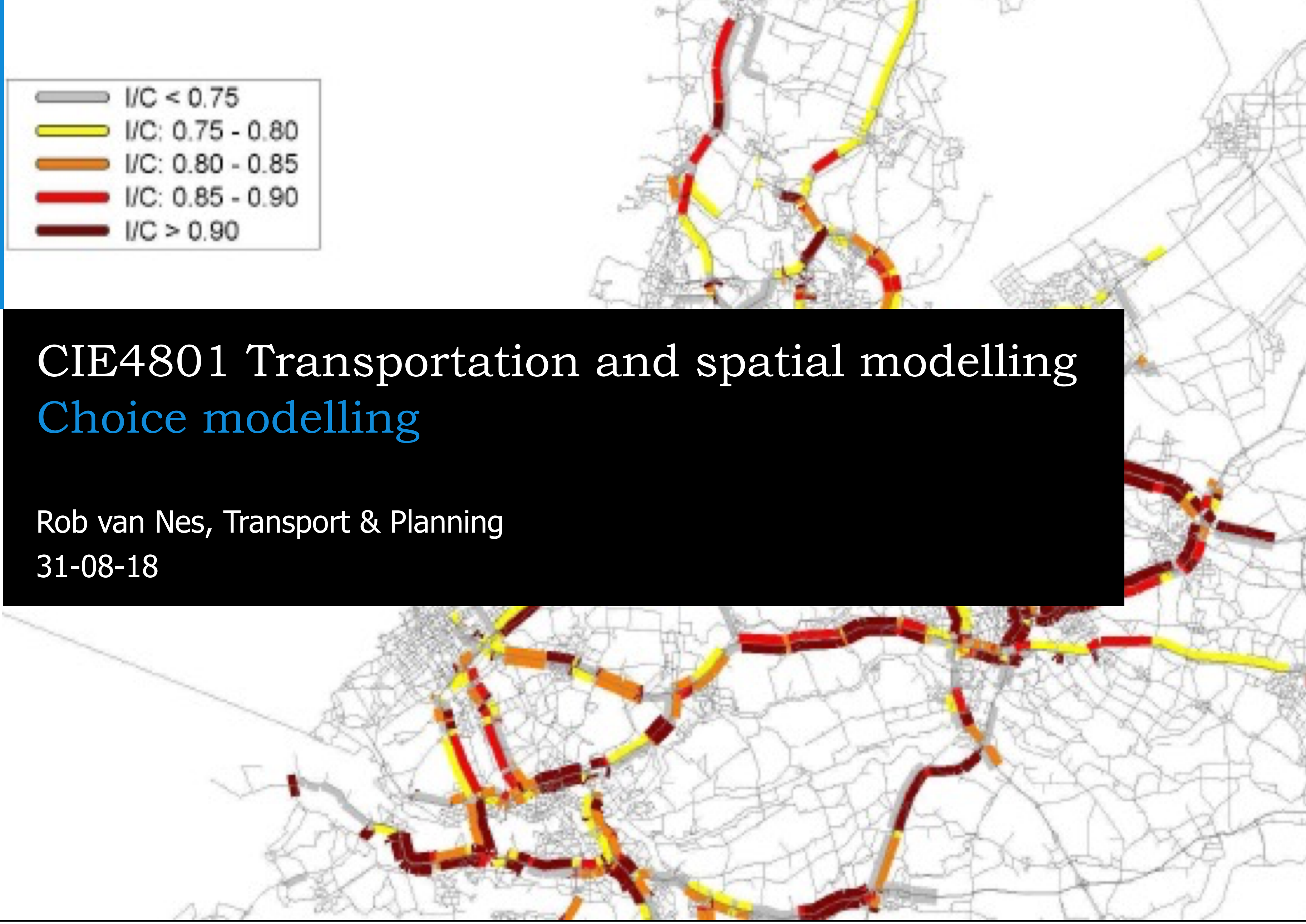
CIE4801 Transportation and spatial modelling
Choice modelling

Rob van Nes, Transport & Planning
31-08-18



Agenda

- Questions previous lecture
- Data on travel behaviour
- Modelling travel choice behaviour
  - Discrete choice modelling

2.

Data on travel behaviour

Data needed for modelling

- Zonal data
- Network data

- Data from other models
  - E.g. regional model as input/constraint for an urban model
  - OD-matrix trucks from a freight transport model

- Date for modelling travel behaviour

- Data for modelling travel choice behaviour

Data sources

- Traffic/Passenger counts
  - Road
  - Public transport
- Surveys
  - Roadside
  - Public transport
  - License plate
  - Household
- New data sources
  - Cell phones
  - Route planners
  - Chip cards

Counts versus surveys

- Counting seems simple
  - In practice quite a difference in quality
  - Limited number of locations
- Just numbers, no information on traveller
- Surveys focus on travellers
  - Road side surveys or PT surveys are still limited
  - Limited number of locations
- Household (or person) survey are most informative

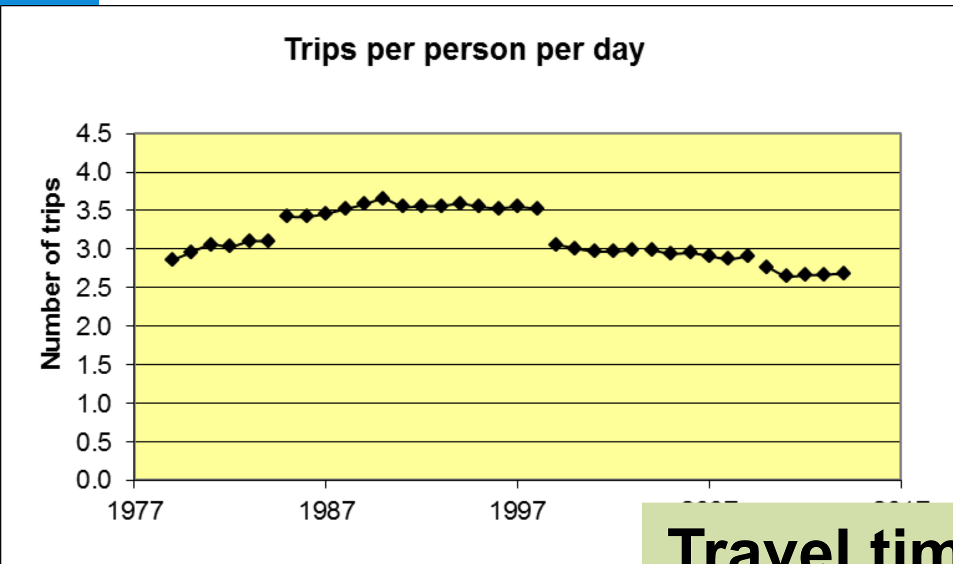
Example travel pattern from a survey

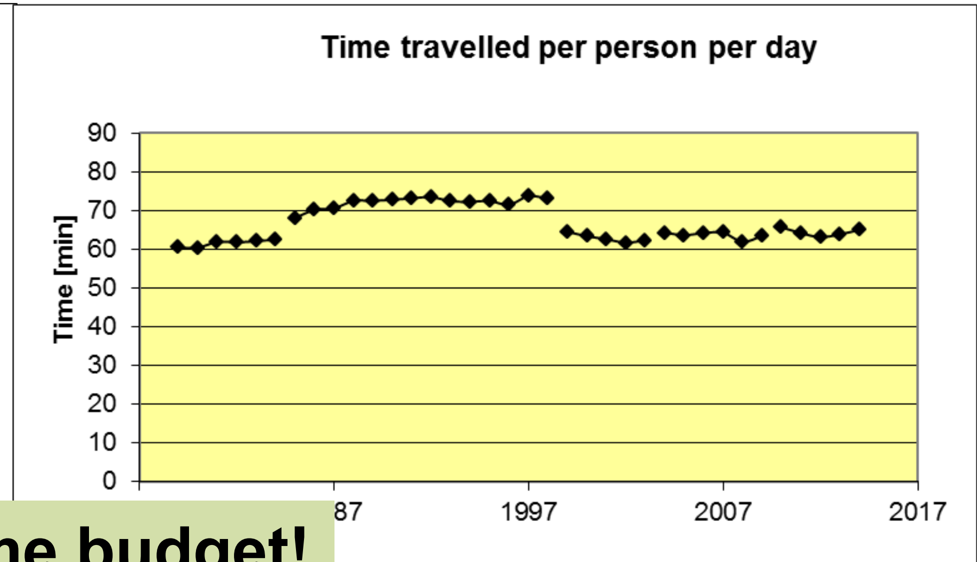
Departure time	Destination	Modes	Distance	Arrival time
06:32	To work	Bike, train, walk	28 km	7:20
	1. Train station	*Bike*	*2 km*	*6:40*
	2. Train station	*Train*	*25 km*	*7:10*
	3. Work	*Walk*	*1 km*	*7:20*
14:30	To home	Walk, train, bike	28 km	15:18
	1. Train station	*Walk*	*1 km*	*14:40*
	2. Train station	*Train*	*25 km*	*15:10*
	3. Home	*Bike*	*2 km*	*15:18*
15:23	Pick up kid from school	Walk	0,6 km	15:30
15:35	To home	Walk	0,6 km	15:44
19:27	Tour with a friend	Bike	19 km	20:43
23:55	Walk the dog	Walk	2 km	0:20

Survey issues

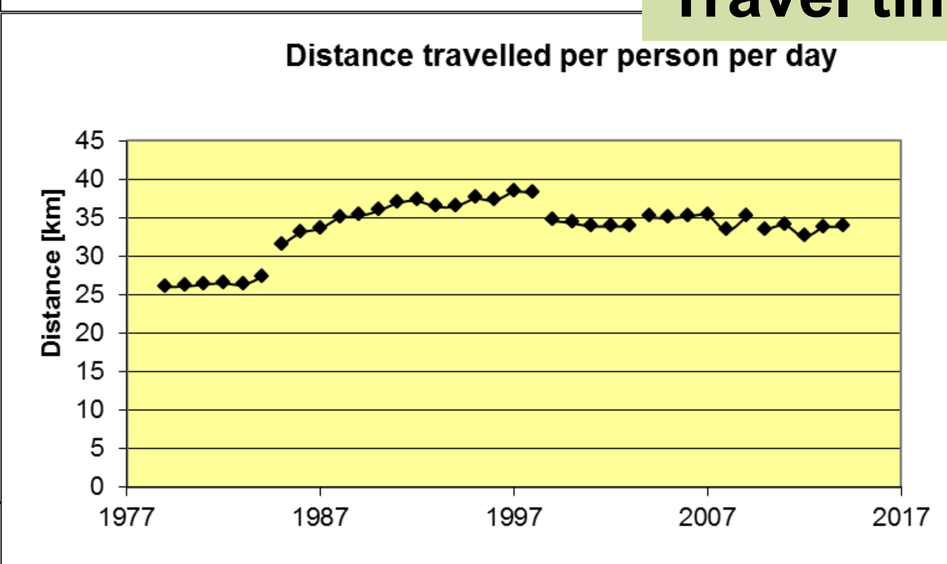
- Non-response / Non-reporting
  - Persons / specific trips
- In-/excluding kids <12 year
- Inconsistency in definitions and phrasing of the questions over time
- Splitting roundtrips or not
- Registration of frequent (professional) trips
- Pedestrians/cyclists

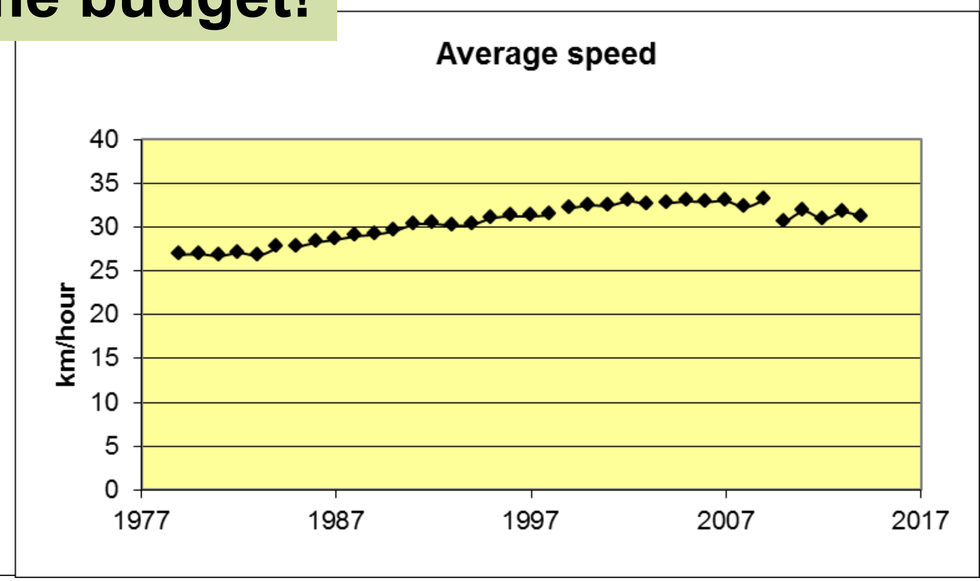
Travel characteristics Netherlands (MON)



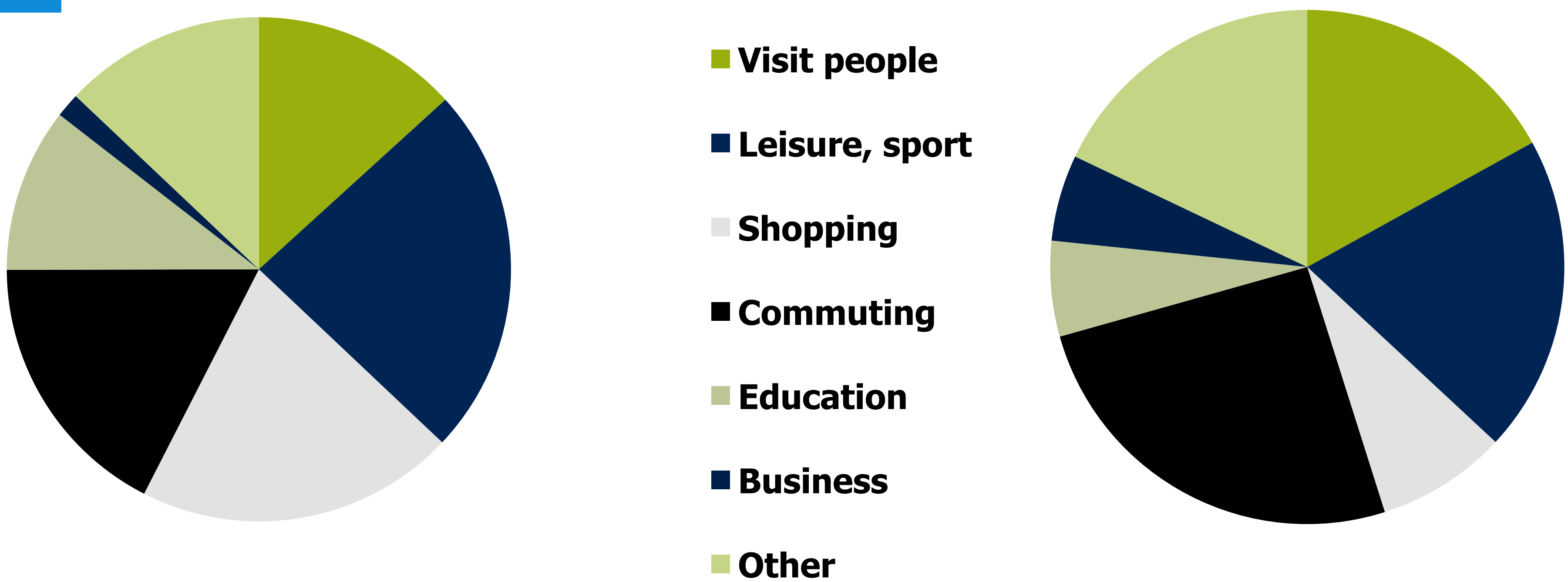


Travel time budget!



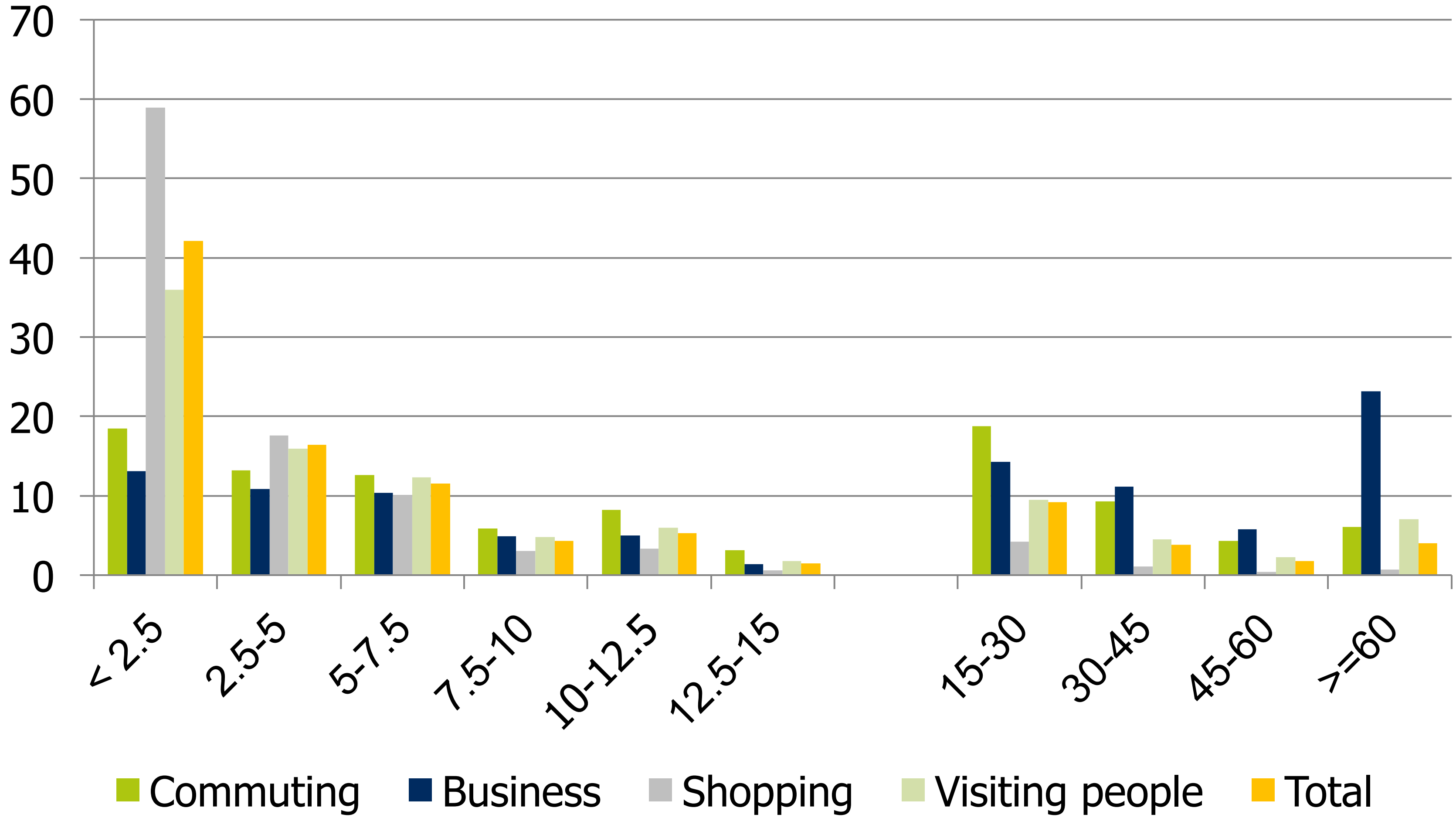


Trip purpose (Netherlands)
Trips and trip kilometres for an average day

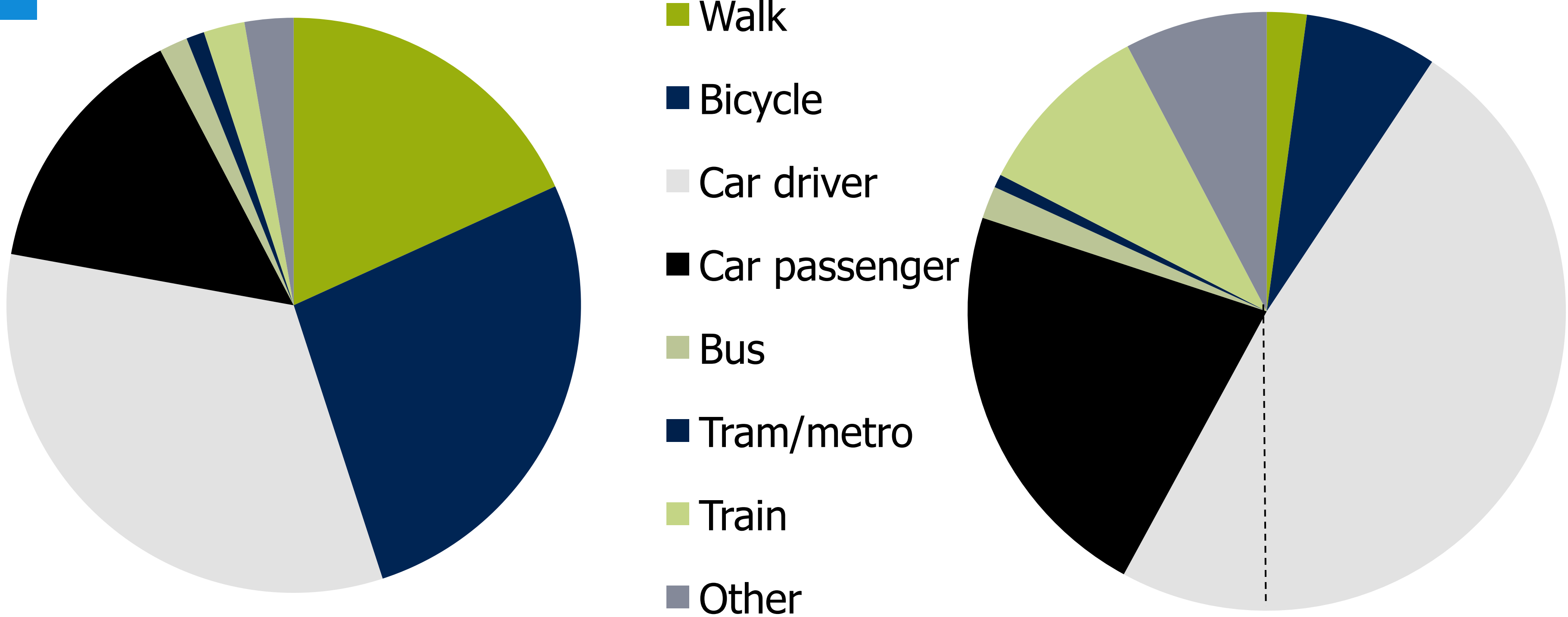


Trip purpose is defined by the activity at the destination, except when the destination is home, then the activity at the origin is decisive

Trip length distributions (Netherlands)



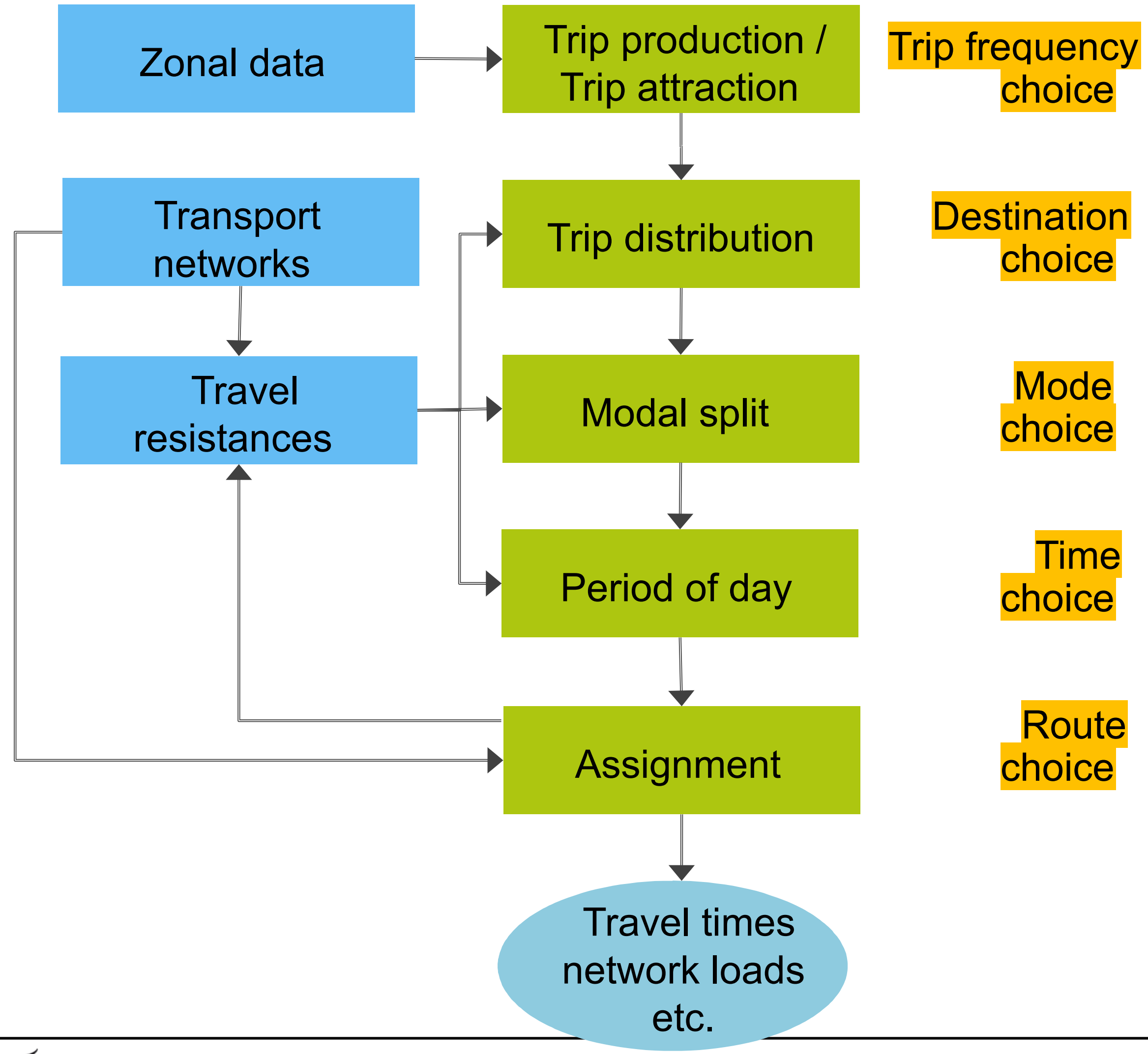
Modal split (Netherlands)
Trips and trip kilometres for an average day



3.1

Discrete choice modelling

Framework for transport modelling

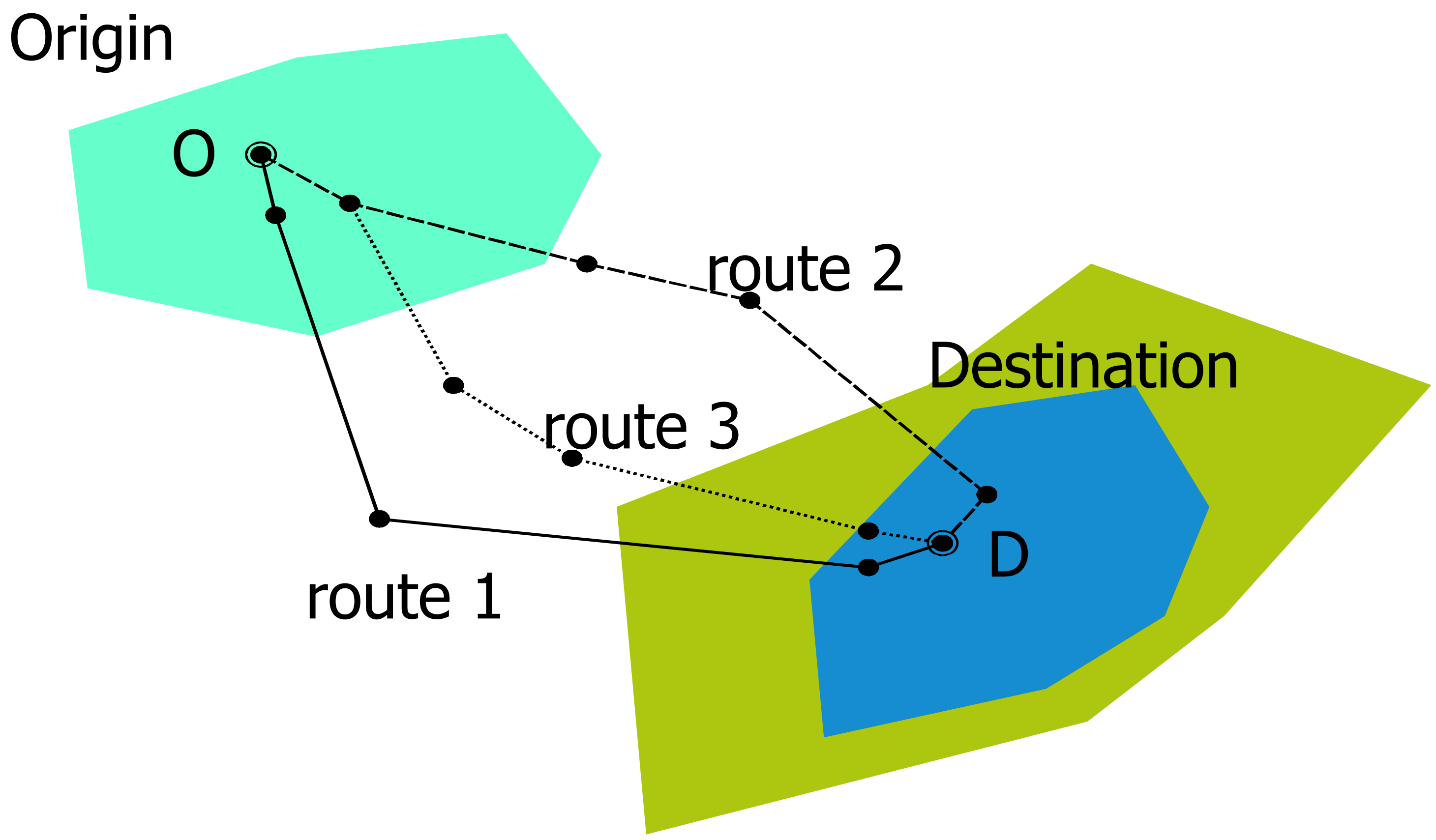


Key building block of transport models

- All kind of choices
  - Trip choice (stay/go)
  - Destination choice
  - Mode choice
  - Time-of-day choice
  - Route choice
  - Departure time choice
  - Move choice (stay/move)
  - Location choice

- Discrete choice modelling is used in other disciplines as well, e.g. marketing

Example route choice



What does it mean?

- Model to describe choice behaviour in situations where people have to choose from a set of distinct alternatives
- Key: individuals only pick one alternative

Key elements for decision making

- Decision maker: individual person or a group of people
- Alternatives: nonempty set of feasible and known alternatives to the decision makers
- Attributes of alternatives
- Decision rule

Decision rule

- Utility Theory: majority of choice models in transportation are based on the utility maximization assumption
- Travellers act rationally
- Travellers have well defined preferences
- Maximize the utility U_j of choosing alternative j

Random utility models (RUM)

- The individuals are assumed to select the alternative with the highest utility
- Inconsistencies in choice behaviour are assumed to be a result of observational deficiencies on the part of the analyst
- The utilities are unknown to the analyst. Thus, they are treated as random variables

$P(i|C)=Pr(U_i \geq U_j, \forall j \in C);$ *i,j: alternatives, C: choice set*

$$U_i = V_i + \varepsilon_i$$

V_i : systematic component of the utility

ε_i : random part of the utility

Basic case (binary choice)

Example Mode choice

$$\begin{aligned} &\textit{Car:} && U_c = \theta_1 T_c + \varepsilon_c \\ &\textit{Transit:} && U_t = \theta_1 T_t + \varepsilon_t \end{aligned}$$

Where T_c is the travel time with car and T_t the travel time with transit

$$\begin{aligned} P\left(c \mid \{c,t\}\right) &= P\left(U_c \geq U_t\right) \\ &= P\left(\theta_1 T_c + \varepsilon_c \geq \theta_1 T_t + \varepsilon_t\right) \\ &= P\left(\theta_1 T_c - \theta_1 T_t \geq \varepsilon_t - \varepsilon_c\right) \\ &= P\left(\theta_1 \left(T_c - T_t\right) \geq \varepsilon\right) \end{aligned}$$

Error term: mean

- Ideal model: mean is zero
- Can be guaranteed by introducing an alternative specific constant (ASC) for all alternatives except 1
  - Car $U_c = ASC + \theta_1 T_c + \varepsilon_c$
  - Transit $U_t = \theta_1 T_t + \varepsilon_t$
- If car is preferred over transit ASC is positive, otherwise ASC is negative

Error term: distribution

- Assumption 1: ε_x is the sum of many random variables capturing unobservable attributes
  => central limit theorem: Normal distribution

$$p_1 = \int_{-\infty}^{\infty}\int_{-\infty}^{V_1 - V_2 + x_1} \frac{\mathrm{e}^{\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1}{\sigma_1}\right)^2 - \frac{2\rho x_1 x_2}{\sigma_1\sigma_2} + \left(\frac{x_2}{\sigma_2}\right)^2\right]\right\}}}{2\pi\sigma_1\sigma_2\sqrt{(1-\rho^2)}} \mathrm{d}x_1\,\mathrm{d}x_2$$

- Assumption 2: ε_x is the maximum of many random variables capturing unobservable attributes
  => Gumbel theorem: Extreme value distribution

$$p_1 = \frac{e^{\beta V_1}}{e^{\beta V_1} + e^{\beta V_2}}$$

Given an Extreme value distribution....

- If ε_c and ε_t are independent and identically distributed (i.i.d.)

  $\varepsilon_c \sim EV(0,\beta)$

  $\varepsilon_t \sim EV(0,\beta)$

  Then: $\varepsilon \sim Logistic(0,\beta)$

Note that it is also assumed that the variances of both alternatives are equal

- For $Logistic\ (0,\beta)$ we have

$$P(c \geq \varepsilon) = \frac{1}{1+e^{-\beta c}}$$

$$P(c \mid \{c,t\}) = P(V_c - V_t \geq \varepsilon)$$

$$= \frac{1}{1+e^{-\beta(V_c - V_t)}}$$

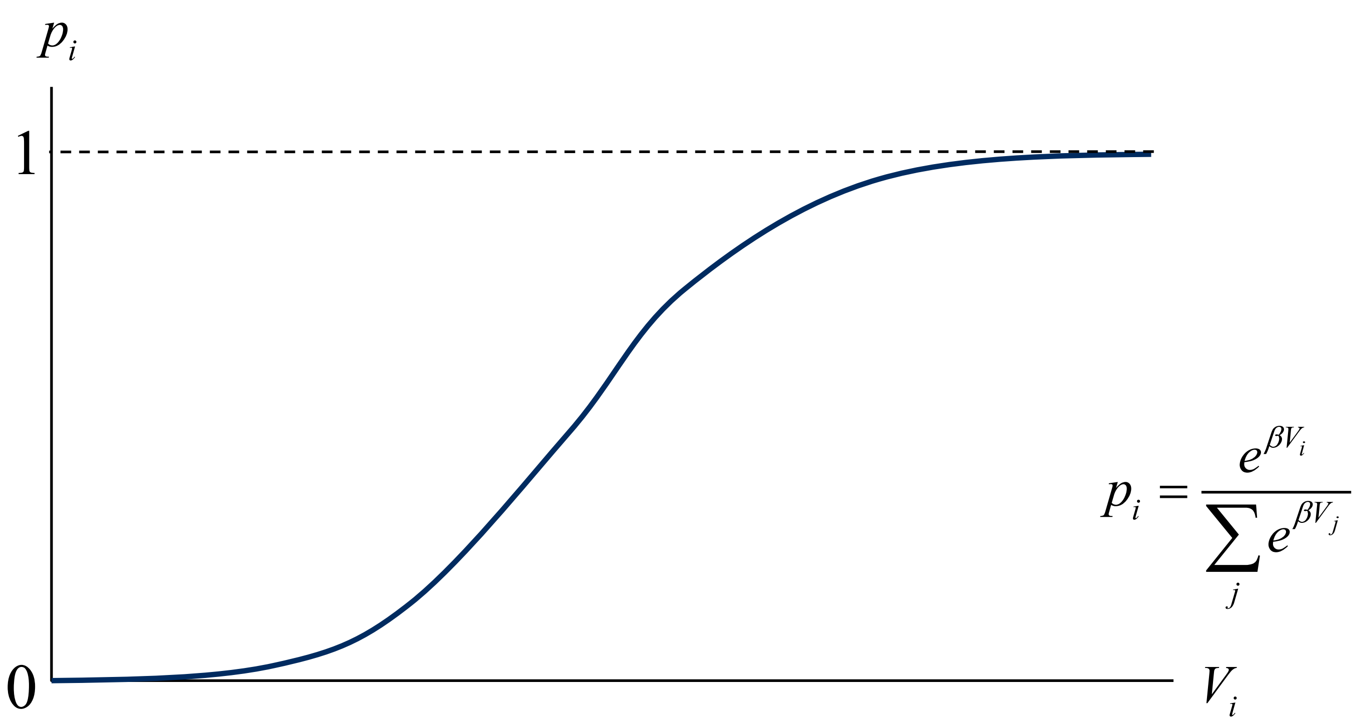
$$= \frac{e^{\beta V_c}}{e^{\beta V_c} + e^{\beta V_t}}$$

Logit model

- Binary case: $$P\left(c \mid \{c,t\}\right) = \frac{1}{1+e^{-\beta\left(V_c - V_t\right)}} = \frac{e^{\beta V_c}}{e^{\beta V_c} + e^{\beta V_t}}$$

- Note that difference is decisive!
- Parameter β describes sensitivity for differences:
  - β is zero: not sensitive
  - β is large: very sensitive (“all or nothing”)
- Multinomial case: $$P\left(i \mid \{alt_1, \ldots, alt_n\}\right) = \frac{e^{\beta V_i}}{\sum_{j=1}^{n} e^{\beta V_j}}$$

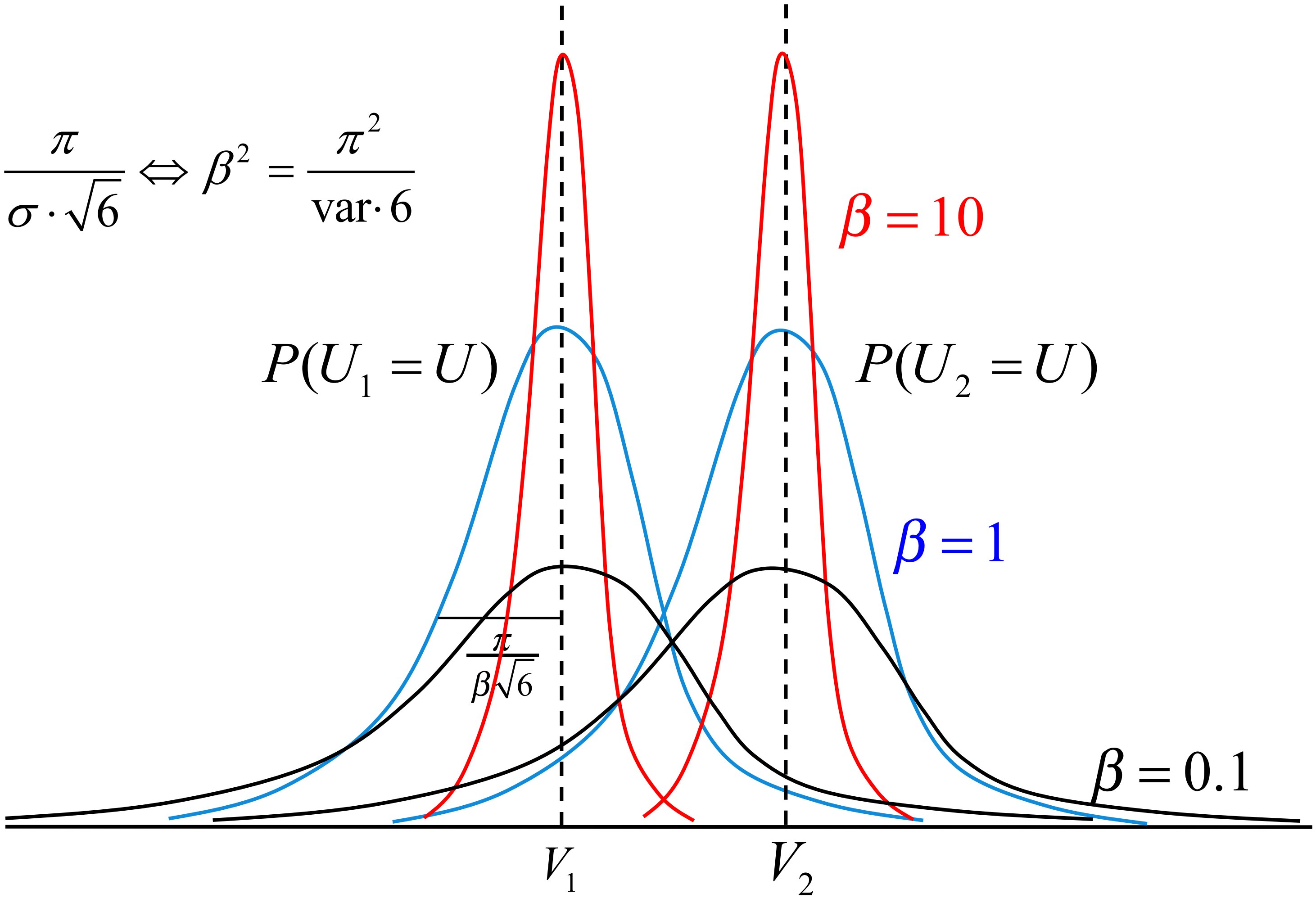
Shape of logit function



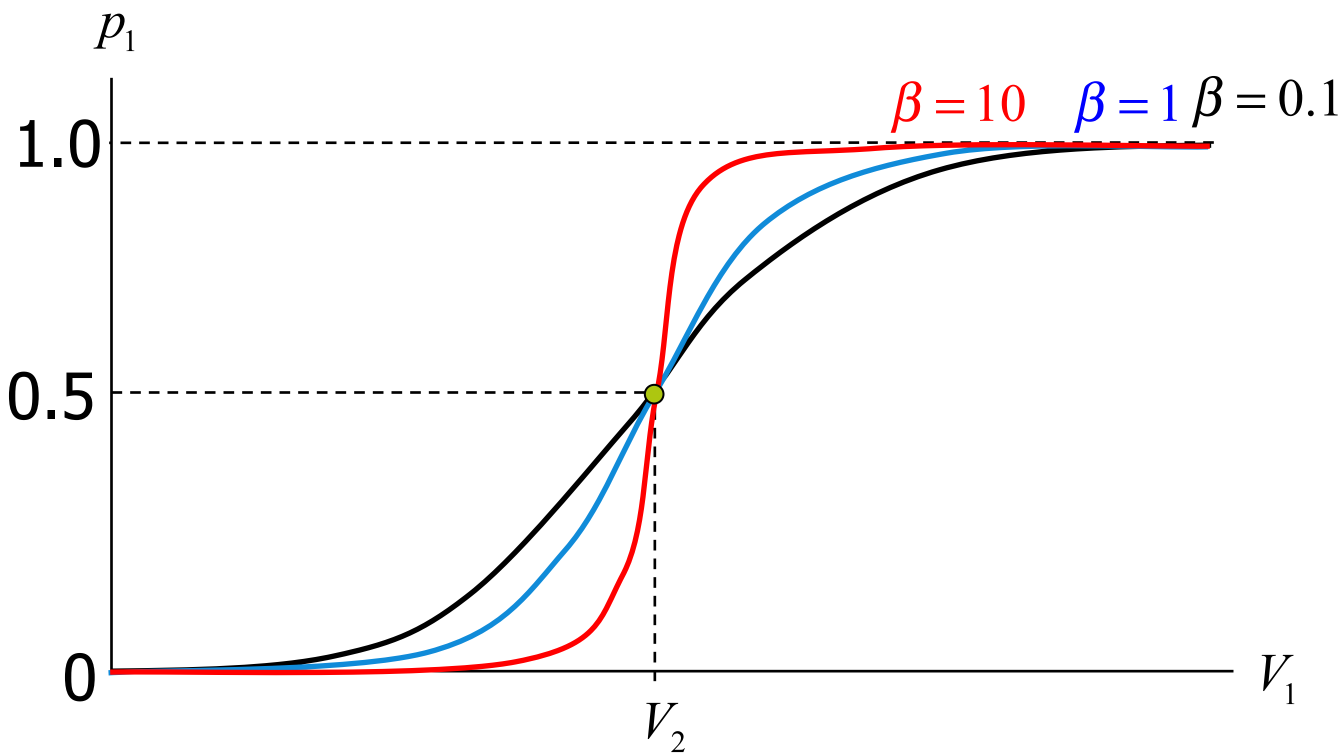
p_i = probability for choosing alternative *i*

Scale parameter and distribution

$$\beta = \frac{\pi}{\sigma \cdot \sqrt{6}} \Leftrightarrow \beta^2 = \frac{\pi^2}{\text{var} \cdot 6}$$



Impact of the scale parameter



The lower the scale parameter β, the higher the variance or 'spread' in the choice proportions and vice versa.

2.2

Application of the Logit model

Example mode choice

#	Time car	Time transit	Choice
1	52.9	24.4	T
2	14.1	28.5	T
3	14.1	86.9	C
…	…	…	…
10	95.0	43.5	T

$V_t = ASC + \theta_1 T_t$ $V_c = \theta_1 T_c$

$ASC = 0.5$ $\theta_1 = -0.1$

$V_{c,2} = -0.1 * 14.1 = -1.41$

$V_{t,2} = 0.5 - 0.1 * 28.5 = -2.35$

The V values are meaningless!

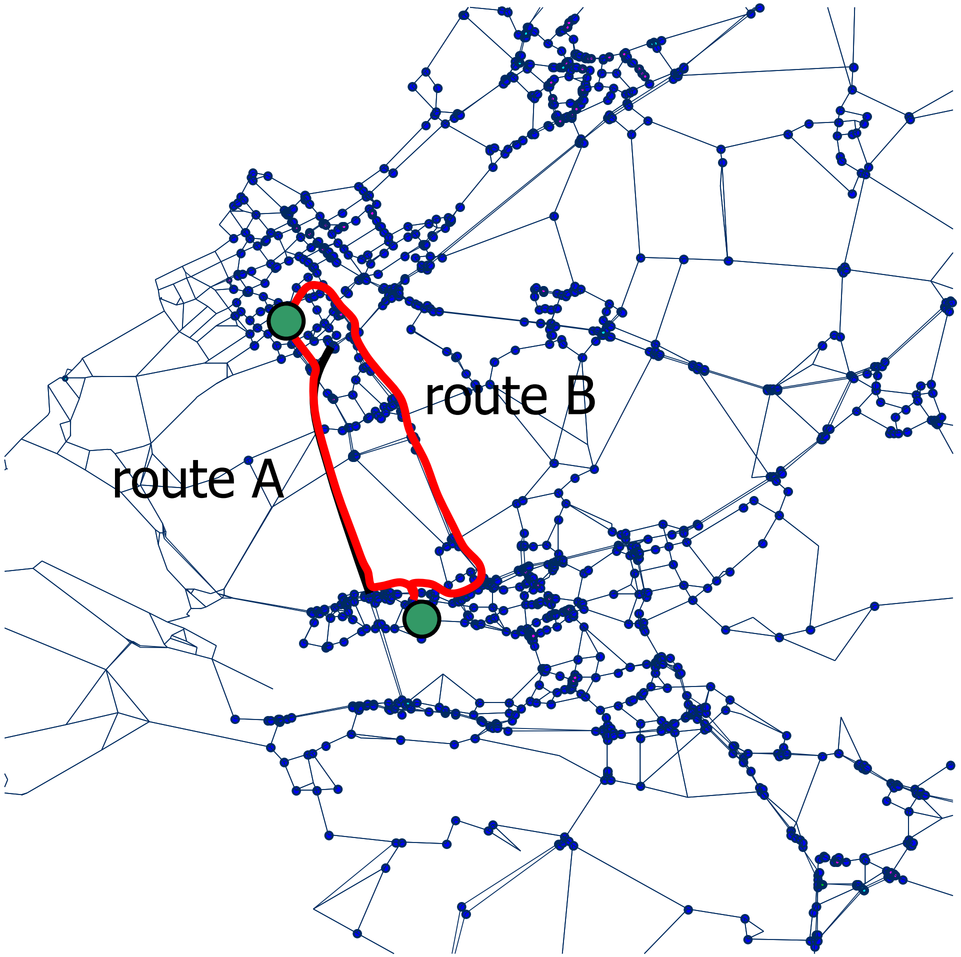
They make sense only as interpretation of the utility function

Probability of individual 2 to choose transit:

Assume: $\beta = 1$

$$P_{t,2} = \frac{e^{-2.35}}{e^{-1.41} + e^{-2.35}} = 0.28$$

Example Route choice



Extension A4

Schiedam - Den Haag
route A: 20 min. + € 2
route B: 35 min.

(dis)utility function:

$V_i = -(time + 5 * toll)\ [min]$

$\beta = 0.1$

$P_A = 62\%$

Where do the parameters come from?

- You need data on actual choice behaviour
  - Chosen alternative
  - Non-chosen alternatives
  - Including the (possibly) relevant attributes
- Typical data collection methods are
  - Revealed preference (i.e. observed behaviour)
  - Stated preference of Stated choice
- Search for the best model by specifying, estimating and assessing utility specifications
  - Using special software, e.g. ALOGIT, NLOGIT or BIOGEME
  - Using statistical tests and travel behaviour theory

Estimation of choice models

- What are the best values for the parameters, e.g. ASC and θ_1?
- Single observation: maximise probability chosen alternative (bit trivial, just define ASC)
- Two observations: maximise probability of observing both choices simultaneously, e.g. max: $P_1(T)*P_2(T)$
- Set of observations: max: $P_1(T)*P_2(T)*P_3(C)\ldots P_{10}(T)$
- Likelihood maximisation or, for numerical reasons, Log-likelihood maximisation

#	Time car	Time transit	Choice
1	52.9	24.4	T
2	14.1	28.5	T
3	14.1	86.9	C
…	…	…	…
10	95.0	43.5	T

Scale and utility parameters

- When estimating a model you determine the best value for $\beta\theta$
- In practice it is thus impossible to identify what the value of β or θ is
- Solution in practice is setting β (or one of the θ's) equal to 1
- This identification problem makes it difficult to compare parameters of different models
- Solution here is to compare ratio's of parameters, e.g. $\beta\theta_t/\beta\theta_c$ (=Value of time)

A comment on the sign of the θ's

- The main framework is utility maximisation, thus you would expect the sign to be positive
- This is true in case of the utility of a destination
  - Benefits are related to the activity to be performed
- Travel time and travel costs, however, reduce this utility
- Therefore travel time and travel costs are likely to have negative parameter values
- Consequence is that in cases where the positive utility of the activity does not play a role, e.g. mode choice or route choice, negative parameters are used

Some comments on the standard logit model

- Logit is commonly used, but isn't perfect
- Logit is sensitive for differences between utilities, independent of the absolute value of the utility
- How to take constraints into account?
- What to do if alternatives are not independent?
  - Route overlap
  - Red/Blue bus problem

2.3

Nested logit

Classic example

- Red and blue bus problem

Red and blue bus problem

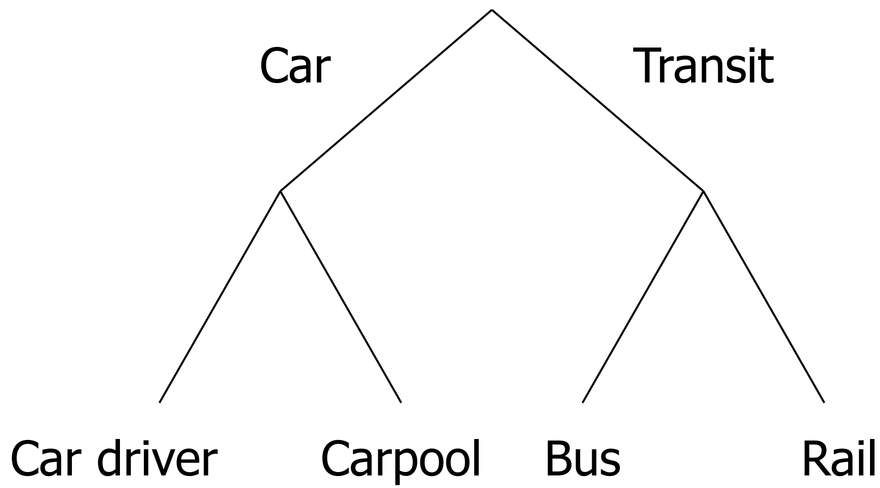
- Assume a simple mode choice problem: car versus bus, e.g. 75% car and 25% bus
- A new company enters having identical buses, except for the colour (i.e. blue instead of red), and having an identical schedule. So now we have 3 modes: car, red bus, blue bus.
- What is the share of car now?
  1. Still 75%
  2. Decreases to 60% (i.e. 0.75/(0.75+0.25+0.25))
  3. Other

Coping with correlations: Nesting

- Nesting accounts for (unobserved) similarities within nests: mix of correlation, simultaneousness and hierarchy
  - It does not necessarily imply a sequential order of choices!

- Special application/interpretation: Conditional choice:
  - Choice for alternative given choice for nest
  - Lower level choice options are part of higher level utility

Typical example



Nests k
Scale parameter β

Alternatives i
Scale parameter λ_k

$$P_i = \frac{e^{\beta V_i}}{\sum_j e^{\beta V_j}}$$

$$P(i,k) = P(i \mid k) P(k) = \frac{e^{\lambda_k V_{i|k}}}{\sum_{j \in k} e^{\lambda_k V_{j|k}}} \cdot \frac{e^{\beta V_k}}{\sum_{l \in K} e^{\beta V_l}}$$

Decomposition in two logits

Split utility in two parts:

- variables describing attributes for nests (aggregate level): W_k
- variables describing attributes within nest: Y_j

$$U_i = W_k + Y_i + \varepsilon_i \quad i \in B_k$$

Probability alternative is product of probability of alternative within nest and probability of nest

$$P_i = P_{i|B_k} P_{B_k}$$

Decomposition in two logits
Resulting formulas

$$P_{B_k} = \frac{e^{\beta \cdot (W_k + I_k)}}{\sum_{l=1}^{K} e^{\beta \cdot (W_l + I_l)}}$$

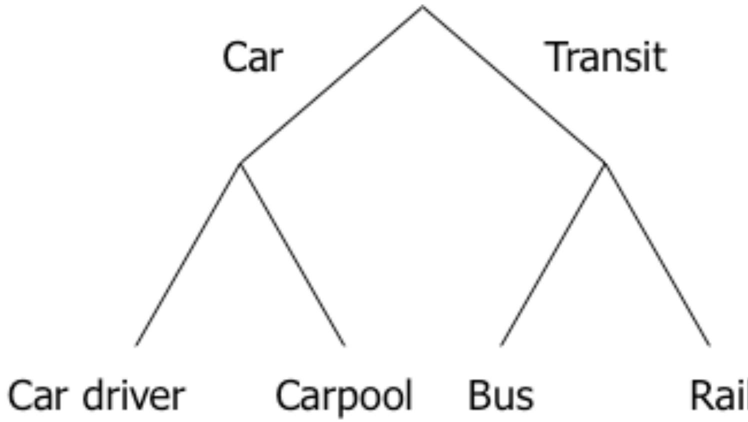
$$P_{i|B_k} = \frac{e^{\lambda_k \cdot Y_i}}{\sum_{j \in B_k} e^{\lambda_k \cdot Y_j}}$$

$$I_k = \frac{1}{\lambda_k} \ln \sum_{j \in B_k} e^{\lambda_k \cdot Y_j}$$

Logsum

$$P_{i|B_k} \cdot P_{B_k} = \frac{e^{\lambda_k \cdot Y_i}}{\sum_{j \in B_k} e^{\lambda_k \cdot Y_j}} \cdot \frac{e^{\beta \cdot \left(W_k + \frac{1}{\lambda_k} \ln \sum_{j \in B_k} e^{\lambda_k \cdot Y_j} \right)}}{\sum_{l=1}^{K} e^{\beta \cdot \left(W_l + \frac{1}{\lambda_l} \ln \sum_{j \in B_l} e^{\lambda_l \cdot Y_j} \right)}}$$

Typical conditions for nested logit



Nests k
Scale parameter β

Alternatives i
Scale parameter λ_k

$$P_{i|B_k} \cdot P_{B_k} = \frac{e^{\lambda_k \cdot Y_i}}{\sum_{j \in B_k} e^{\lambda_k \cdot Y_j}} \cdot \frac{e^{\beta \cdot \left(W_k + \frac{1}{\lambda_k} \ln \sum_{j \in B_k} e^{\lambda_k \cdot Y_j} \right)}}{\sum_{l=1}^{K} e^{\beta \cdot \left(W_l + \frac{1}{\lambda_l} \ln \sum_{j \in B_l} e^{\lambda_l \cdot Y_j} \right)}}$$

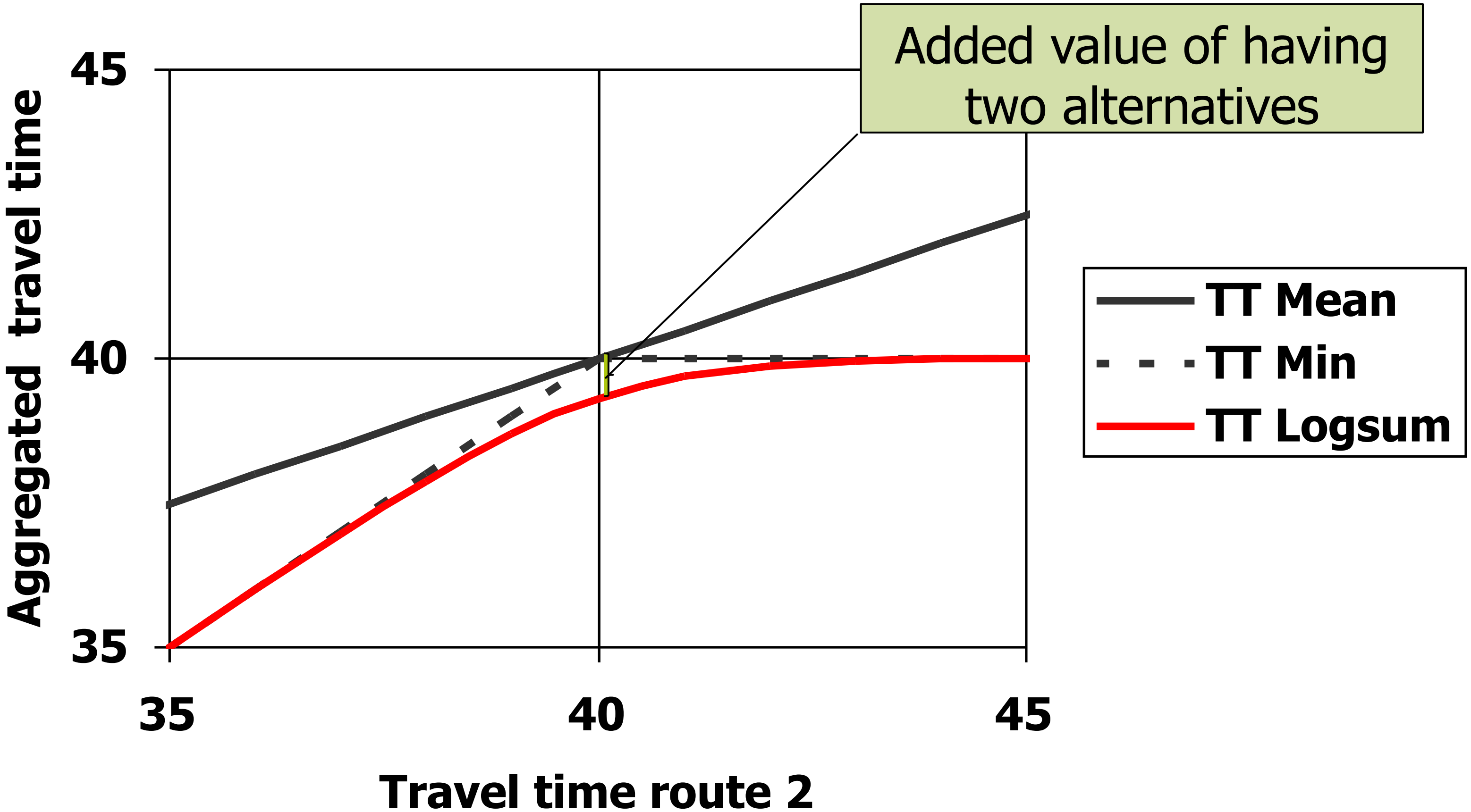
Define the parameter

$$\mu_k = \frac{\beta}{\lambda_k}$$

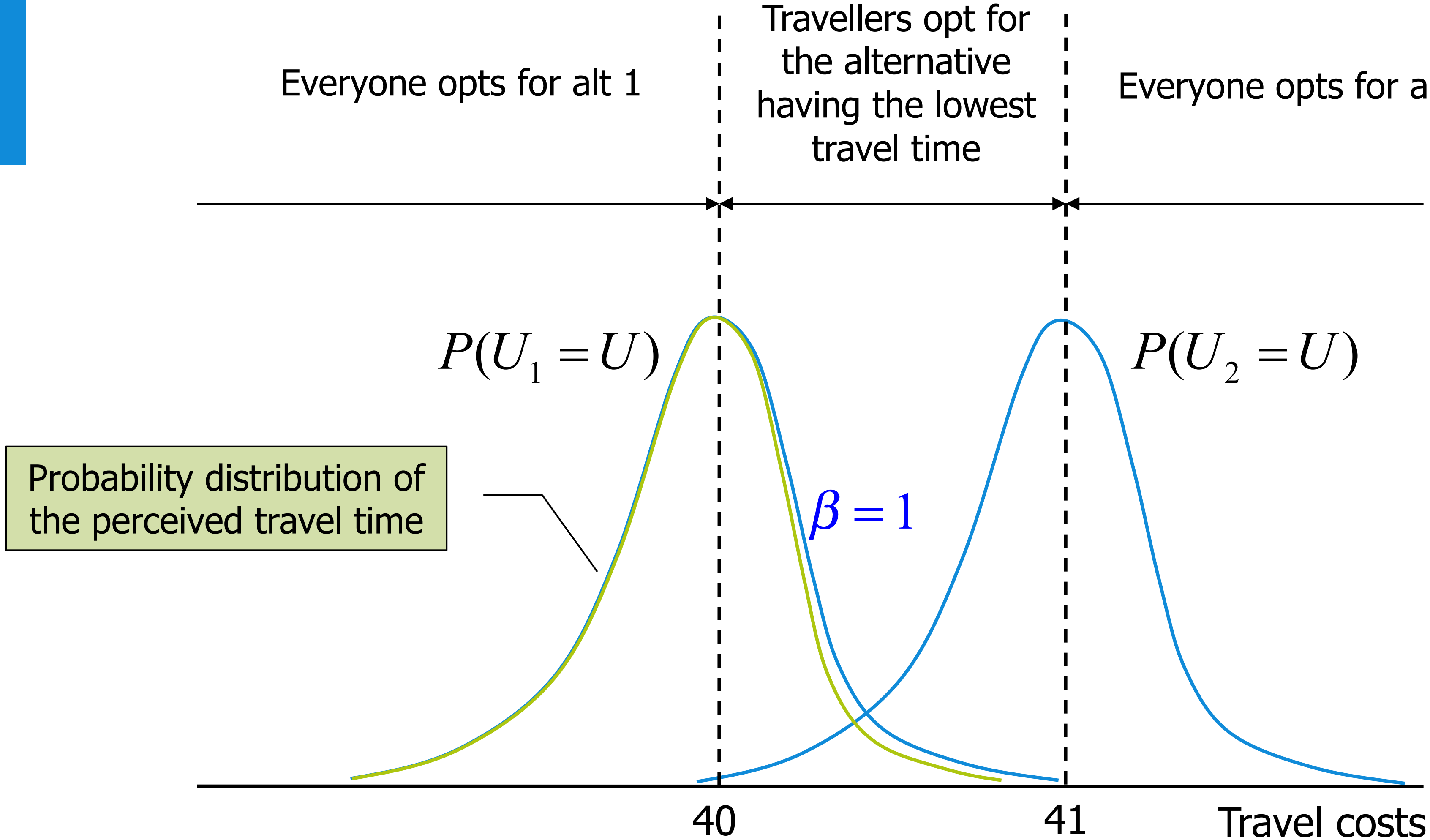
- It is required that $\mu_k \leq 1$
- Note that if $\mu_k = 1$ this expression collapses to the standard logit model
- If $\mu_k \rightarrow 0$, the nest is reduced to the alternative having the highest utility, i.e. the other alternatives in the nest have no additional value

Example route choice with 2 routes

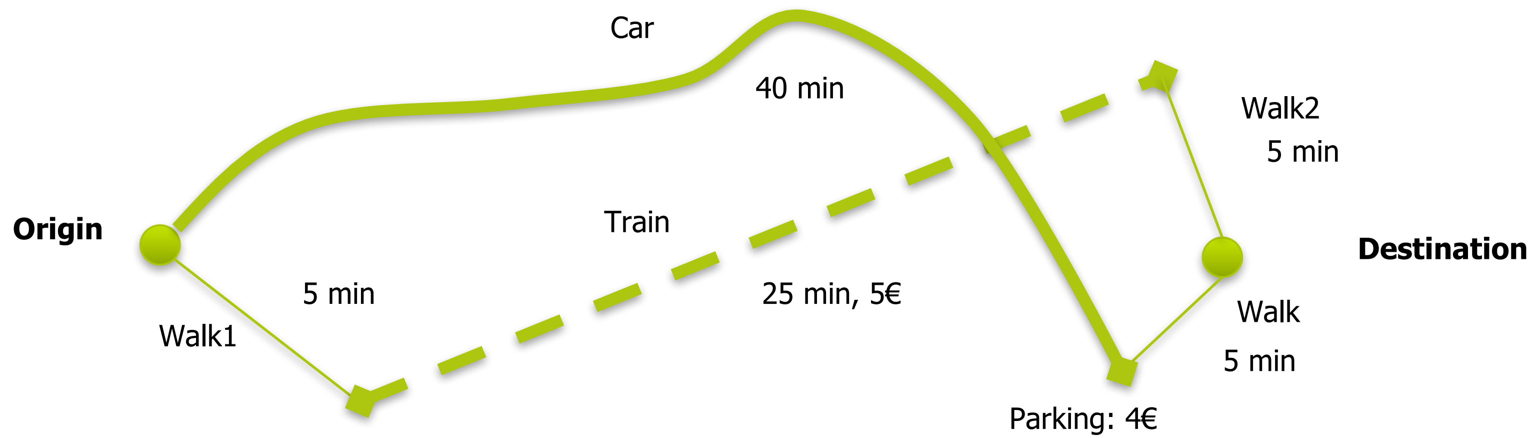
Travel time route 1 is 40 minutes, travel time route 2 varies



Why is there an added value?



Example for P+R facility



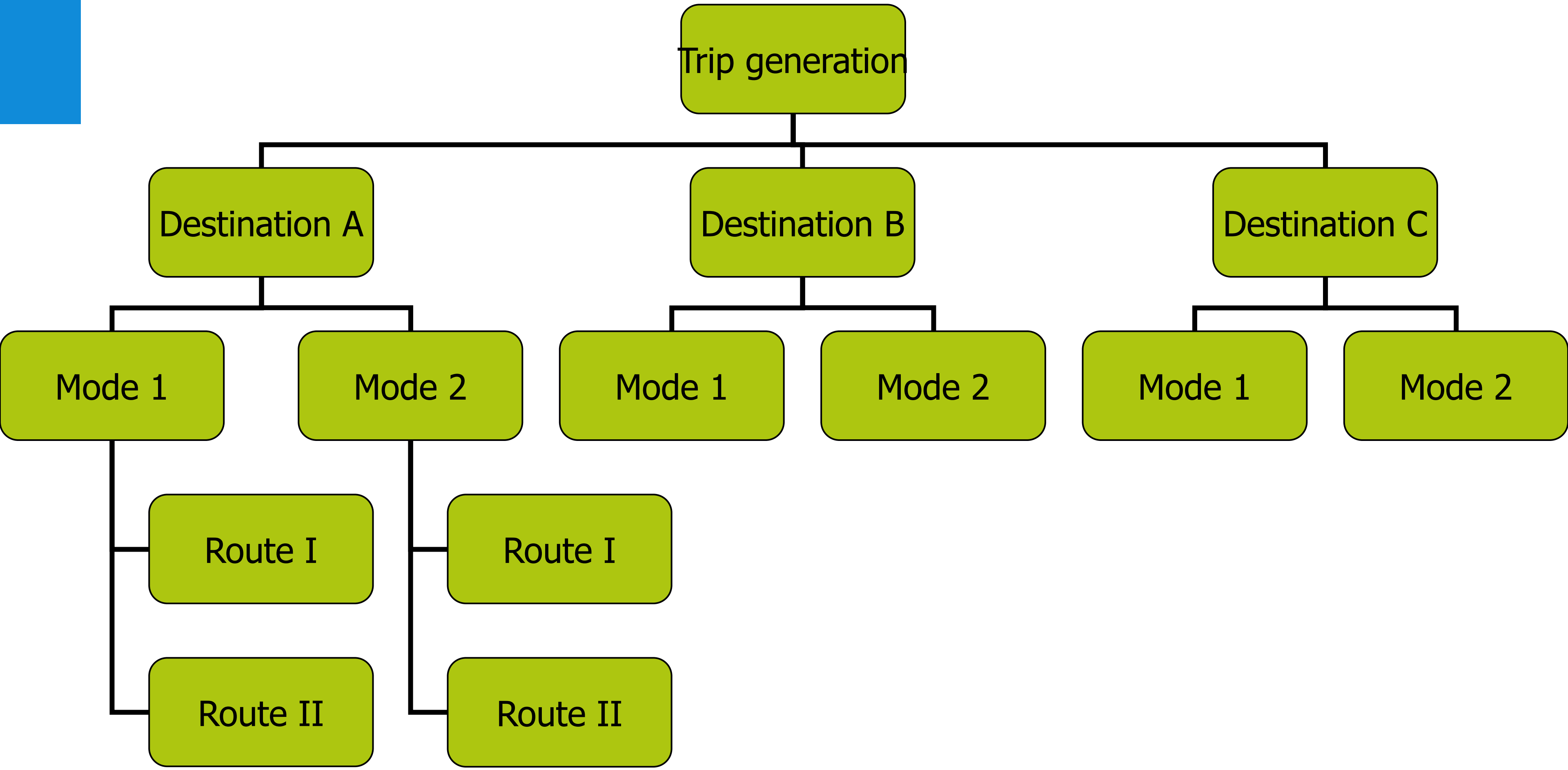
- See spreadsheet on Blackboard

  Analyse the spreadsheet and experiment with the values of β and λ

Other examples of nested models

- Logsum over routes in mode choice
- Logsum over modes in destination choice
- Dutch National Model considers nesting when modelling destination and mode choice (and tour generation)

Four stage model and logsums



Nested logit: to conclude

- Nested logit modelling proved to be a powerful tool for travel behaviour modelling
- Limitations: an alternative can only be allocated to a specific nest
- Possible extensions:
  - Cross-nested logit
  - Generalised nested logit
  - Network GEV (Generalised Extreme Value)