Discrete choice (2)

Applied Econometrics for Spatial Economics

Hans Koster
Professor of Urban Economics and Real Estate

VU VRIJE UNIVERSITEIT AMSTERDAM
tinbergen institute

1. Introduction

- Topics:
  1. Discrete choice
     - Random utility framework, estimating binary and multinomial regression models
  2. Spatial econometrics
     - Spatial data, autocorrelation, spatial regressions
  3. Identification
     - Research design, IV, OLS, RDD, quasi-experiments, standard errors
  4. Hedonic pricing
     - Theory and estimation
  5. Quantitative spatial economics
     - General equilibrium models in spatial economics

1. Introduction

Wednesday

09:30-10:30	Lecture 1	Discrete Choice I (The random utility framework)
10:45-11:45	Lecture 2	Discrete Choice II (Estimating discrete choice models)
12:00-13:00	Lecture 3	Spatial Econometrics I (Spatial data)
14:00-15:30	Tutorial 1	Assignment 1

Thursday

09:30-10:30	Lecture 4	Spatial Econometrics II (Spatial autocorrelation)
10:45-11:45	Lecture 5	Spatial Econometrics III (Spatial regressions)
12:00-12:30	Lecture 6	Identification I (Research design)
13:30-14:00	Tutorial 2	Discussion of Assignment 1
14:00-15:00	Tutorial 3	Assignment 2

Friday

09:30-10:00	Lecture 7	Identification II (RCTs, OLS, IV, quasi-experiments)
10:00-10:30	Lecture 8	Hedonic pricing I (Theory)
10:45-11:45	Lecture 9	Hedonic pricing II (Estimation)
12:00-12:30	Tutorial 4	Discussion of Assignment 2

1. Introduction

- How to estimate binary discrete choice models?

- Three main options
  1. Linear probability model
  2. Logit
  3. Probit

2. Linear probability model

- **Regress 0/1 variable on characteristics of that choice and use OLS:**
  $\Pr(d_j = 1) = \beta' x_j$

- **Dataset example:**

Chosen	Price	Time
1	14	12
0	25	5
0	15	15
1	15	13
1	4	45
1	3	40
0	20	10

VRIJE UNIVERSITEIT AMSTERDAM

2. Linear probability model

Advantages:

- **Consistent when** $0 \leq \hat{y}_j \leq 1 \; \forall \; j$

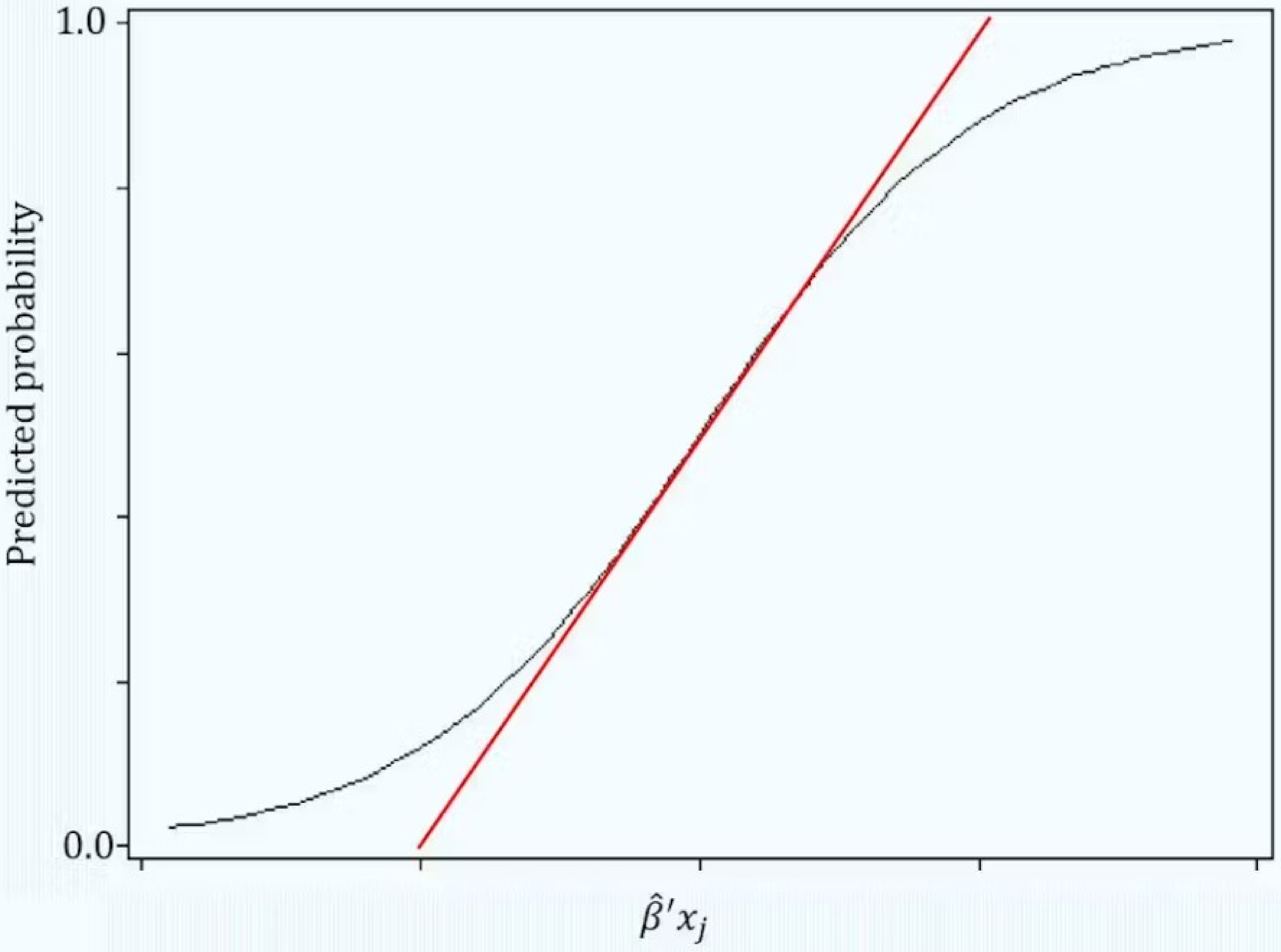
VU VRIJE UNIVERSITEIT AMSTERDAM

2. Linear probability model

VU VRIJE UNIVERSITEIT AMSTERDAM

Advantages:

- Consistent when $0 \leq \hat{y}_j \leq 1 \; \forall \; j$



2. Linear probability model

Advantages:

- **Consistent when** $0 \leq \hat{y}_j \leq 1 \; \forall \, j$

- **Easy to interpret**
  - **Say that** $\beta = -0.25$ **and** x **is measured in €, then for each euro increase in** x**, the probability to choose alternative** j **decreases by 25 percentage points**
  - $\frac{\partial \Pr(d_j = 1)}{\partial x} = \beta$

VU VRIJE UNIVERSITEIT AMSTERDAM

2. Linear probability model

Advantages:

- **Consistent when** $0 \leq \hat{y}_j \leq 1 \; \forall \; j$
- **Easy to interpret**
  - $\frac{\partial \Pr(d_j=1)}{\partial x} = \beta$
- **Computationally feasible**
  - **Important for large panel datasets**
- **In practice, leads to very similar results as Logit and Probit**

VU VRIJE UNIVERSITEIT AMSTERDAM

Disadvantages:

- No direct link with structural parameters of utility function
  - e.g. not able to calculate aggregate utility from choice set
- Biased for small samples and possibly inconsistent marginal effects
  - Linearity?
- Not suitable for multinomial choices

VU VRIJE UNIVERSITEIT AMSTERDAM

3. Logit

- **Let's define**

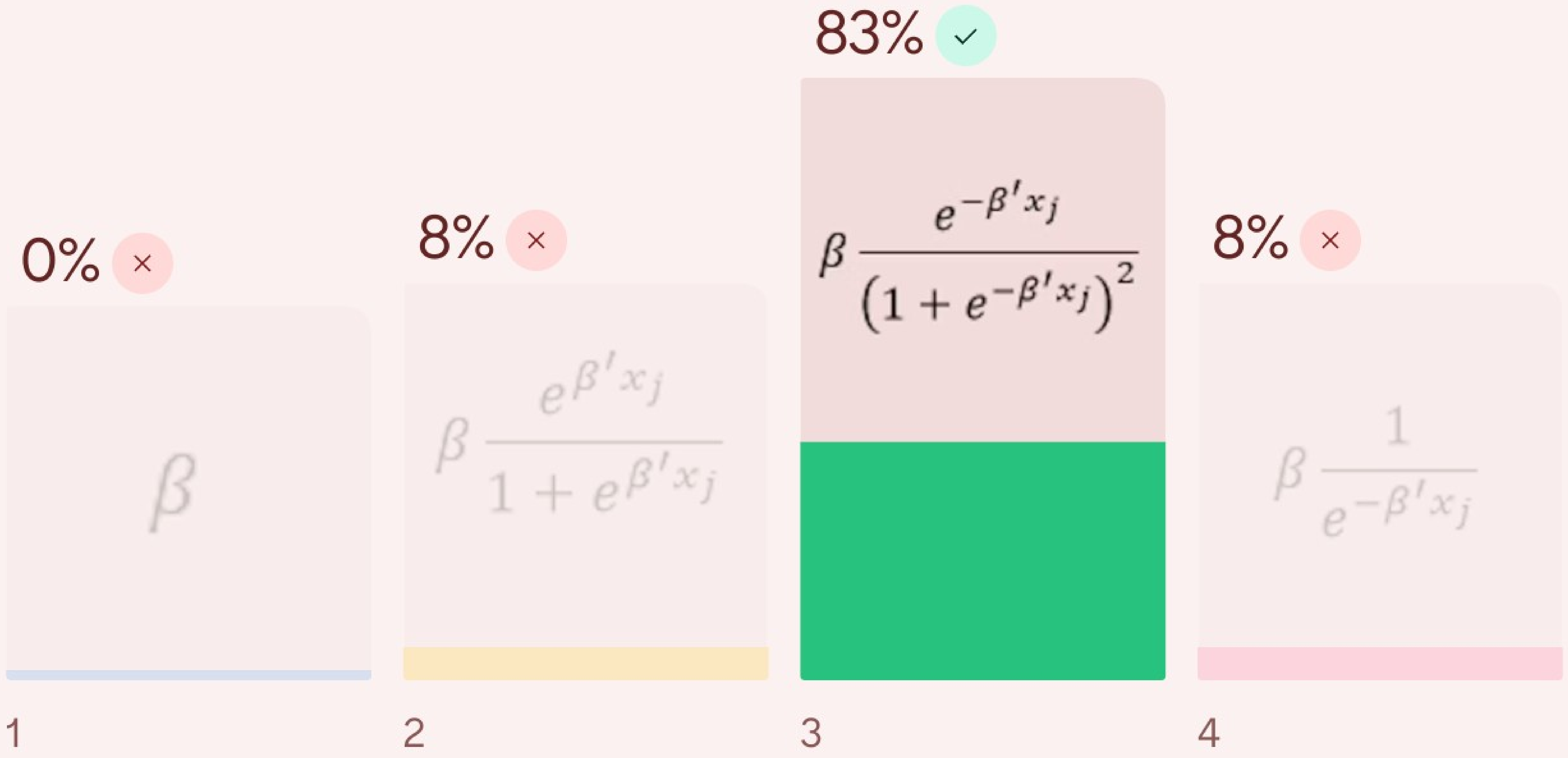
$$\Pr(d_j = 1) = \frac{1}{1+e^{-\beta' x_j}}$$

- **Example: regress 0/1 variable on *differences* in characteristics of the alternatives**

Chosen_B	$\text{Price}_B\text{-Price}_A$	$\text{Time}_B\text{-Time}_A$
1	-14	5
0	5	0
0	15	-20
1	-8	13
1	-10	3
1	3	-5
0	20	10

VU VRIJE UNIVERSITEIT AMSTERDAM

What is the marginal effect on the probabiblity of one unit increase in x, so $\frac{\partial \Pr(d_j=1)}{\partial x_j}$?



3. Logit

- **Marginal effects:**
  - **Use chain rule of differentiation**
  - $$\frac{\partial \Pr(d_j=1)}{\partial x_j} = -\left(1 + e^{-\beta' x_j}\right)^{-2} \times e^{-\beta' x_j} \times -\beta$$
  - $$\frac{\partial \Pr(d_j=1)}{\partial x_j} = \beta \frac{e^{-\beta' x_j}}{\left(1+e^{-\beta' x_j}\right)^2}$$

VU VRIJE UNIVERSITEIT AMSTERDAM

3. Logit

- **The change in the probability for one unit increase in x**
  - $\frac{\partial \Pr(d_j=1)}{\partial x_j} = \beta \frac{e^{-\beta' x_j}}{\left(1+e^{-\beta' x_j}\right)^2}$

- **Marginal effect depends on x_j, so is not constant/linear**
  - **For example, evaluate at mean values of x**

VU VRIJE UNIVERSITEIT AMSTERDAM

3. Logit

- **Software**
  - LOGIT **or** LOGISTIC **in STATA**
  - REGRESSION – BINARY LOGISTIC **in SPSS**

- **In STATA you can select to report marginal effects**
  - **Use** MARGINS **after** LOGIT **command**
  - **Choose at which x the values are evaluated (e.g. at means)**

VRIJE UNIVERSITEIT AMSTERDAM

3. Logit

Advantages of Logit:

- Predicted probability is always between one and zero
- Clear link to random utility framework
  - Log-sum may be used for welfare calculations
- Closed-form marginal effects
  - Usually leads to very similar results as Probit
- Can include 'fixed effects' (XTLOGIT in STATA)
  - *e.g. to* control for individual heterogeneity

3. Logit

Disadvantages of Logit:

- Why Extreme Value Type I distribution for ϵ?

- Maximum likelihood / non linear model

4. Probit

- We may also assume that ϵ_j is normally distributed, so $\epsilon_j = N(0, \sigma^2)$
  - This implies $\Pr(d_j = 1) = \Phi(\beta' x_j)$
  - However, no closed-form for cumulative normal distribution!

- Marginal effects:

$$\frac{\partial \Pr(d_j=1)}{\partial x_j} = \beta\phi(\beta x_j)$$

where $\phi(\cdot)$ is the density function of the normal distribution

VRIJE UNIVERSITEIT AMSTERDAM

4. Probit

Advantages:

- Normal distribution for ϵ_j may seem more reasonable
- Probability is always between one and zero

Disadvantages:

- No closed-form marginal effects
- Hard to include many fixed effects

4. Probit

- How to choose between the three models?
  - Probit estimates ≈ Logit estimates
  - Check for robustness of marginal effects
  - Large sample and interested in marginal effects?
    - → Usually linear probability model!
    - → There is an ongoing debate in economics on this issue

5. Application

- Koster and Koster (2014)
  - Estimate the value of time and unreliability
  - Uses a stated choice experiment

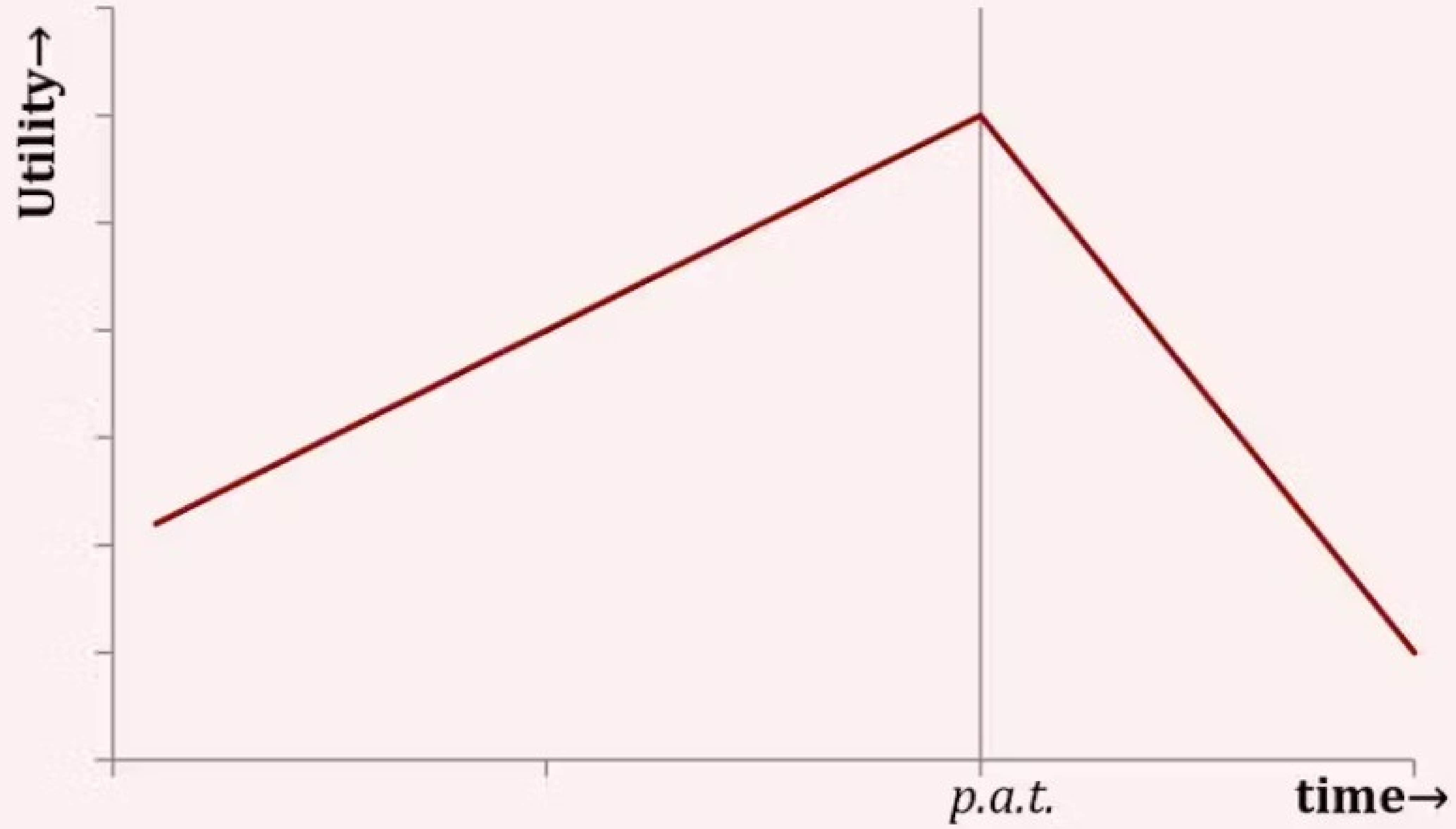
- Stated-choice experiment about preferences of morning commuters
  - "*Spitsmijden*" (Peak-avoidance project)
  - People get a reward if they avoid the peak

  - But: they may be too early or late at work!
  - Trade-off

VU VRIJE UNIVERSITEIT AMSTERDAM

5. Application

- Value of time
- Value of schedule delay early
- Value of schedule delay late
  - *p.a.t.* = preferred arrival time



5. Application

- **Example of a choice with two alternatives and uncertainty**

Your preferred arrival time if there is no delay is: 8:40.

	Alternative 1		Alternative 2	
Departure time from home	6:05		6:50	
Probability	*80%*	*20%*	*90%*	*10%*
Total travel time	30 min	40 min	20 min	35 min
Arrival time at work	6:35	6:45	7:10	7:25
Reward	4 euro	4 euro	0 euro	0 euro

5. Application

- **Utility is specified as follows**
  - $U_{icj} = \beta^R R_{icj} + \beta^T T_{icj} + \beta^{SDE} SDE_{icj} + \beta^{SDL} SDL_{icj} + \epsilon_{icj}$

i	**individual**
c	**choice**
j	**alternative**
R_{icj}	**expected *reward***
T_{icj}	**expected travel time**
SDE_{icj}	**expected time before *p.a.t.***
SDL_{icj}	**expected time after *p.a.t.***
ϵ_{icj}	**random taste variation, Extreme Value Type I distributed**

5. Application

- **Utility is specified as follows**
  - $U_{icj} = \beta^R R_{icj} + \beta^T T_{icj} + \beta^{SDE} SDE_{icj} + \beta^{SDL} SDL_{icj} + \epsilon_{icj}$
  - $\Delta u_{inj} = 0 = \beta^R \Delta R_{icj} + \beta^T \Delta T_{icj} + \beta^{SDE} \Delta SDE_{icj} + \beta^{SDL} \Delta SDL_{icj}$

- **Estimate the value of time and unreliability**

  - **Value of time (VOT):**

$$-\beta^R \Delta R_{icj} = \beta^T \Delta T_{icj} \rightarrow \Delta T_{icj} = -1 \rightarrow -\Delta R_{icj} = -\frac{\beta^T}{\beta^R}$$

Note that we look at the willingness to pay. Because the experiment focuses on rewards, we have $-\Delta R_{icj}$

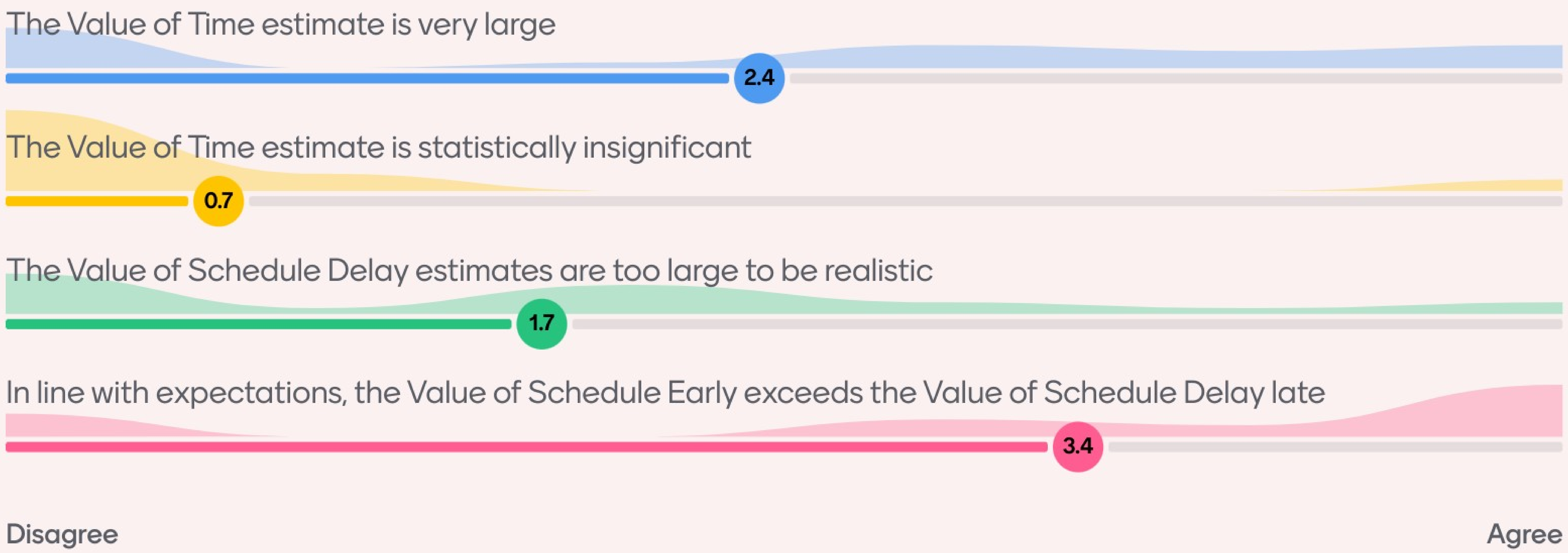
  - **Value of schedule delay early (VSDE):**

$$-\beta^R \Delta R_{icj} = \beta^{SDE} \Delta SDE_{icj} \rightarrow \Delta SDE_{icj} = -1 \rightarrow -\Delta R_{icj} = -\frac{\beta^{SDE}}{\beta^R}$$

  - **Value of time (VSDL):**

$$-\beta^R \Delta R_{icj} = \beta^{SDL} \Delta SDL_{icj} \rightarrow \Delta SDL_{icj} = -1 \rightarrow -\Delta R_{icj} = -\frac{\beta^{SDL}}{\beta^R}$$

What is your interpretation of the results?



19

5. Application

- Results *(s.e.'s between parentheses)*

VOT	€ 35.05 (€ 4.158)
VSDE	€ 23.22 (€ 2.211)
VSDL	€ 17.16 (€ 1.621)

- Willingness to pay estimates are high
  - People are more sensitive to tolls
  - Relatively high share of high income households
- VSDE>VSDL?
  - Constraints in the morning rather than at work

6. Summary

Today:

- How to estimate binary choice models?
  - Use LPM, Logit or Probit

- Application to measure value of time, value schedule delay early and schedule delay late

Discrete choice (2)

Applied Econometrics for Spatial Economics

Hans Koster
Professor of Urban Economics and Real Estate

Discrete choice (3)

Applied Econometrics for Spatial Economics

Hans Koster
Professor of Urban Economics and Real Estate

VU VRIJE UNIVERSITEIT AMSTERDAM
tinbergen institute

1. Introduction

- Topics:
  1. Discrete choice
     - Random utility framework, estimating binary and multinomial regression models
  2. Spatial econometrics
     - Spatial data, autocorrelation, spatial regressions
  3. Identification
     - Research design, IV, OLS, RDD, quasi-experiments, standard errors
  4. Hedonic pricing
     - Theory and estimation
  5. Quantitative spatial economics
     - General equilibrium models in spatial economics

1. Introduction

Wednesday

09:30-10:30	Lecture 1	Discrete Choice I (The random utility framework)
10:45-11:45	Lecture 2	Discrete Choice II (Estimating discrete choice models)
12:00-13:00	Lecture 3	Spatial Econometrics I (Spatial data)
14:00-15:30	Tutorial 1	Assignment 1

Thursday

09:30-10:30	Lecture 4	Spatial Econometrics II (Spatial autocorrelation)
10:45-11:45	Lecture 5	Spatial Econometrics III (Spatial regressions)
12:00-12:30	Lecture 6	Identification I (Research design)
13:30-14:00	Tutorial 2	Discussion of Assignment 1
14:00-15:00	Tutorial 3	Assignment 2

Friday

09:30-10:00	Lecture 7	Identification II (RCTs, OLS, IV, quasi-experiments)
10:00-10:30	Lecture 8	Hedonic pricing I (Theory)
10:45-11:45	Lecture 9	Hedonic pricing II (Estimation)
12:00-12:30	Tutorial 4	Discussion of Assignment 2

1. Introduction

- **How to estimate these types of models?**

- **Overview**

	# Alternatives	Coefficients
1. Binary Logit	2	Homogeneous
2. Multinomial Logit with alternative specific parameters	>2, <~10	Differ between alternatives
3. Nested Logit	>2, <~10	Usually homogeneous
4. Conditional Logit	>2	Homogeneous

2. Multinomial logit

- **Recall:**

$$\Pr(Y = A) = \frac{e^{\beta x_A}}{e^{\beta x_A} + e^{\beta x_B} + e^{\beta x_C}}$$

But now let the coefficients be alternative-specific:

$$\Pr(Y = A) = \frac{e^{\beta_A x_A}}{e^{\beta_A x_A} + e^{\beta_B x_B} + e^{\beta_C x_C}}$$

- **We cannot identify all the coefficients β_A, β_B, β_C, because <u>we compare the results to a reference category</u>**
  - » **Think of dummies**

- **Illustration: we can write the probability only in terms of differences with respect to one reference category, e.g.:**

$$\Pr(Y = A) = \frac{1}{1 + e^{\beta_B x_B - \beta_A x_A} + e^{\beta_C x_C - \beta_A x_A}}$$

3. Nested logit

- Independence of irrelevant alternatives
  - Adding an alternative does not affect the relative odds between two other options considered
  - Solution: use Nested Logit
    → Allows for correlation within nests

- Software
  - NLOGIT in STATA
  - Use Biogeme software
  - Limdep/nlogit

4. Conditional logit

- Often, the number of alternatives is very large
  - Location choice
  - Route choice
  - Holiday destinations
  - Choice of car
  - Partner choice
  - ...

- With Multinomial Logit this becomes infeasible
  - Unique coefficients for each alternative
  - Not necessary for large choice sets

- Conditional Logit:

$$\Pr(d_j = 1) = \frac{e^{\beta' x_j}}{\sum_{k=1}^{J} e^{\beta' x_k}}$$

VRIJE UNIVERSITEIT AMSTERDAM

4. Conditional logit

- **How to deal with large choice sets?**
  - **Number of observations in your regressions is** ***number of alternatives*** **×** ***respondents***

1. **Model aggregate choices**
2. **Random selection of alternatives**
3. **Estimate count data models (Poisson)**

4. Conditional logit

1. Model aggregate choices

- Modelling location choice
  - Focus on aggregate areas (*e.g.* municipalities)

- Choice of cars
  - Only distinguish between brands

- However, lack of detail makes results less credible

2. Random selection of alternatives

- McFadden (1978)
  - Choose a random subset of J alternatives for each choice set, including the chosen option
  - This should not affect the *consistency* of the estimated parameters
  - Small-sample properties are yet unclear

- How large should J be?

- Applied in many good papers
  - e.g. Bayer et al. (2007, *JPE*)

VRIJE UNIVERSITEIT AMSTERDAM

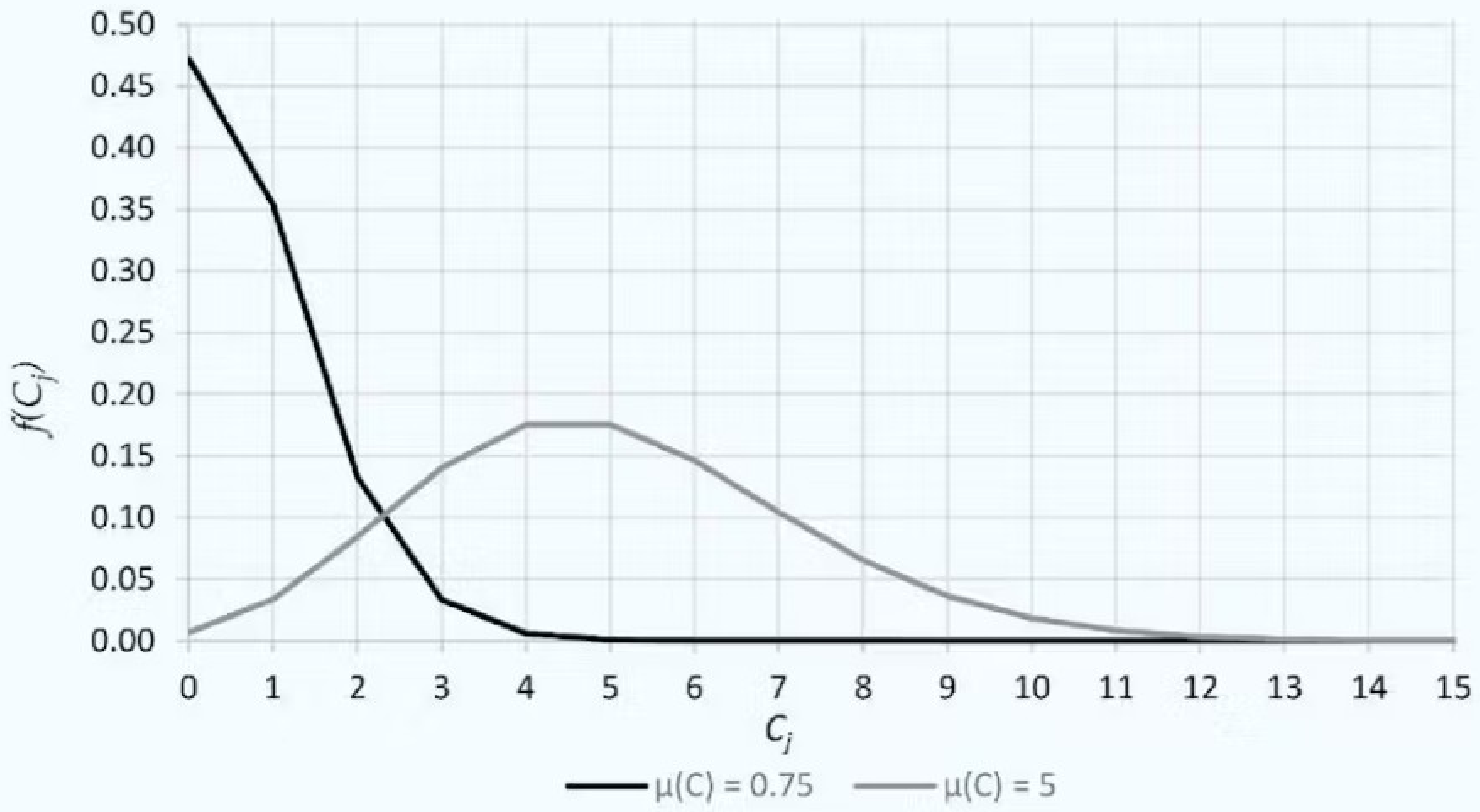
3. Estimate count data models

- Estimate Conditional Logit by means of a Poisson model

- A Poisson regression is a count data model
  - Dependent variable is integer
  - … and should be Poisson distributed

3. Estimate count data models

- Example of a Poisson distribution



- Equidispersion: $\bar{y} = \sigma_y$

VRIJE UNIVERSITEIT AMSTERDAM

3. Estimate count data models

- **Estimate Conditional Logit by means of a Poisson model**

- **A Poisson regression is a count data model**
  - **Dependent variable is integer**
  - **… and should be Poisson distributed**
  - $C_j = e^{\beta' x_j} + \epsilon$
    where C_j is the # of decision makers that have chosen a certain alternative

- **Convenient interpretation of β**
  - **When x_j increases with one, C_j increases with $\beta \times 100$ percent**

VRIJE UNIVERSITEIT AMSTERDAM

3. **Estimate count data models**

- **A Poisson model should give identical parameters to the Conditional Logit**
  - **Maximum likelihood functions are identical *up to a constant***
  - **Guimarães et al. (2003)**

- **Hence, group observations based on their chosen alternatives**
  - **... the number of firms choosing a certain location**
  - **... the number of people buying a certain car**

4. Conditional logit

3. Estimate count data models

- Implications
  - You cannot include characteristics of the decision maker *(because you sum over decision makers)*!
  - Homogeneous parameters across the population

- Extensions
  - Include fixed effects
  - Negative binomial regression
  - Zero-inflated models
  - See Guimarães et al. (2004) for details

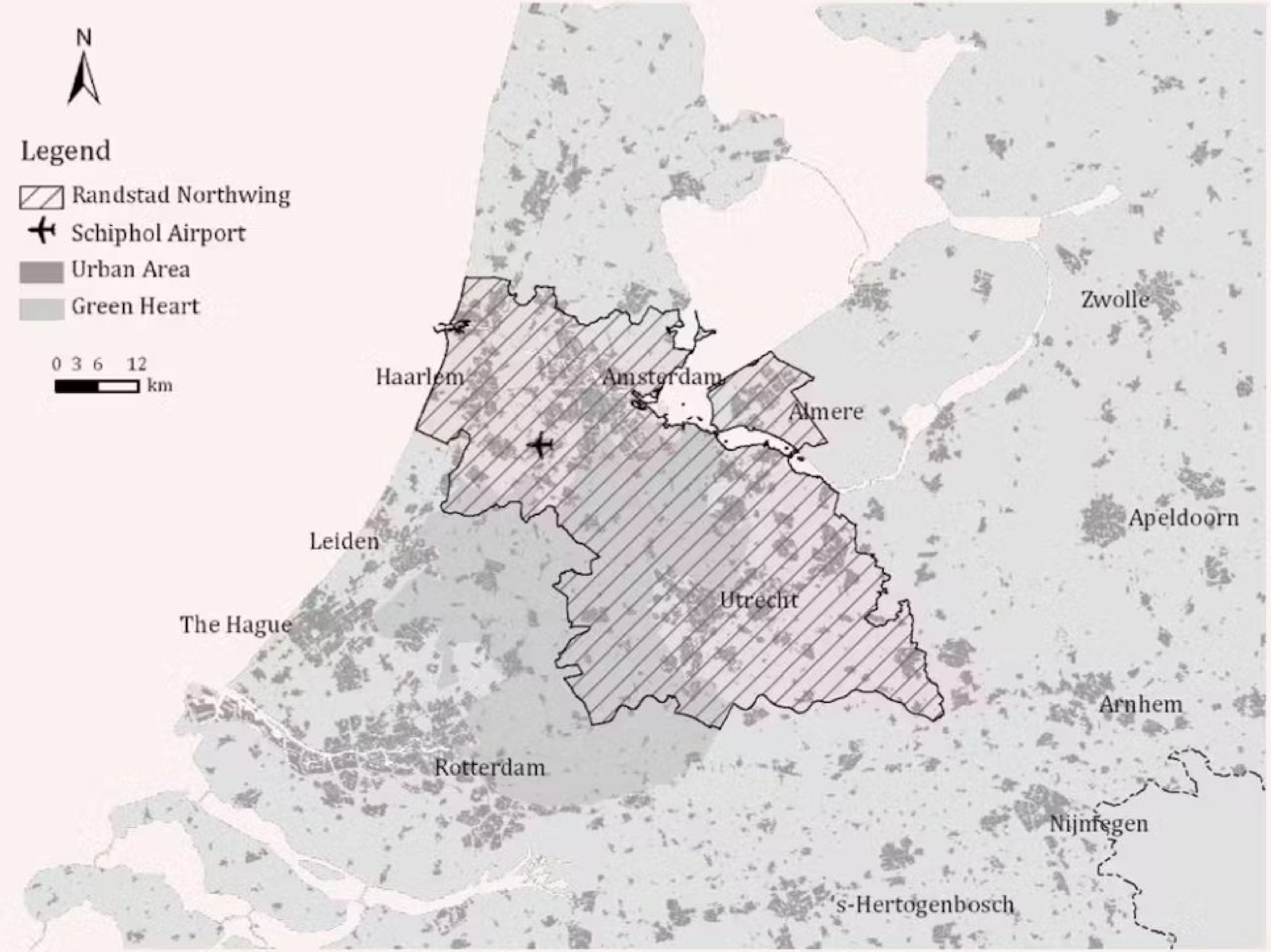
4. Conditional logit

- Jacobs *et al.* (2013)
  - Analyse location choices of business start-ups
  - Investigate the impact of *multinationals* on the number of business start-ups
  - In the Randstad Northwing

- Multinationals may generate:
  - Knowledge spillovers
  - Spin-offs
  - Potential customers *(output sharing)*

4. Conditional logit

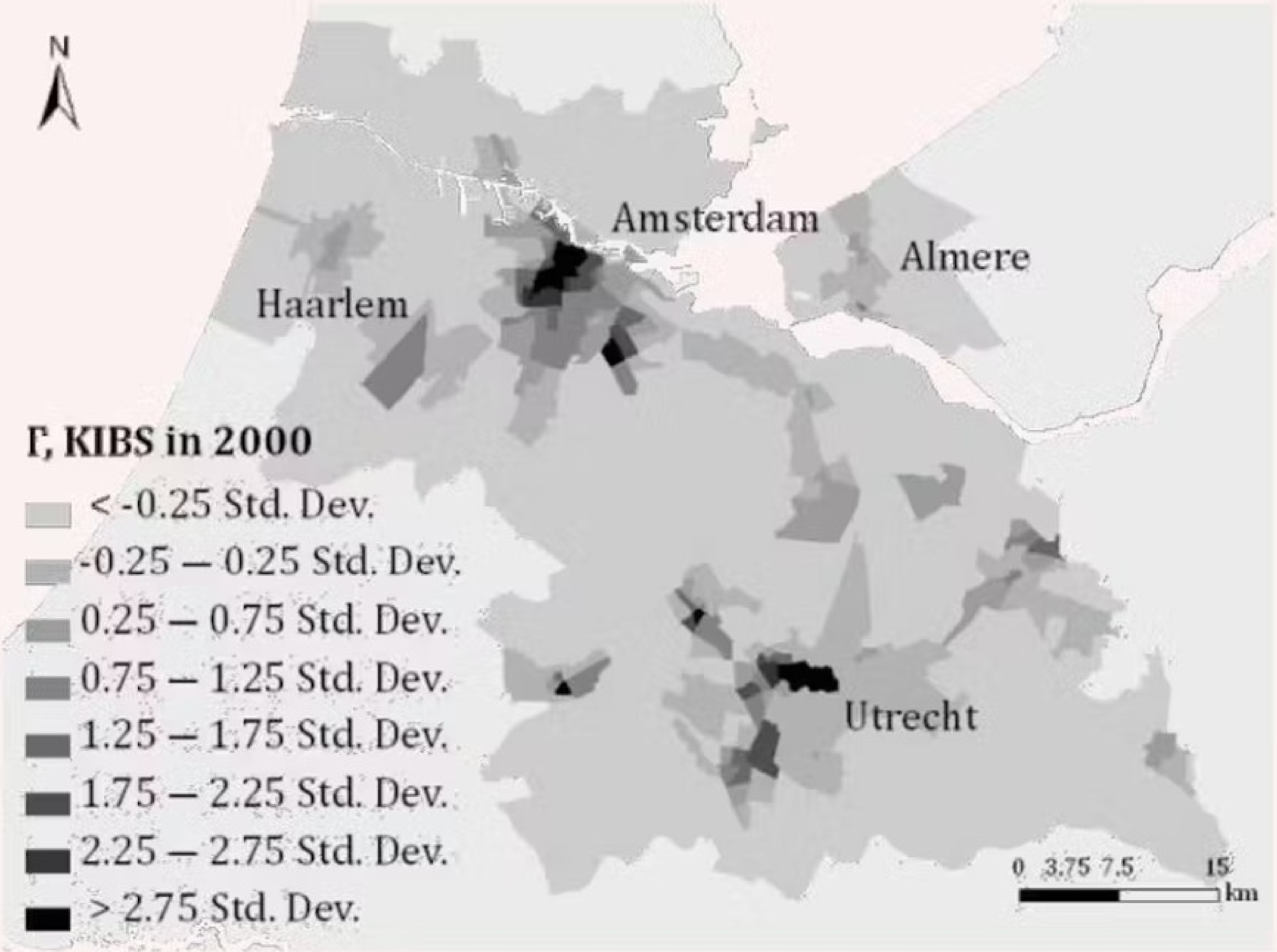
- **The Randstad Northwing**



VU VRIJE UNIVERSITEIT AMSTERDAM

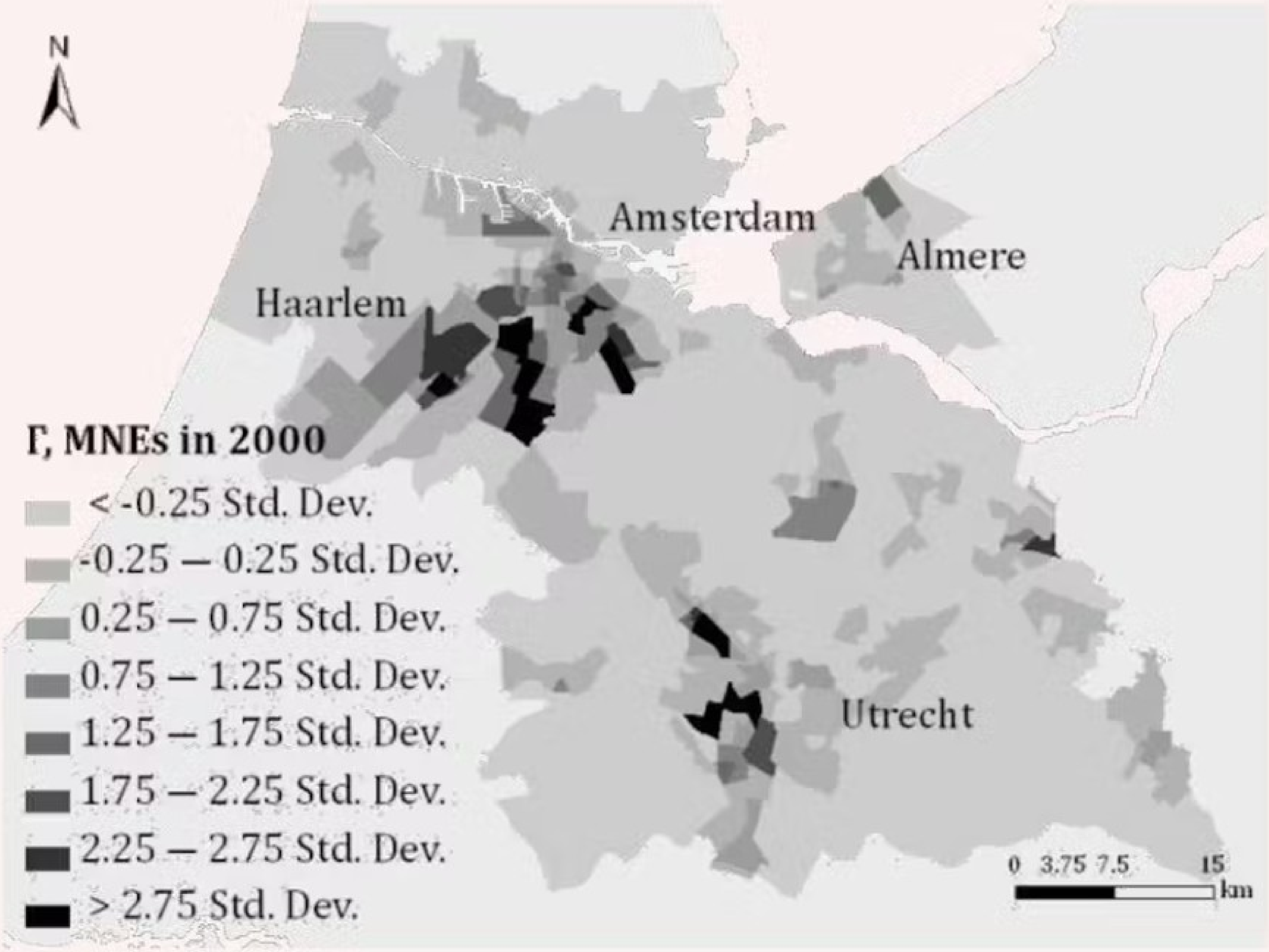
4. Conditional logit

- **Business services in the Randstad Northwing**



4. Conditional logit

- **Multinationals in the Randstad Northwing**



4. Conditional logit

- **Location choice model:**

$$\pi_{ij} = \alpha + \beta e_j^{MNE} + \gamma e_j^{BSF} + \delta e_j^{OF} + \zeta X_j + \eta_{j \in M} + \epsilon_{ij}$$

i	**firm**
j	**PC6 location (alternatives)**
e_j^{MNE}	**multinational employment**
e_j^{BSF}	**business services employment**
e_j^{OF}	**other employment**
X_j	**control variables**
$\eta_{j \in M}$	**municipality fixed effects**

VU VRIJE UNIVERSITEIT AMSTERDAM

4. Conditional logit

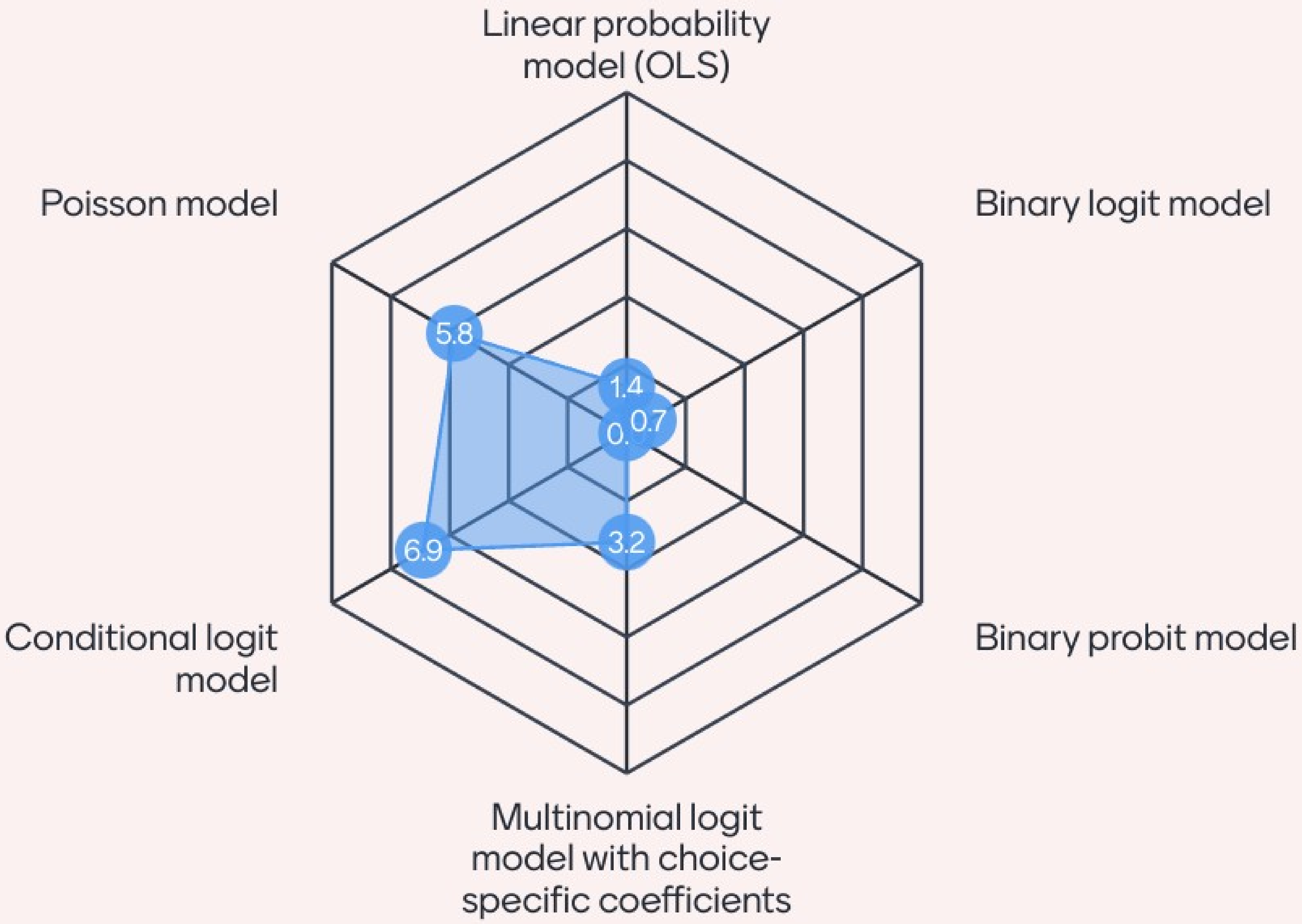
- **Probability Π that i chooses k:**

$$\Pr(d_j = 1) = \frac{e^{\alpha+\beta e_j^{MNE}+\gamma e_j^{BSF}+\delta e_j^{OF}+\zeta X_j+\eta_{j\in M}}}{\sum_{\kappa=1}^{J} e^{\alpha+\beta e_k^{MNE}+\gamma e_k^{BSF}+\delta e_k^{OF}+\zeta X_k+\eta_{k\in M}}}$$

- **There are 13,655 locations**

→ **How would you estimate this model?**

What regression method would you use to estimate this model?



4. Conditional logit

- Problem: many locations
  - Use count data to estimate this model
    → There are no individual firm characteristics
  - Dependent variable: # start-ups per location

VRIJE UNIVERSITEIT AMSTERDAM

4. Conditional logit

- **Results**

Table – A POISSON MODEL

(Dependent variable: The number of business services start-ups per location)

	(1)	(2)	(3)
Multi-national employment density *(log)*	0.0709***	0.0422***	0.0772***
	(0.0151)	(0.0092)	(0.0121)
Business services employment density *(log)*	0.4304***	0.4374***	0.3821***
	(0.0240)	(0.0162)	(0.0214)
Other employment density *(log)*	-0.2242***	-0.2203***	-0.1352***
	(0.0162)	0.0071	(0.0178)
Control variables (9)	No	Yes	Yes
Municipality fixed effects (61)	No	No	Yes
Number of locations	13,655	13,655	13,655
Log-likelihood	-13,146.903	-13,051.163	-12,709.249

Notes: We include locations with at least 10 employees in 2000. The coefficients can be interpreted as elasticities and differ from Jacobs et al. (2013) because of a slightly different set of controls and because we estimate Poisson models instead of Negative-binomial regressions. Robust standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

→ **Please interpret the results in column (3)**

VU VRIJE UNIVERSITEIT AMSTERDAM

4. Conditional logit

- Multinationals attract new business services
  - The effect of other business services on start-ups, is however, much larger

- Coefficients are convenient to interpret
  - *e.g.* a 1% increase in multinational empl. leads to an increase of start-ups of 0.077% (in Column (1))

6. Summary

Discrete choice:

- Random utility framework

- Generalisations of logit models
  - LPM $(J = 2)$
  - Binary logit/probit $(J = 2)$
  - Multinomial logit $(2 < J < 10)$
  - Nested logit $(2 < J < 10)$
  - Conditional logit $(J > 2)$

- Conditional Logit models can be estimated by count data models
  - Cannot include characteristics of the decision maker

Discrete choice (3)

Applied Econometrics for Spatial Economics

Hans Koster
Professor of Urban Economics and Real Estate

VU VRIJE UNIVERSITEIT AMSTERDAM
tinbergen institute

5. Revealed preference and stated preference data

Types of data

- Revealed preference (RP) data
  - Observed or reported actual behaviour

- Stated preference (SP) data
  - Respondents are confronted with hypothetical choice sets

- Combinations of RP and SP

5. Revealed preference and stated preference data

Advantages of RP data

- Based on actual behaviour!!
- Use existing (large) data sources
  - Cheaper
  - No expensive experiments
- Panels of the same individuals over a long time

5. Revealed preference and stated preference data

Disadvantages of RP data

- Lack of variability
- Collinearity (e.g. price and travel times)
- Lack of knowledge on the choice set
- Not possible with new choice alternatives
- Actual behaviour may not be first choice
  - University numerus fixus
- Perception errors and imperfect information
  - Airline tickets

5. Revealed preference and stated preference data

- Example of stated preference question
  - Different from contingent valuation!

Suppose you have to ship a product from A to B			
Option 1		**Option 2**	
Price:	€ 1,000	Price:	€ 750
Handling time:	3 days	Handling time:	1 week
% does not arrive:	1.0%	% does not arrive:	1.3%
What alternative will you choose?			

5. Revealed preference and stated preference data

Advantages of SP data

- New alternatives
- New attributes
- Large variability is possible
- Problems of collinearity can be solved
  - 'Orthogonal design'
- Choice set is clearly defined

VU VRIJE UNIVERSITEIT AMSTERDAM

Disadvantages of SP data

- Information bias
- Starting point bias
- Hypothetical bias
- Strategic bias
- Errors

Disadvantages of SP data

- Information bias
  - The respondent has incorrect information on the context
  - Make your experiment as realistic as possible

- Starting point bias
  - Respondents are influenced by the set of available responses to the experiment
  - Test your design and choose realistic attribute values

5. Revealed preference and stated preference data

Disadvantages of SP data

- Hypothetical bias
  - Individuals tend to respond differently to hypothetical scenarios than they do to the same scenarios in the real world.
  - Cognitive incongruity with actual behaviour
  - Again: make your experiment as realistic as possible
  - But otherwise hard to mitigate…

- Strategic bias
  - Respondent wants a specific outcome
  - (S)he fills in answers that are in line with desired outcomes

Disadvantages of SP data

- Unintentional biases
  - Information, starting point, hypothetical bias
- Intentional biases
  - Strategic bias
- Errors
  - Boredom
  - Respondents do not carefully read instructions
  - Respondents do not understand the questions

If there is good data available, I would prefer RP *(personal opinion)*

Would you use SP or RP in the following cases:

Investigate the preference of households for hydrogen-powered cars

Test the impact of a (past) industrial policy on firm location choices

Measure costs and benefits of the construction of a high-speed rail line to Shikoku

Measure the economic impact of existing power plants on nearby residential neighbourhoods

0.0

Stated preference — **Revealed preference**

6. Summary

Today:

- Generalisations of logit models
    - Multinomial logit
    - Nested logit
    - Conditional logit
- Conditional Logit models can be estimated by count data models
    - Cannot include characteristics of the decision maker
- Data
    - Stated preference or revealed preference data

6. Summary

Discrete choice:

- Random utility framework
- Generalisations of logit models
  - LPM ($J = 2$)
  - Binary logit/probit ($J = 2$)
  - Multinomial logit ($2 < J < 10$)
  - Nested logit ($2 < J < 10$)
  - Conditional logit ($J > 2$)
- Conditional Logit models can be estimated by count data models
  - Cannot include characteristics of the decision maker
- Data
  - Stated preference or revealed preference data

VRIJE UNIVERSITEIT AMSTERDAM

Discrete choice (3)

Applied Econometrics for Spatial Economics

Hans Koster

Professor of Urban Economics and Real Estate