Integration of human factors in pedestrian crossing choice models

Eleonora Papadimitriou, PhD

Workshop on ‘Observation and Modelling of Pedestrian Behaviour - IFSTTAR, 4 May 2014

Background

- Analysis of pedestrians crossing behavior in urban areas:
  - Understanding the way pedestrians interact with the road and traffic environment, other pedestrians,
  - Understanding the way they balance the need for comfort and safety with the cost of delays, within existing traffic rules.
- Better adjustment of urban road networks to pedestrians' needs.
- Signalized junctions provide pedestrians a protected crossing phase
- Mid-block crossing and diagonal crossing are common practice among pedestrians aiming to save travel time
- Pedestrians experience smaller delays compared to other road users, but increased road accident risk

Existing models

- Gap acceptance models, in which each pedestrian is associated with a critical gap for road crossing.
- Level of service approach, in which the difficulty to cross is used as a measure of for pedestrian level of service
- Crossing choices among a set of discrete alternatives are often modeled on the basis of utility theory.

- A distinct part of existing research on pedestrian crossing behavior is devoted to analyses of psychological, attitudinal, perceptual and motivational factors.
- Human factors are seldom incorporated in pedestrian behavior and safety models

Objectives

The analysis of pedestrians' crossing behavior along entire trips in urban road networks in relation to road, traffic and human factors.

- Use data from a dedicated survey combining observed behaviour and declared attitudes, perceptions, motivations, behaviours etc.
- Develop choice models for estimating the probability to cross at each location along a pedestrian trip in relation to roadway design, traffic flow and traffic control
- Introduce and integrate human factors in the choice models, as latent variables (measured through sets of indicators).

Survey scenarios

- Pedestrians were followed along urban trips, and their crossing behavior was recorded, together with features of the road environment and the traffic conditions.

- Crossing an main urban road with signal controlled and uncontrolled crosswalks: scenarios (i) and (viii);
- Crossing a minor (residential) road with or without marked crosswalks: scenarios (ii), (v), (vi) and (vii);
- Crossing a major urban arterial with signal controlled crosswalks: scenarios (iii) and (iv).

Observed crossing behaviour data

- Static data: characteristics of the trips, street names, road geometry and traffic control available;

- Dynamic data: the walking and crossing characteristics of the participants, recorded in real time conditions while following the pedestrian:
  - Data recorded for each road link, e.g. walking time and length, traffic volume, number and duration of crossing attempts etc.
  - Additional data recorded for road links with a primary crossing, e.g. crossing location (junction or mid-block), crossing type (diagonal), signal display (red / green), following another pedestrian etc.

Declared attitudes and behaviours

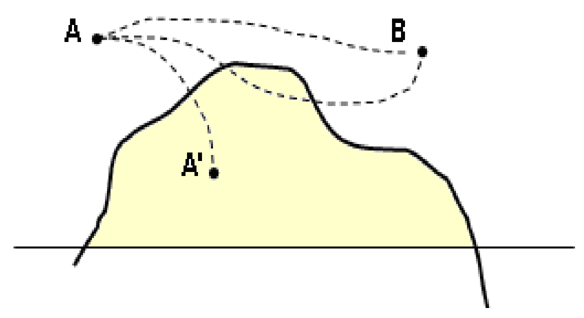
B	**How many times per week do you travel by each one of the following modes*:**
B1_i	Public transport (metro, bus, trolley bus, tramway)
B1_ii	Pedestrian
B1_iii	Passenger car (driver or passenger)
	Last week, how many kilometers did you travel by each one of the following modes:**
B2_i	Passenger car (driver or passenger)
B2_ii	Pedestrian
B2_iii	Public transport (metro, bus, trolley bus, tramway)
	As a pedestrian, how much would you agree with each one of the following statements*:**
B3_i.	I walk for the pleasure of it
B3_ii	I walk because it is healthy
B3_iii	In short trips, I prefer to walk
B3_iv	I prefer taking public transportation (buses, metro, tramway, etc.) than my car
B3_iv	I walk because I have no other choice
C	**As a pedestrian, how much would you agree with each one of the following statements***:**
C1_i.	Crossing roads is difficult
C1_ii.	Crossing roads outside designated locations increases the risk of accident
C1_iii.	Crossing roads outside designated locations is wrong
C1_iv	Crossing roads outside designated locations saves time
C1_v	Crossing roads outside designated locations is acceptable because other people do it
C2_i	I prefer routes with signalized crosswalks
C2_ii	I try to make as few road crossings as possible
C2_iii	I try to take the most direct route to my destination
C2_iv	I prefer to cross diagonally
C2_v	I try to take the route with least traffic to my destination
C2_vi	I am willing to make a detour to find a protected crossing
C2_vii	I am willing to take any opportunity to cross
C2_viii	I am willing to make dangerous actions as a pedestrian to save time
D	Compared to other pedestrians, how much do you agree that***:
D_i	I am less likely to be involved in a road crash than other pedestrians
D_ii	I am faster than other pedestrians
D_iii	I am more careful than other pedestrians

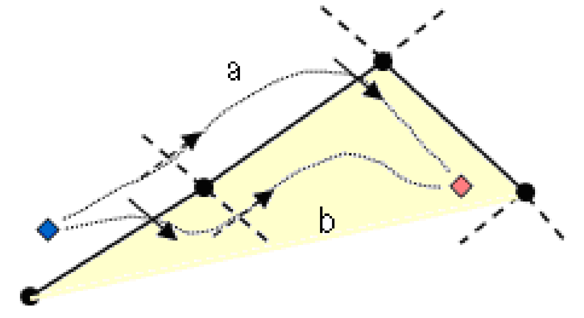
Declared attitudes and behaviours (cont.)

E	As a pedestrian, how often do you adopt each one of the following behaviors****:
E1_i.	I cross diagonally
E1_ii	I cross at midblock at major urban arterials
E1_iii	I cross at midblock at urban roads
E1_iv	I cross at midblock in residential areas
E1_v	I cross at midblock when I am in a hurry
E1_vi	I cross at midblock when there is no oncoming traffic
E1_vii	I cross at midblock when I see other people do it
E1_viii	I cross at midblock when my company prompts me to do it
E1_ix	I prompt my company to cross at midblock
E1_x	I cross at midblock when there is a shop I like on the other side
E1_xi	I cross even though the pedestrian light is red
E1_xii	I walk on the pavement rather than on the sidewalk
E2_i	I cross between vehicles stopped on the roadway in traffic jams
E2_ii	I cross without paying attention to traffic
E2_iii	I am absent-minded while walking
E2_iv	I cross while talking on my cell phone or listing to music on my headphones
E2_v	I cross even though obstacles (parked vehicles, buildings, trees, etc.) obstruct visibility
E2_vi	I cross even though there are oncoming vehicles
F	**As a pedestrian, how much would you agree with each one of the following statements***:**
F1_i	Drivers are not respectful to pedestrians
F1_ii	Drivers drive too fast
F1_iii	Drivers are aggressive and careless
F1_iv	Drivers should always give way to pedestrians
F1_v	When there is an accident, it is the driver's fault most of the times
F1_vi	I let a car go by, even if I have right-of-way

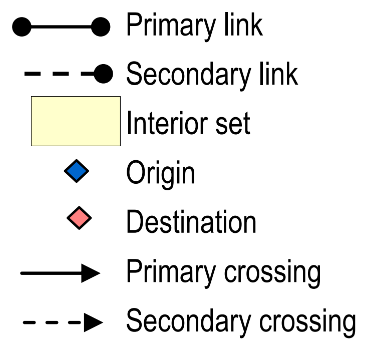
Parameterisation of crossing behaviour

- A topological analysis of pedestrian trajectories on the urban road network (based on the 'Jordan curve theorem')
- Primary crossings are defined in previous research as crossings that take place across the pedestrian trajectory and their choice is stochastic (i.e. pedestrian may choose from a number of alternative locations).
- There are other crossings whose choice is deterministic, referred to as 'secondary' crossings.





- ***The survey scenarios were designed so that only one crossing of interest will take place for each scenario, namely a 'primary' crossing.***



The ‘classical’ choice model

- For each road link of each walking scenario, different options are available (choice utilities) e.g. “cross at mid-block”, “cross at junction”, “not cross at all”.
- The utility of each alternative is conditional on the availability of the alternative

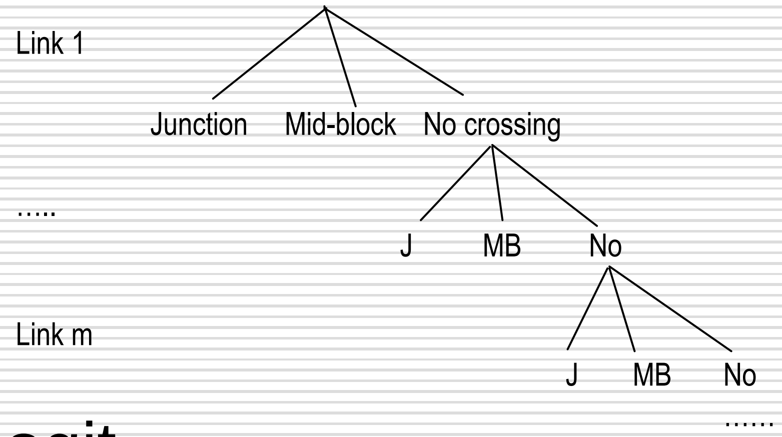
$U_{in} = V_{in} + \varepsilon_{in}$

$V_{in} = \beta X_{in}$ systematic part of the utility, ε_{in} stochastic part Extreme Value distributed $\sim(0,\mu)$

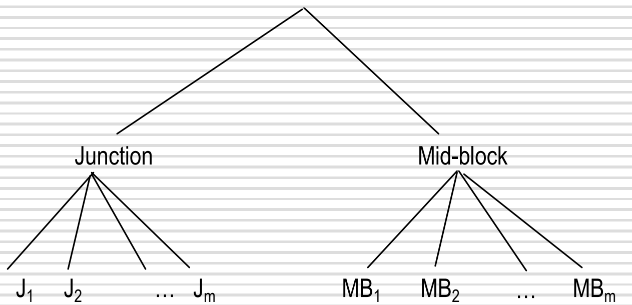
- Variables can be:
  - generic, with a common B coefficient for all alternatives, (typically, characteristics of the choice maker)
  - alternative-specific, i.e. with different B coefficients for each alternative, (typically, characteristics (‘attributes’) of the alternatives.
- A random ‘panel’ effect can be examined, in order to capture heterogeneity due to unobserved differences between respondents.

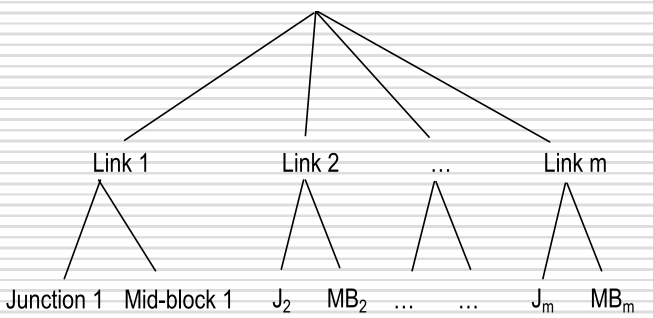
Pedestrian choice hypotheses

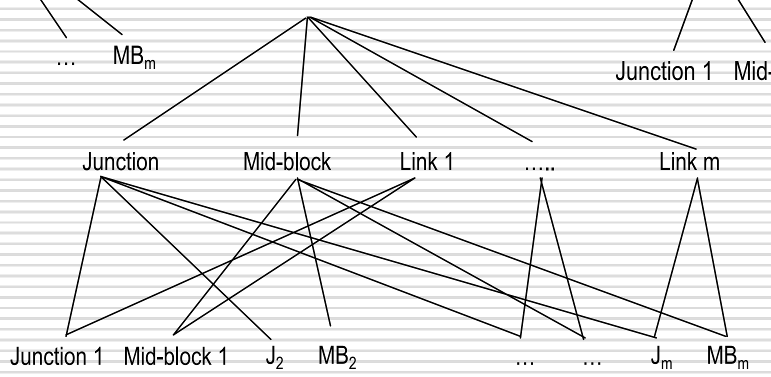
- Sequential choice:
  - Multinomial logit



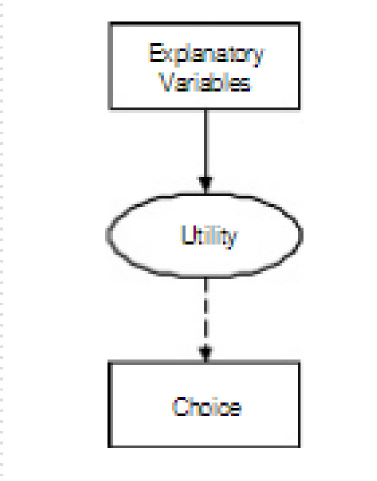
- Hierarchical choice
  - Nested or cross-nested logit



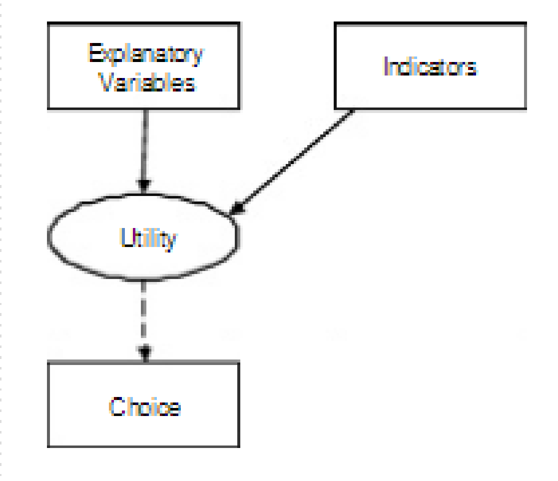




Steps for an integrated model



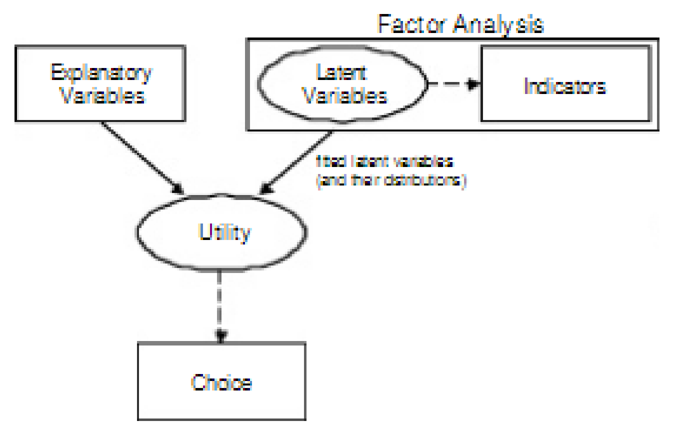
a. Choice model



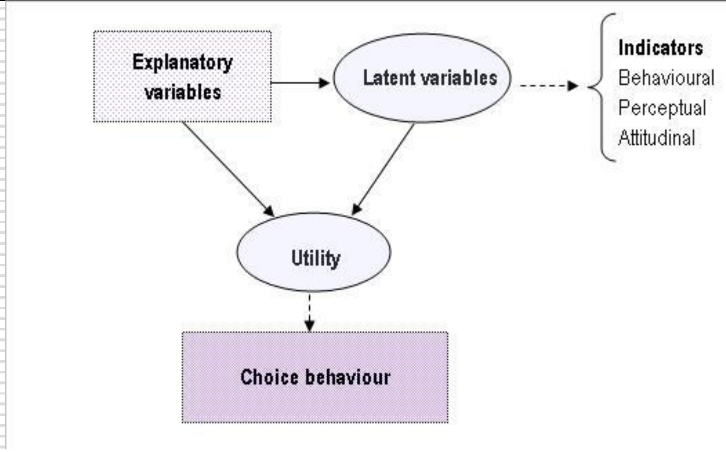
b. Choice model with indicators directly included in utility

- In Figure a, important latent variables are omitted, leading to the standard discrete choice model, in which parameters estimates may be inconsistent.
- In Figure b, the observed variables (indicators) may be directly inserted in the choice model; however they are highly correlated, they are not causal, and they are highly dependent on the phrasing of the survey question.

Steps for an integrated model (cont.)



c. Sequential estimation: PCA followed by a choice model



d. Integrated choice and latent variables model

- In Figure c, a two-stage approach: a principal component analysis to estimate the latent variables “components”, and their (mean) scores are introduced in the choice model. Their variance is not included, leading to measurement errors and inconsistent estimates.
- In Figure d, the latent variable model is composed of a group of “structural equations” describing the latent variables as a function of observable exogenous variables, and a group of “measurement equations”, linking the latent variables to the observable indicators. The key feature is that the latent variables can be calculated from the observable variables once the model parameters are estimated (integration).

Example (binary choice, 4 latent variables)

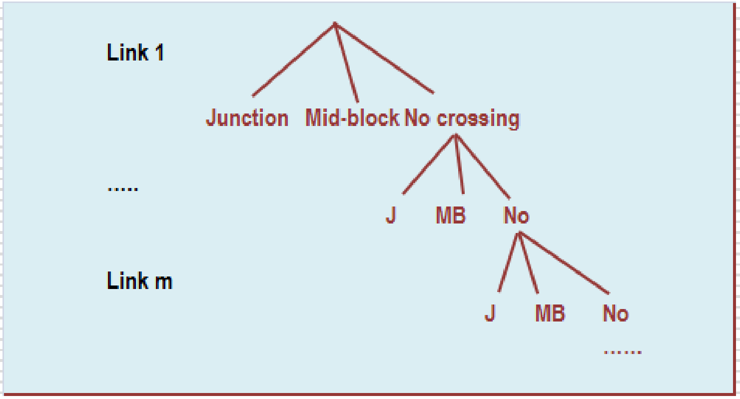
Integrated choice – latent variables model	**Latent Variables model**
Structural equations $U_{in} = b'X_{in} + c_1\tilde{Z}_{1n} + c_2\tilde{Z}_{2n} + c_3\tilde{Z}_{3n} + c_4\tilde{Z}_{4n} + \varepsilon_{in}$ $U_{jn} = b'X_{jn} + \varepsilon_{jn}$ Measurement equation $y_n = \begin{cases} 1, if\ U_{in} > U_{jn} \\ 0, otherwise \end{cases}$	Structural equations $Z_{1n} = W_{1n}\lambda_1 + \omega_{1n}$ $Z_{2n} = W_{2n}\lambda_2 + \omega_{2n}$ $Z_{3n} = W_{3n}\lambda_3 + \omega_{3n}$ $Z_{4n} = W_{4n}\lambda_4 + \omega_{4n}$ Measurement equations $I_{1rn} = a_{1r}Z_{1n} + \upsilon_{1rn}, r = 1,2,3,4$ $I_{2kn} = a_{2k}Z_{2n} + \upsilon_{2kn}, k = 1,2,3,4$ $I_{3ln} = a_{3l}Z_{3n} + \upsilon_{3ln}, l = 1,2,3,4$ $I_{4mn} = a_{4m}Z_{4n} + \upsilon_{4mn}, m = 1,2,3,4$

- U_{in}, U_{jn} denote the utility of each alternative respectively, for individual *n*;
- X_{in}, X_{jn} are sets of observed variables;
- $Z_{1n}, Z_{2n}, Z_{3n}, Z_{4n}$ are the latent variables (actually the components accounting for most of the variability of the respective latent variables)
- $I_{1n}, I_{2n}, I_{3n}, I_{4n}$ are sets of the indicators of the latent variables $Z_{1n}, Z_{2n}, Z_{3n}, Z_{4n}$ respectively;
- $\tilde{Z}_{1n}, \tilde{Z}_{2n}, \tilde{Z}_{3n}, \tilde{Z}_{4n}$ are the fitted values of the latent variables, once they are estimated by the structural equations of the latent variable model;
- $W_{1n}, W_{2n}, W_{3n}, W_{4n}$ are sets of observed variables (characteristics of respondent *n*);
- $\omega_{1n}, \omega_{2n}, \omega_{3n}, \omega_{4n}$ and $\upsilon_{1rn}, \upsilon_{2kn}, \upsilon_{3ln}, \upsilon_{4mn}$ are sets of (normally distributed) errors;
- $b, a_{1r}, a_{2k}, a_{3l}, a_{4m}$, are sets of unknown parameters;

$\lambda_1, \lambda_2, \lambda_3, \lambda_4, c_1, c_2, c_3, c_4$ are unknown parameters;

Results: the choice model

- Sequential choice behaviour
- Mixed sequential logit model (with random heterogeneity)



Utility functions		
0 (cross at mid-block)	=	ASC0 * one + B0_first * first + B0_majorroad * majorroad + B0_secondaryroad * secondaryroad + B0_minorroad * minorroad + B0_trafficempty * trafficempty + B0_trafficlow * trafficlow + B0_traffichighcong * traffichighcong + ZERO [SIGMA] * one
1 (cross at junction)	=	ASC1 * one + B1_first * first + B1_signal * L_signal + B1_barriers * L_barriers + ZERO [SIGMA] * one
2 (no crossing)	=	ASC2 * one

Results: estimation of human 'factors'

PEDMOD

- Categorical Principal Component Analysis (CATPCA)
- Three components of pedestrian behaviour:
  - Risk taker & optimiser
  - Conservative & public transport user
  - Pedestrian for pleasure

Component 1: Risk taker & optimizer	Loadings
Crossing roads outside designated locations increases the risk of accident	-0.568
Crossing roads outside designated locations is wrong	-0.509
Crossing roads outside designated locations is acceptable because other people do it	0.418
I prefer to cross diagonally	0.633
I am willing to make a detour to find a protected crossing	-0.564
I am willing to take any opportunity to cross	0.636
I am willing to make dangerous actions as a pedestrian to save time	0.526
I am faster than other pedestrians	0.473
I cross diagonally	0.674
I cross at midblock at major urban arterials	0.579
I cross at midblock at urban roads	0.739
I cross at midblock in residential areas	0.723
I cross at midblock when I am in a hurry	0.825
I cross at midblock when there is no oncoming traffic	0.602
I cross at midblock when I see other people do it	0.467
I cross at midblock when my company prompts me to do it	0.575
I prompt my company to cross at midblock	0.746
I cross even though the pedestrian light is red	0.593
I cross between vehicles stopped on the roadway in traffic jams	0.658
I cross even though obstacles (parked vehicles, buildings, trees, etc.) obstruct visibility	0.548
I cross even though there are oncoming vehicles	0.683

Component 2: Conservative & public transport user	Loadings
Weekly travel by Public transport	0.698
Weekly travel by Pedestrian	0.470
Weekly travel by Passenger car	-0.534
Weekly Km of travel by Passenger car	-0.475
Weekly Km of travel by Public transport	0.724
I prefer taking public transportation than my car	0.493
Crossing roads is difficult	0.558
I try to make as few road crossings as possible	-.463
I prefer to cross diagonally	-.503
I am less likely to be involved in a road crash than other pedestrians	-.452

Component 3: Pedestrian for pleasure	Loadings
Weekly travel by Pedestrian	0.570
Weekly travel by Passenger car (driver or passenger)	-0.593
WeeklyKm of travel by Passenger car (driver or passenger)	-0.534
WeeklyKm of travel by Pedestrian	0.583
I walk for the pleasure of it	0.562
I walk because it is healthy	0.628
I prefer routes with singalised crosswalks	0.419
I am willing to make a detour to find a protected crossing	.417
I cross at midblock when there is a shop I like on the other side	.425
When there is an accident, it is the driver's fault most of the times	.478

Introducing human factors in the choice

Utility functions		
0 (cross at mid-block)	=	ASC0 * one + B0_first * first + B0_majorroad * majorroad + B0_secondaryroad * secondaryroad + B0_minorroad * minorroad + B0_trafficempty * trafficempty + B0_trafficlow * trafficlow + B0_traffichighcong * traffichighcong + B0_comp1 * Comp1 + B0_comp3 * Comp3 + ZERO [SIGMA] * one
1 (cross at junction)	=	ASC1 * one + B1_first * first + B1_signal * L_signal + B1_barriers * L_barriers + ZERO [SIGMA] * one
2 (no crossing)	=	ASC2 * one

Utility parameters				
Name	**Value**	**Std. error**	**t-test**	**P-value**
ASC0	-3.890	0.457	-8.510	0.000
ASC1	-2.040	0.230	-8.880	0.000
ASC2	0.000	--fixed--		
B0_comp1	0.201	0.107	1.880	0.060
B0_comp3	-0.161	0.114	-1.410	0.160
B0_first	0.893	0.252	3.550	0.000
B0_majorroad	0.000	--fixed--		
B0_minorroad	0.631	0.300	2.100	0.040
B0_secondaryroad	1.630	0.374	4.370	0.000
B0_trafficempty	1.360	0.395	3.450	0.000
B0_traffichighcong	0.000	--fixed--		
B0_trafficlow	0.664	0.317	2.100	0.040
B1_barriers	0.936	0.205	4.570	0.000
B1_first	0.978	0.206	4.750	0.000
B1_signal	0.177	0.177	1.000	0.320
SIGMA	-0.371	0.122	-3.050	0.000
ZERO		--fixed--		

Variance of normal random coefficients			
Name	**Value**	**Std.error**	**t-test**
ZERO_SIGMA	0.138	0.104	1.320

Model's fit			
Number of estimated parameters	13	Nulllog-likelihood	-1043.86
Number of observations	1048	Finallog-likelihood	-812.475
Numberofindividuals	74	Likelihoodratiotest	461.223

Towards the integration

- Latent variable **'risk taking & optimising' ('Risk')** measurement equations
  - On the basis of PCA results
  - Example: *C2_vii "I am willing to take any opportunity to cross", E1_iii "I cross at mid-block at urban roads"*

$$I\_C2\_vii = a_1 * \text{'risk'} + r_1 * u_1$$

$$I\_E1\_iii = a_2 * \text{'risk'} + r_2 * u_2$$

- Latent variable structural equation
  - Risk taking & optimising indicators are correlated with pedestrian speed, age, gender, income

$$\text{'Risk'} = \lambda_1 * speed + \lambda_2 * age + \ldots + \omega$$

Towards the integration

- Choice utility

  - Integrates the latent variable ‘risk’

Utility functions		
0 (cross at mid-block)	=	ASC0 * one + B0_first * first + B0_majorroad * majorroad + B0_secondaryroad * secondaryroad + B0_minorroad * minorroad + B0_trafficempty * trafficempty + B0_trafficlow * trafficlow + B0_traffichighcong * traffichighcong + **B0_Risk** * Risk + ZERO [SIGMA] * one
1 (cross at junction)	=	ASC1 * one + B1_first * first + B1_signal * L_signal + B1_barriers * L_barriers + ZERO [SIGMA] * one
2 (no crossing)	=	ASC2 * one

  - Estimated by structural and measurement equations simultaneously with the choice model

Next steps

- Development of basic latent variable models
- Dedicated optimization package (pythonbiogeme)

- Testing simple latent variables specifications upon the classical choice model
- More sophisticated models depending on first results