Discrete choice (1)

Applied Econometrics for Spatial Economics

Hans Koster
Professor of Urban Economics and Real Estate

VRIJE UNIVERSITEIT AMSTERDAM

tinbergen institute

1. Introduction

- **Today:**
  1. Spatial econometrics
  2. Discrete choice
  3. Identification
- **Tomorrow:**
  4. Hedonic pricing
  5. Quantitative spatial economics

- **Today:**
  1. Spatial econometrics
  2. Discrete choice
     - Random utility framework, estimating binary and multinomial regression models
  3. Identification

- **Tomorrow:**
  4. Hedonic pricing
  5. Quantitative spatial economics

1. Introduction

VRIJE UNIVERSITEIT AMSTERDAM

1. Introduction

1. Introduction

- *Continuous* choice: estimate marginal changes in behaviour
  - E.g. "when fuel price increases by 10%, the demand for fuel will decrease by 2%"
  - Standard micro-economic theory applies

- Transport demand often has a discrete (binary) nature
  - Some x impacts a discrete y
  - Then use discrete choice methods

1. Introduction

- Discrete choice methods
  - *Dependent* variable y_i is discrete

- Why not use OLS?

- Let's have the standard OLS equation

$$y_i = \beta x_i + \epsilon_i \qquad (1)$$

where i indexes the individual

1. Introduction

- **OLS may be consistent for binary choice**
  - ***But,*** y_i ***(and therefore*** ϵ**) is *not* normally distributed**

- **Horrace and Oaxaca (2006)**
  - **Leads to biased and inconsistent estimates if** $\hat{y}_i$ **lies 'often' outside the [0,1] interval**
  - **I show later today why that is an issue…**

- **OLS does not necessarily provides a link with economic theory**

- **Not suitable for multinomial choice**

VRIJE UNIVERSITEIT AMSTERDAM

2. The random utility framework

- **Indirect utility may be given by:**

  $$U_{iA} = V_A(travel\ time_A) + \epsilon_{iA} \quad (2)$$

  $$U_{iB} = V_B(travel\ time_B) + \epsilon_{iB} \quad (3)$$

  - V_A, V_B ➔ **deterministic utility**

- **Random terms:** $\epsilon_{iA}, \epsilon_{iB}$: **random taste variation**
  - **Random utility model (RUM)**
  - **Note that the levels of U_{iA} and U_{iB} are not directly observed!**

- $\Pr(Y = A) = \Pr(U_{iA} > U_{iB})$
- $\Pr(V_A + \epsilon_{iA} > V_B + \epsilon_{iB}) = \Pr(V_A - V_B > \epsilon_{iB} - \epsilon_{iA})$

VRIJE UNIVERSITEIT AMSTERDAM

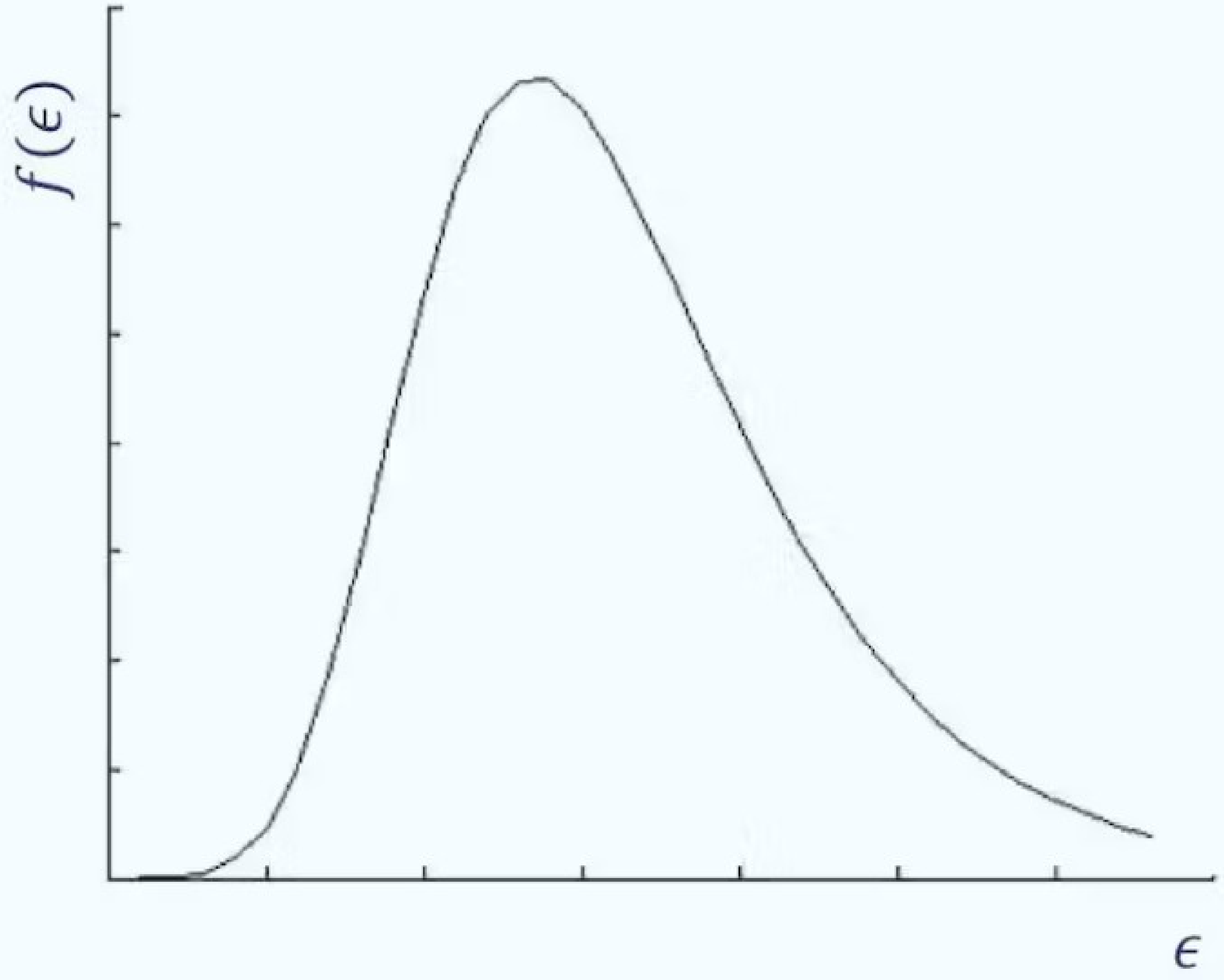
2. The random utility framework

- Two things are unknown
  - Which distribution for ϵ's?
  - What is the functional form for V_A and V_B?

2. The random utility framework

- **Which distribution for ϵ's?**
  - **ϵ's are unobserved**
  - **You draw them from a distribution**

  - **Logit: Extreme Value Type I distribution**



2. The random utility framework

- **Which distribution for ϵ's?**
  - **Extreme Value Type I distribution**
  - **Generates simple closed-form solutions!**

    $\rightarrow \quad \Pr(V_A - V_B > \epsilon_{iB} - \epsilon_{iA})$
  - **Daniel McFadden (1964)**

2. Random utility framework

- **It appears that:**

$$\Pr(Y = A) = \frac{e^{V_A}}{e^{V_A} + e^{V_B}} \tag{4}$$

- **With two alternatives this can be written as:**

$$\Pr(Y = A) = \frac{1}{1 + e^{V_B - V_A}}$$

2. The random utility framework

- Which functional form for V_A and V_B?
  - Can be any function
  - Linear function is often assumed
  - Can be extended with multiple variables

    $$U_{jA} = \beta p_{jA} + \kappa t_{jA} + \epsilon_{jA} \quad (5)$$

    $$U_{jB} = \beta p_{jB} + \kappa t_{jB} + \epsilon_{jB} \quad (6)$$

    where p_{jA} is the price of a trip and t_{jA} is travel time of alternative j
  - $\beta < 0, \kappa < 0$

- Recall (from previous slide):
  - $\Pr(Y = A) = \frac{1}{1+e^{\beta(p_{jB}-p_{jA})+\kappa(t_{jB}-t_{jA})}}$

3. Value of time

- Important concept in Transport Economics: *Value of Time* (VOT)
  - "How much are you willing to pay to reduce your travel time with one hour, *holding utility constant*"

- Let's take the deterministic utility function

$$U_{jA} = \beta p_{jA} + \kappa t_{jA} + \varepsilon_{jA} \quad (7)$$

- When t_{jA} is measured in hours, the VOT can be written as κ/β

3. Value of time

- **Important concept in Transport Economics: *Value of Time* (VOT)**
  - **"How much are you willing to pay to reduce your travel time with one hour, *holding utility constant*"**

- **Let's take the deterministic utility function**
  $U_{iA} = \beta p_{iA} + \kappa t_{iA} + \gamma x_{iA} + \varepsilon_{iA}$
  - $\Delta U_{iA} = 0$ (7)

→ **Show that when t_{iA} is measured in hours, the VOT can be written as κ/β**

With $U_{iA} = \beta p_{iA} + \kappa t_{iA} + \gamma x_{iA} + \varepsilon_{iA}$, show that when t_{iA} is measured in hours, the Value of Time can be written as κ/β.

7

I'm ready!

7

→ **Show that when t_{jA} is measured in hours, the VOT can be written as κ/β**

Let's look at a change in utility:

$$\Delta U_{iA} = \beta\Delta p_{iA} + \kappa\Delta t_{iA} + \Delta\varepsilon_{iA}$$

- **We hold utility constant, so** $\Delta U_{iA} = 0$
- **Because** $\mathrm{E}[\varepsilon_{iA}] = 0$, $\Delta\varepsilon_{iA} = 0$
- $-\beta\Delta p_{iA} = \kappa\Delta t_{iA}$
- $\Delta t_{iA} = -1$ **(WTP for one hour reduction in travel time)**
- $\Delta p_{iA} = \frac{\kappa}{\beta}$

VU VRIJE UNIVERSITEIT AMSTERDAM

3. Value of time

- Value of time is often used in cost benefit analyses

- VOT depends on trip purpose
  - Business €26.25/h
  - Commuting €9.25/h
  - Social purpose €7.50/h

- VOT depends on income
  - About 50% of net income

3. Value of time

- Other applications
  - Value of a Statistical Life (VSL)
    The VSL is the local tradeoff between fatality risk and money
  - Value of schedule delay (VSD)
    The VSD is the local tradeoff between being too early/late and money
  - Etc.
  - … What is necessary is a cost/reward parameter in the discrete choice experiment

4. Multiple alternatives

- **The choice probability for two alternatives:**

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

- **Usually there are <u>more alternatives in the choice set</u>**
  - **Train, bus, car**
  - **Rotterdam, Antwerp, Hamburg**
  - **Routes to the VU**

- **Simply extend the logit formula:**

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

4. Multiple alternatives

- **The aggregate utility derived from the choice set is summarised by the logsum:**

  $$\mathrm{E[CS]} = \frac{1}{\upsilon}\ln\left(\mathrm{e}^{\beta x_A} + \mathrm{e}^{\beta x_B} + \mathrm{e}^{\beta x_C}\right)$$

  - υ **is the marginal utility of income**
  - **Can be used in welfare estimates**

- **Assume** $\beta x_A = \beta x_B = 10$

- **Now alternative** C **is added and** $\beta x_C = 1$

- **The average utility per alternative decreases from** 10 **to** 7 **but** $\mathrm{E[CS]}$ **increase**
  - **'Love of variety' effect**

4. Multiple alternatives

- **Property of logit formula:**
  - **The *ratios* of choice probabilities for A and B do not depend on whether or not C is in the choice set**
  - **Independence of irrelevant alternatives**
  - $$\frac{\Pr(Y=A)}{\Pr(Y=B)} = \frac{\left(\frac{e^{\beta x_A}}{e^{\beta x_A}+e^{\beta x_B}+e^{\beta x_C}}\right)}{\left(\frac{e^{\beta x_B}}{e^{\beta x_A}+e^{\beta x_B}+e^{\beta x_C}}\right)} = \frac{e^{\beta x_A}}{e^{\beta x_B}}$$
  - **Let's find out whether this is a desirable property...**

4. Multiple alternatives

- 'The *Red Bus-Blue Bus*' problem

- Choice set 1: Train, red bus, blue bus
- Assume market shares are 70, 15 and 15%

	Train	Red bus	Blue bus
V	2.54	1	1
Prob	0.700	0.150	0.150

- Choice set 2: Train, red bus, so:

	Train	Red bus
V	2.54	1
Prob	0.823	0.177

- Probability to take the bus in choice set 2 is

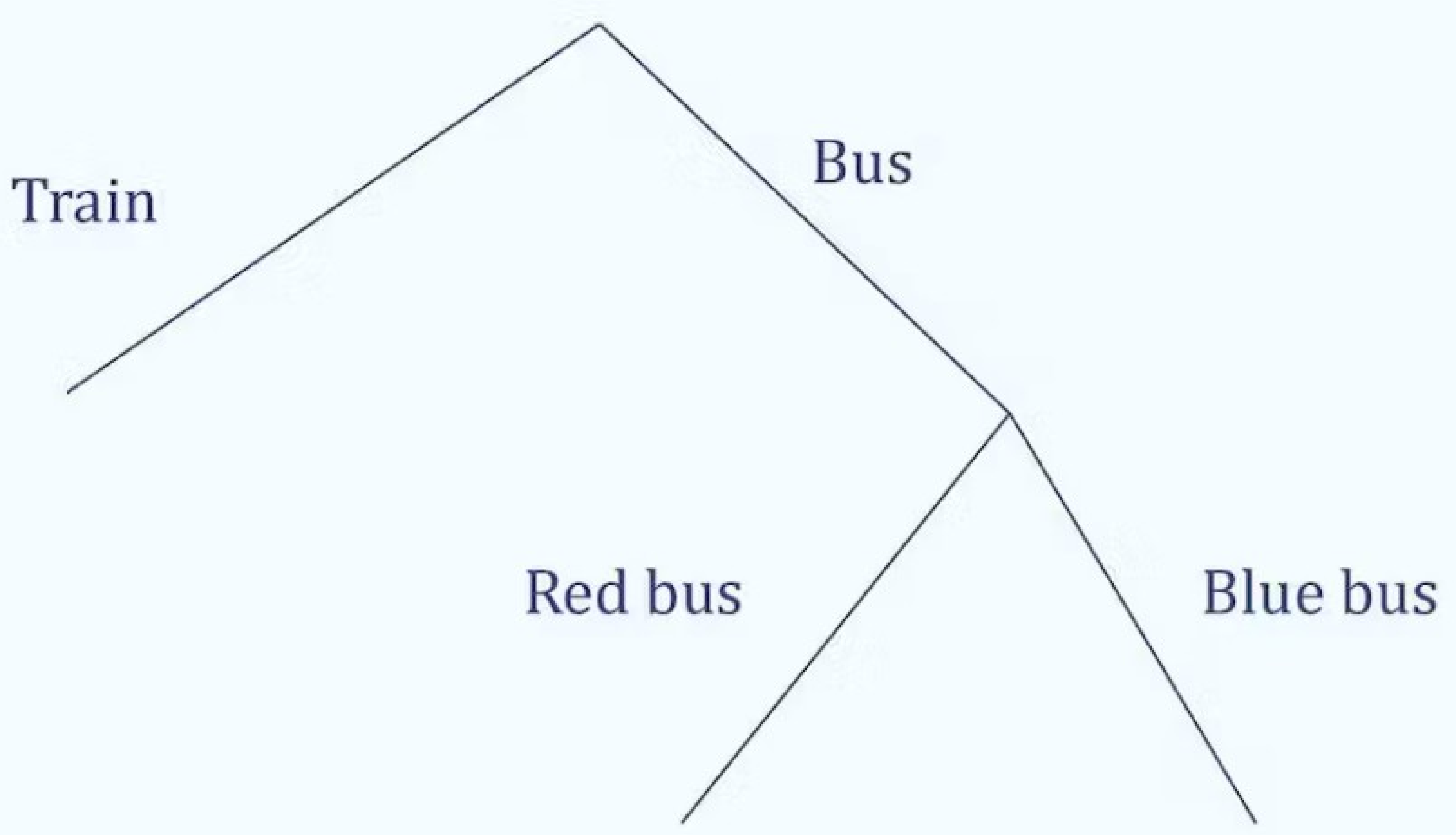
$$\frac{e^{1}}{e^{2.54}+e^{1}} = 0.177$$

  - Higher probability – not very realistic as red buses and blue buses are identical

4. Multiple alternatives

- So, when some alternatives are more similar than other alternatives, the use of multinomial choice model may be misleading

- Use nested logit!



4. Multiple alternatives

- Nested logit takes into account correlation between alternatives
  - But define nests yourself!

2. The random utility framework

- **Let us define utility as follows:**

  $U_{jg} = V_j + W_g + \epsilon_{jg}$

  V_j **only differs within nests between alternatives** j

  W_g **only differs between nests** g

2. The random utility framework

- **We may write the probability to choose an alternative:**
  - $\Pr\left(d_j = 1\right) = \Pr(g) \cdot \Pr(j \mid g)$
  - $\Pr(j \mid g) = \dfrac{e^{V_j/\lambda_g}}{\sum_{k \in g} e^{V_k/\lambda_g}}$
  - $\Pr(g) = \dfrac{e^{W_g + \lambda_g I_g}}{\sum_{\tilde{g}} e^{W_{\tilde{g}} + \lambda_{\tilde{g}} I_{\tilde{g}}}}$

  with $I_g = \log\left(\sum_{j \in g} e^{V_j/\lambda_g}\right)$

- $\lambda_g = 1 \Rightarrow$ **no correlation (multinomial logit)**
- $\lambda_g \to 0 \Rightarrow$ **perfect correlation (red bus/blue bus)**

VRIJE UNIVERSITEIT AMSTERDAM

2. The random utility framework

- **When j and k are in the same nest:**

$$\frac{\Pr(d_j = 1)}{\Pr(d_k = 1)} = \frac{e^{W_g + V_j / \lambda_g}}{e^{W_g + V_k / \lambda_g}} = \frac{e^{W_g + V_j}}{e^{W_g + V_k}} = \frac{e^{V_j}}{e^{V_k}}$$

  - **IIA property holds *within* nests**

- **When $\lambda_g \to 0$:**

  - $\Pr(j \mid g) = \dfrac{e^{V_j/\lambda_g}}{\sum_{k \in g} e^{V_k/\lambda_g}} = 1/k_g$
  - $\Pr(g) = \dfrac{e^{W_g + \lambda_g I_g}}{\sum_{\tilde{g}} e^{W_{\tilde{g}} + \lambda_{\tilde{g}} I_{\tilde{g}}}} = \dfrac{e^{W_g}}{\sum_{\tilde{g}} e^{W_{\tilde{g}}}}$
  - **Hence, multinomial logit *between* nests**

2. The random utility framework

- So, nested logit probability depends on
  - Probability to choose a nest
  - Probability to choose an alternative within the nest

- Note that Nested Logit does not imply a *sequential* choice

Discrete choice (1)

Applied Econometrics for Spatial Economics

Hans Koster
Professor of Urban Economics and Real Estate

Discrete choice (2)

Applied Econometrics for Spatial Economics

Hans Koster
Professor of Urban Economics and Real Estate

1. Introduction

- How to estimate binary discrete choice models?

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

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**

2. Linear probability model

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

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**

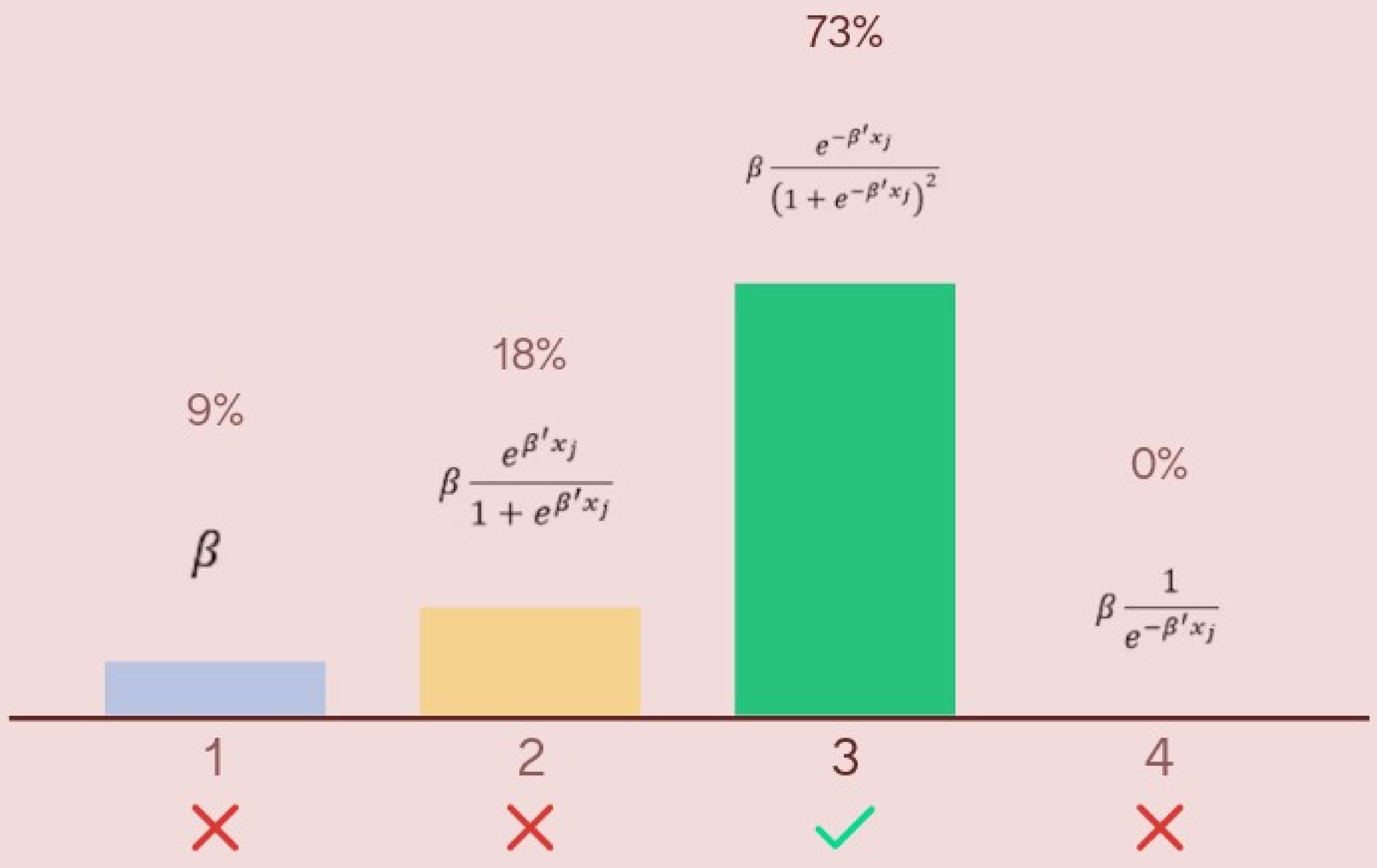
Chosen_B	$\text{Price}_B\text{-Price}_A$	$\text{Time}_B\text{-Time}_A$
1	-14	5
0	5	0
0	15	-20
1	-8	13
1	-10	3
1	3	-5
0	20	10

3. Logit

- **Recall**
  - $\Pr(d_j = 1) = \left(1 + e^{-\beta' x_j}\right)^{-1}$

→ **Calculate the marginal effect on probability of one unit increase in x (so $\frac{\partial \Pr(d_j=1)}{\partial x_j}$)**

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



11

- **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**
  - $$\frac{\partial \Pr(d_j=1)}{\partial x_j} = \beta \frac{e^{-\beta' x_j}}{\left(1+e^{-\beta' x_j}\right)^2}$$

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

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)**

VU VRIJE UNIVERSITEIT AMSTERDAM

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. t*o control for individual heterogeneity

VRIJE UNIVERSITEIT AMSTERDAM

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

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

VU VRIJE UNIVERSITEIT AMSTERDAM

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

VU VRIJE UNIVERSITEIT AMSTERDAM

5. Application

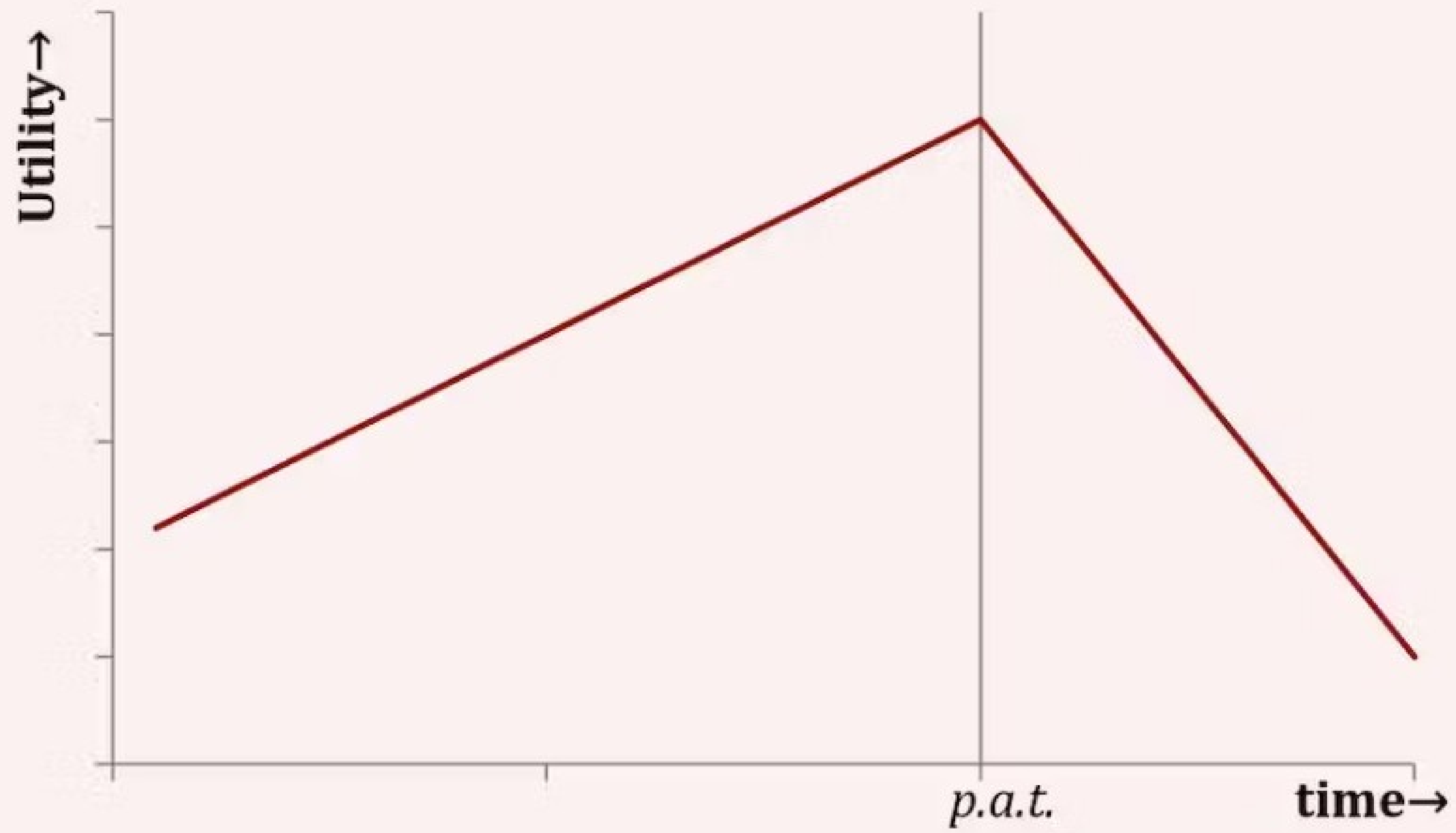
- Koster and Koster (2014)
  - Estimate the value of time and unreliability
  - Uses a stated choice experiment

- 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

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**

VU VRIJE UNIVERSITEIT AMSTERDAM

5. Application

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

VU VRIJE UNIVERSITEIT AMSTERDAM

5. Application

- **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):**

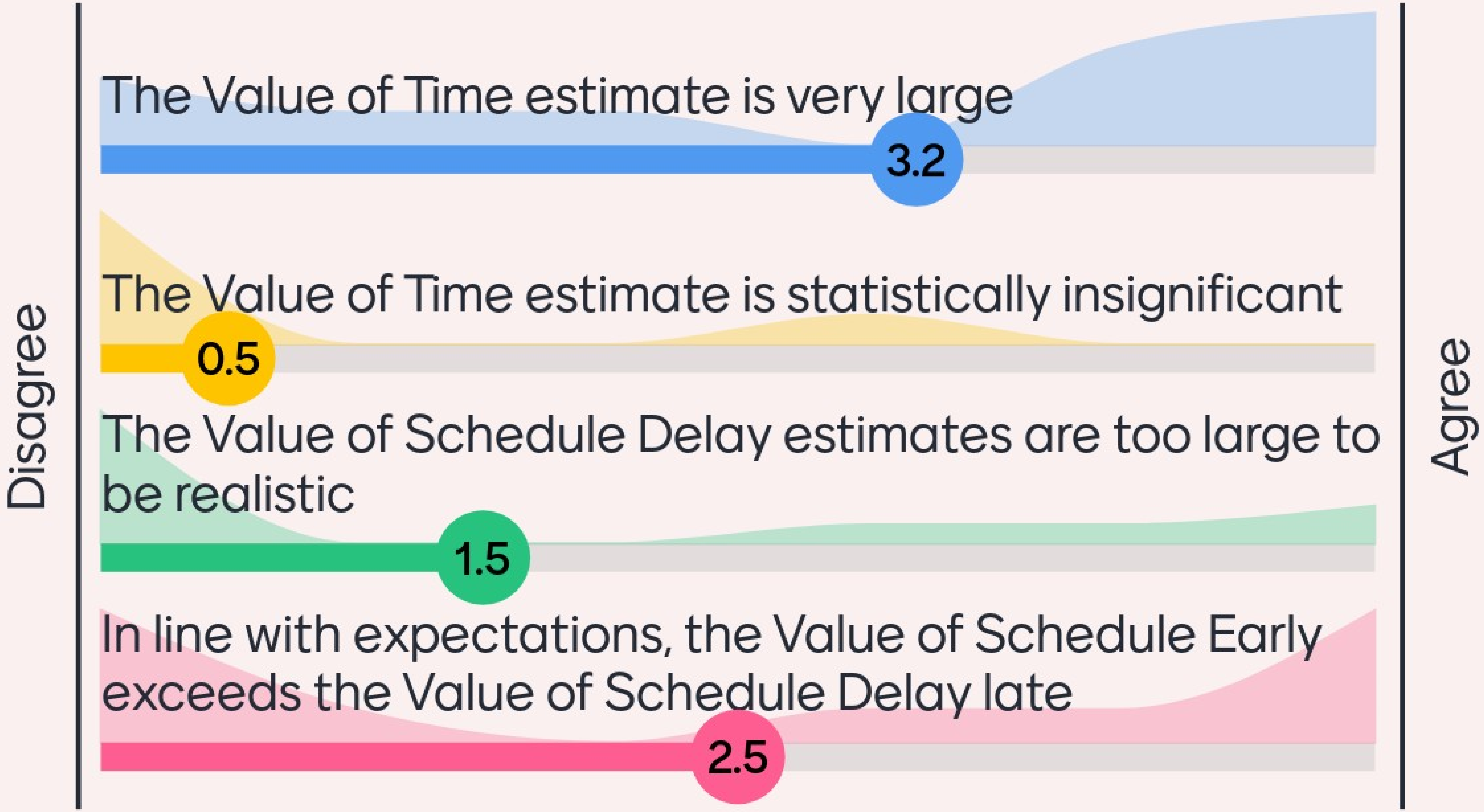
$$-\beta^R \Delta R_{icj} = \beta^{SDE} \Delta SDE_{icj} \rightarrow \Delta SDE_{icj} = -1 \rightarrow -\Delta R_{icj} = -\frac{\beta^{SDE}}{\beta^R}$$

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

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

VRIJE UNIVERSITEIT AMSTERDAM

What is your interpretation of the results?



11

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

Discrete choice (2)

Applied Econometrics for Spatial Economics

Hans Koster

Professor of Urban Economics and Real Estate

VRIJE UNIVERSITEIT AMSTERDAM

tinbergen institute

Discrete choice (3)

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

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.:**

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

2. Multinomial logit

- All the coefficients are compared to one base category!
- Coefficients are different for different alternatives
- Particularly useful when outcomes do not have a logical ordering
  - Bus, car, train
  - Holiday destinations
  - Otherwise: OLS or Ordered Logit
- If the number of alternatives is very large ➔ too many coefficients to interpret meaningfully

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

VRIJE UNIVERSITEIT AMSTERDAM

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:

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

VU 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)

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

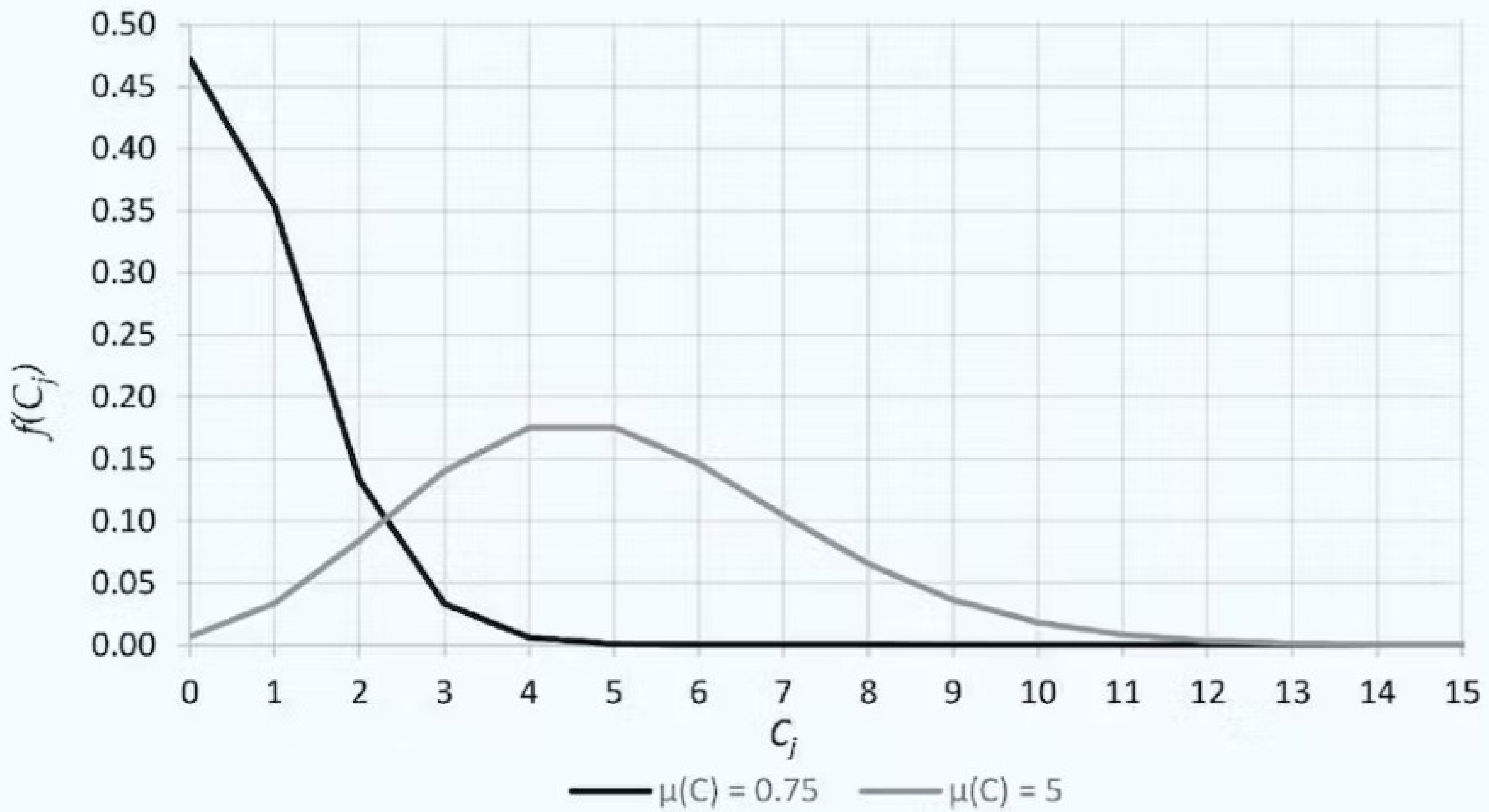
- Estimate Conditional Logit by means of a Poisson model

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

VU VRIJE UNIVERSITEIT AMSTERDAM

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

VU VRIJE UNIVERSITEIT AMSTERDAM

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

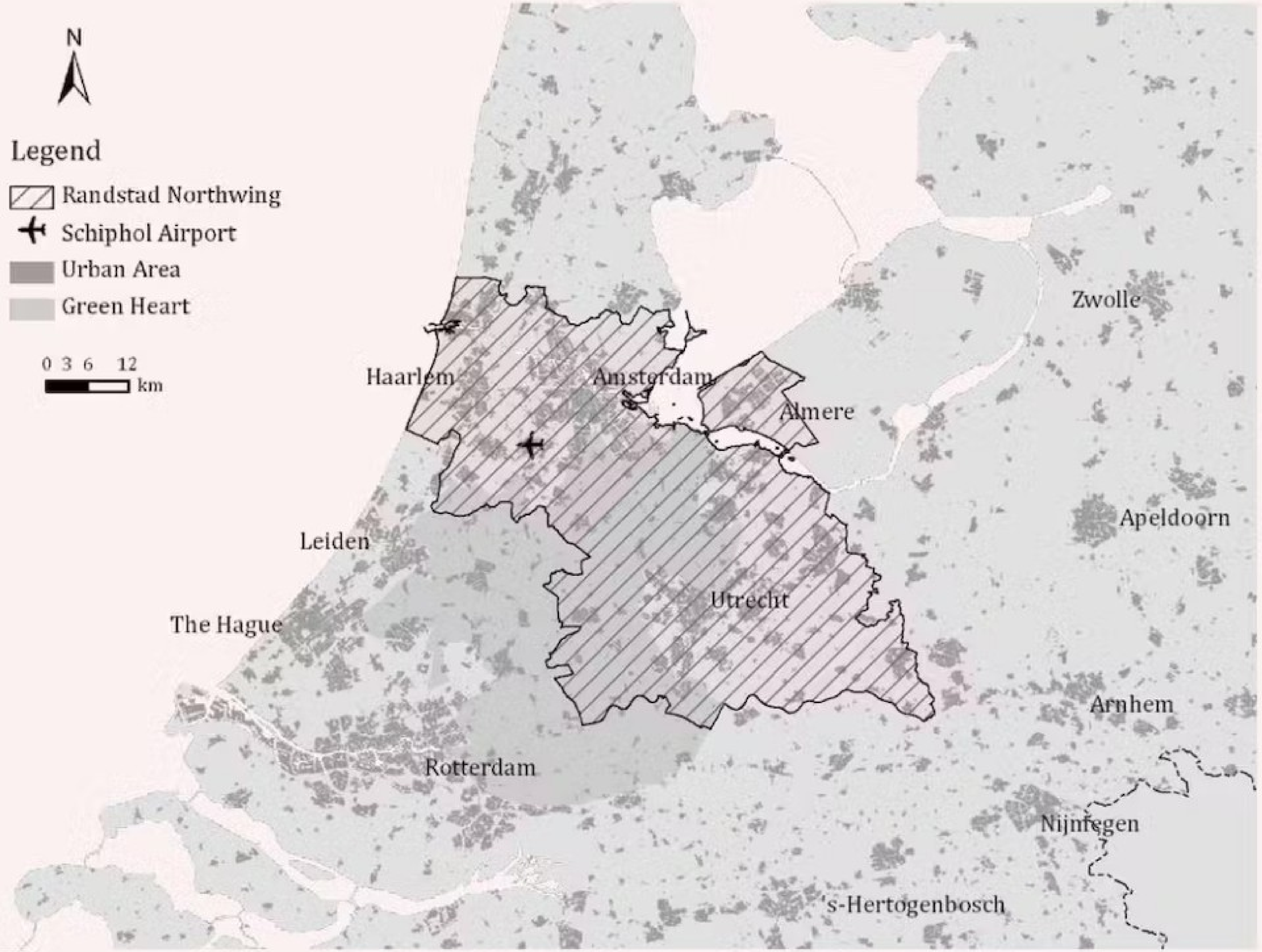
4. Conditional logit

- 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

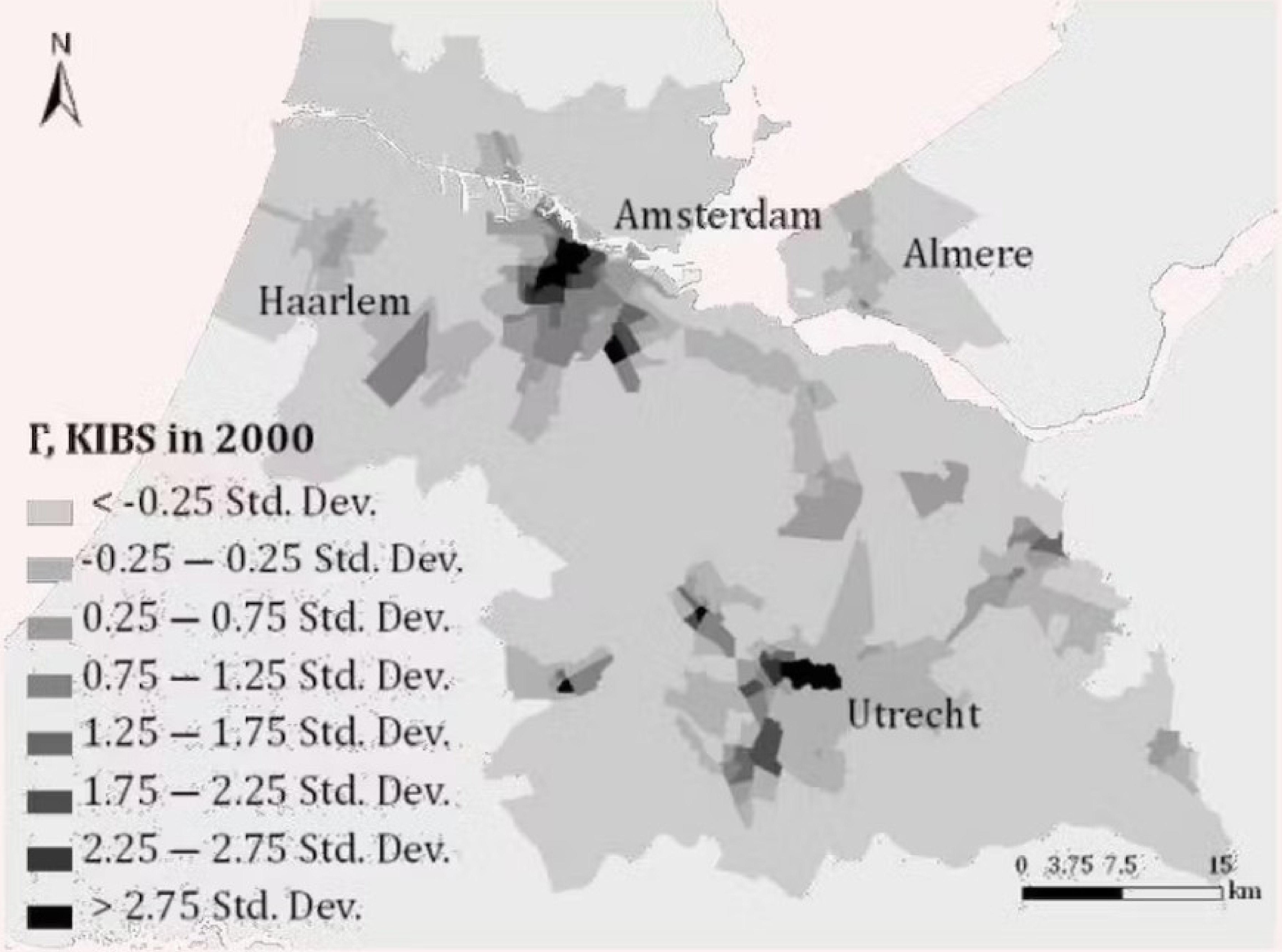
- **The Randstad Northwing**



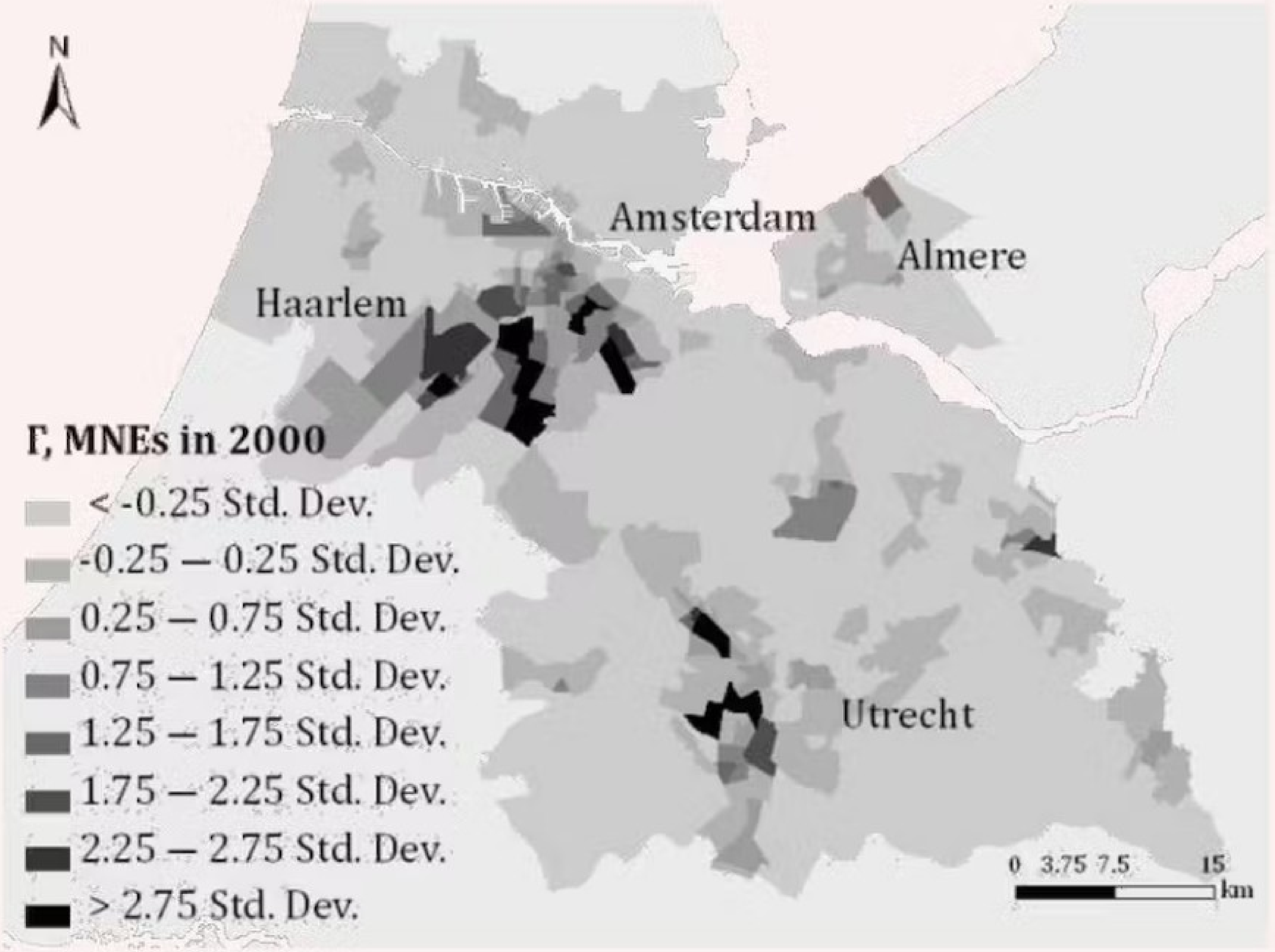
VRIJE UNIVERSITEIT AMSTERDAM

4. Conditional logit

- **Business services in the Randstad Northwing**



- **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**

4. Conditional logit

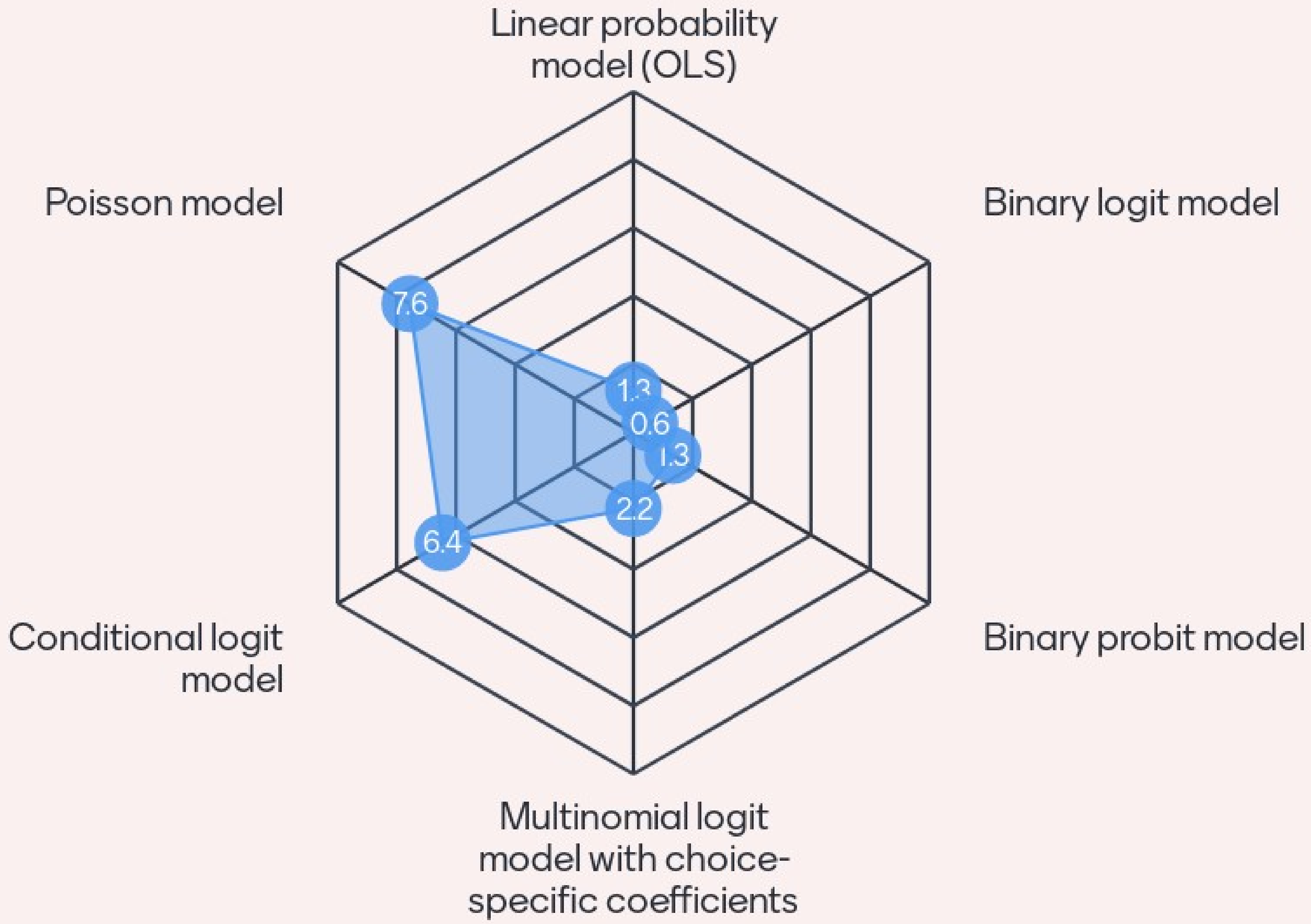
- **Probability Π that i chooses k:**

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

- **There are 13,655 locations**

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

What regression method would you use to estimate this model?



9

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

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))

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

VRIJE UNIVERSITEIT AMSTERDAM

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?			

VRIJE UNIVERSITEIT AMSTERDAM

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

5. Revealed preference and stated preference data

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

- <u>Hypothetical bias</u>
  - 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…

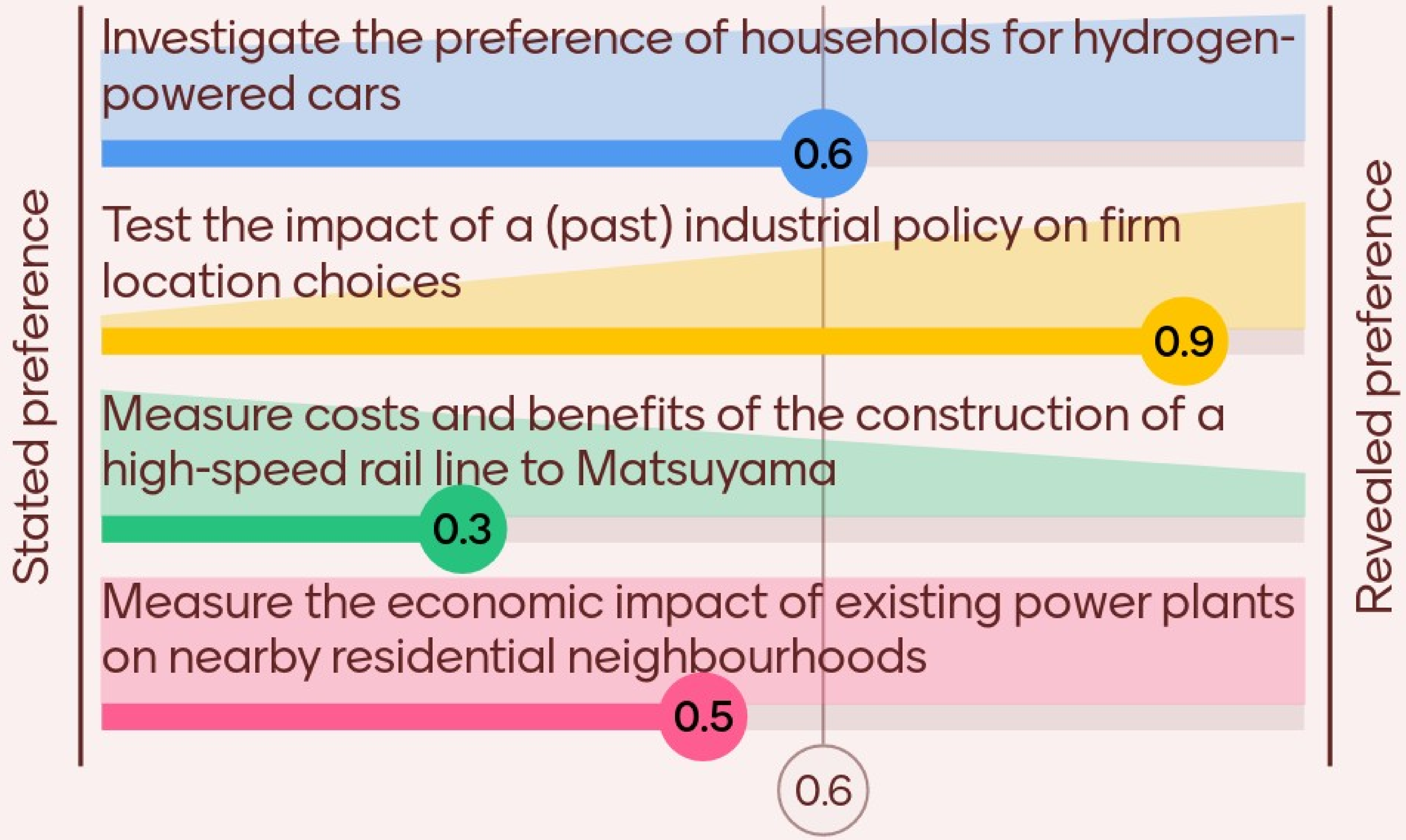
- <u>Strategic bias</u>
  - Respondent wants a specific outcome
  - (S)he fills in answers that are in line with desired outcomes

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:



4
12

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

Discrete choice (3)

Applied Econometrics for Spatial Economics

Hans Koster
Professor of Urban Economics and Real Estate