynamic” factors (crowding)

Revealed Preference Data

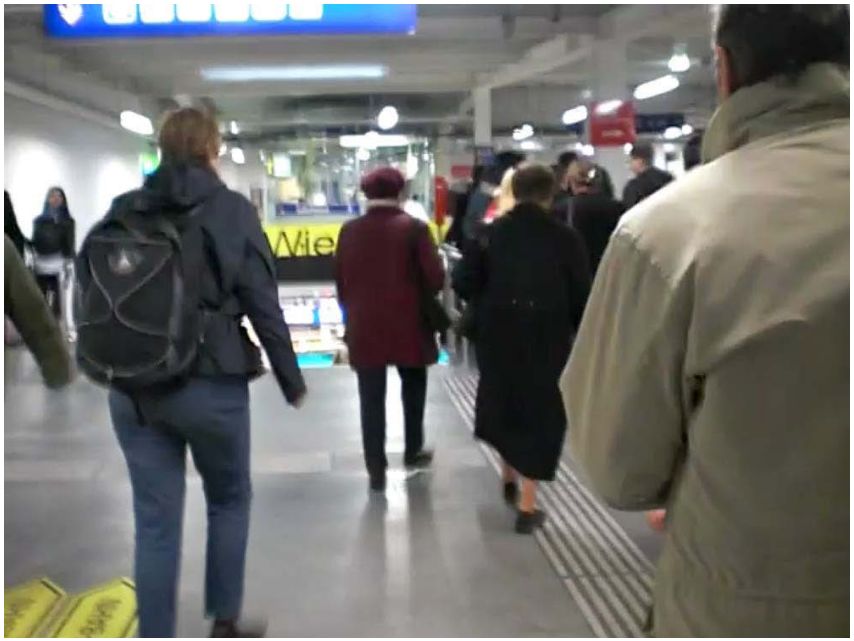
Actual choices made

- RP data are real choices, that reflect **current conditions** only
- Choices were recorded at bottom of the stairs
- Video footage from top was used to determine degree of crowding at time of choice
- The 7 personal factors were collected during an interview immediately after the choice

Stated Preference Data

Hypothetical choices

- SP questions are ”designed” to **meet research needs**, however realism of answers unknown
- Each participant was shown six different video sequences
- The sequences were shot from first-person perspective
- Exhibition of different degrees of crowding



Data Analysis: Using Discrete Choice Models

Standard Binary Logit

- The trip maker can choose from two distinct routes; the probability of choosing one alternative is modelled.
- S/he chooses the alternative with the highest utility, which consists of
  - Observable part: $V_n = \beta' x_n$ with x_n explanatory variables, β parameters to be estimated
  - Unobservable part: unobservable factors ϵ_n with ϵ iid extreme value are treated as random
- Leading to a choice probability

$$P_n = \frac{e^{V_n}}{1 + e^{V_n}}$$

Data Analysis: Using Discrete Choice Models

Standard Binary Logit II

- The values of the parameters have the same impact on all individuals in the sample - *homogeneity in taste*
- This might be a problematic simplification, since not all factors might act the same way (e.g. age)

Data Analysis: Using Discrete Choice Models

Mixed Logit

- Parameters of the explanatory variables may follow a distinct distribution $f(\beta|\theta)$
- Consequently, not the parameter values directly, but the parameters θ of the corresponding distribution are estimated
- Explanatory variables can have a different impact on different individuals - *heterogeneity in taste*
- Furthermore, it accommodates the correlation structure for repeated choice situations

Data Analysis: Using Discrete Choice Models

Joint RP and SP Data Set

- Modification to Logit is necessary
  - Real choices are influenced by unobserved factors
  - In SP respondents ought to treat all decisions similarly
  - Variance of random part of the utility might be different for RP and SP
- Introduction of a scale parameter λ:
- λ scales the utility such that variance of SP and RP is equal
- λ is estimated together with all other parameters

Explanatory Variables

Entering the Discrete Choice Models

- Personal factors
  - Gender (m: 106, f: 94)
  - Age (15-19: 7.76%, 19-30: 46.73%, 31-60: 46.73%, 60-88: 17.66%)
  - Frequency of visit (six levels between *first time visit* and *daily*)
  - Self-reported walking speed (slow, normal, fast)
  - Trip purpose, level of education, reasons for their choice
- ”Dynamic” Factors at time of choice
  - Number of persons (P) queuing with luggage
  - Number of persons (P) queuing without luggage

Results

Separate Data Sets

- Revealed Preferences Only
  - Only *P* with luggage influenced choice significantly (1.7168, t-ratio 4.2973), ASC -2.8295 (-7.8488).
  - Allowing for distributed parameters did not improve results.
- Stated Preferences Only
  - Mixed logit improved results significantly (174.9 points)
  - Parameter age is best to be normally distributed with $\mu = 0.06$ and $\sigma = 0.81$
  - Parameter is for a share of 52.8% negative for taking the stairs
  - Self-reported walking speed, *P* w luggage and *P* w/o luggage are significant and best fixed

In SP age and walking speed have an significant influence.

Comparison of Probabilities

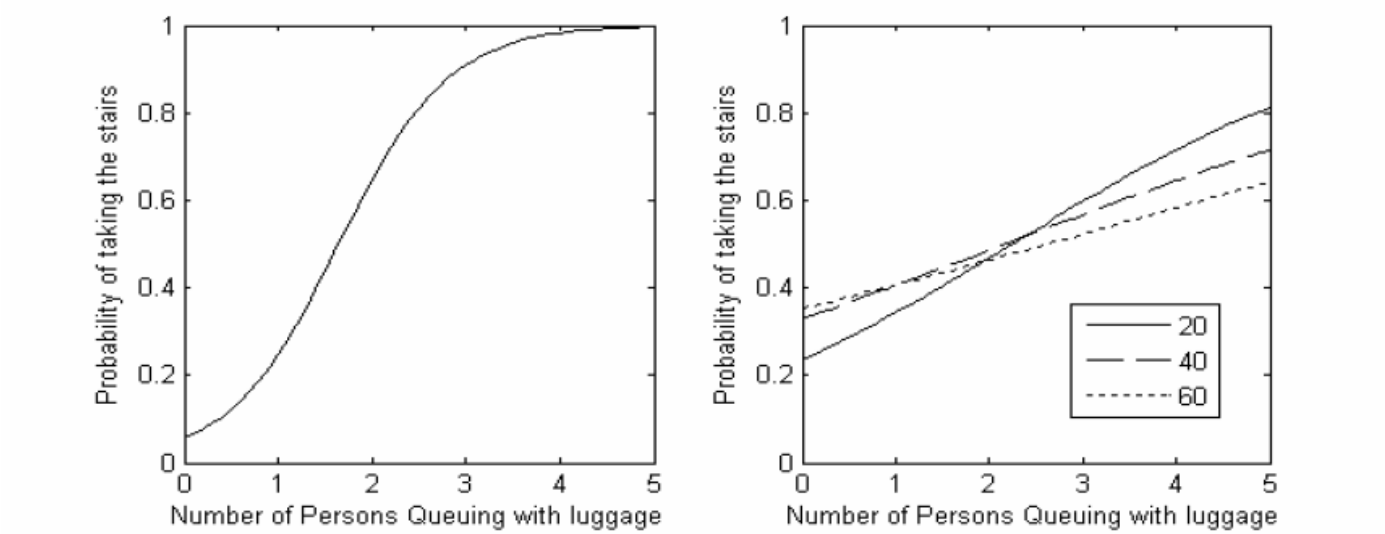


Fig. 2. Probability of taking the stairs as function of the number of persons queuing with luggage, estimated for the separate SP and RP data sets of revealed (left plot) and stated preferences (right plot). Self reported speed has been set to normal; number of persons queuing without luggage is zero. Age has a significant influence in the stated preferences only and moderates the effect of the persons queuing with luggage.

Results

Joint Data Set

- Parameter age is best to be normally distributed with $\mu = -0.264$ and $\sigma = 0.8$
- *P* w/o luggage lognormal with $\mu = 0.8038$ and $\sigma = 0.2582$
- Self-reported walking speed, *P* w luggage are significant and best fixed
- Influence of age is negative for taking the stairs for 63% and positive for 37%
- Bias: using the SP/RP model on the RP data shows an average of 33% taking the stairs, instead of the actual 11%.

Four significant parameters, two show an heterogenous influence.

Conclusions

- Crowding is an important factor, persons carrying luggage should be considered for actual capacity
- There is heterogeneity in taste, so mixed logit is appropriate and should be employed
- The use of SP data is valuable, since
  - It successfully introduced variance in choice
  - It yielded interesting results on self-assessment of participants
- The use of RP data is indispensable for realism of the models.

Thanks!

Results In Detail

SP Data Only

SP Data	Standard Logit		Mixed Logit		
	Estimate	*t-Ratio*	*Estimate*	*t-Ratio*	*Distribution*
c	-0.7061	-18.581	-2.2673	-2.7310	-
Age	-0.1751	-4.607	μ: -0.0608	-0.3447	Normal
			σ: 0.8096	9.0661	
Speed	0.4470	5.4981	0.7060	2.3557	-
#P w Luggage	0.3672	6.0795	0.7489	7.8832	-
#P w/o Luggage	0.1962	7.1087	0.4031	8.9181	-
LL(θ)	-703.1074		-528.2375		

Results In Detail

Joint Data Set

	Standard Logit		Mixed Logit		
	Estimate	*t-Ratio*	*Estimate*	*t-Ratio*	*Distribution*
c^r	-2.3982	-3.1714	-7.6442	-3.4806	-
c^s	-0.7609	-2.6365	-1.7785	-2.3014	
λ	0.9224	2.5587	2.7347	2.7523	
Age	-0.1654	-4.3412	μ: -0.264	-1.7449	Normal
			σ: 0.7994	9.0532	
Speed	0.4549	5.7655	0.689	2.3932	-
#P w Luggage	0.3774	6.4845	μ: -0.2655	-2.0129	Lognormal
			σ: 0.3154	3.1477	
#P w/o Luggage	0.1949	7.3271	0.4079	8.9846	-
LL(θ)	-767.8122		-585.3091		