The luxury (and curse) of variety:
choosing between data on real, hypothetical and virtual travel behaviour, and avoiding fake data

Stephane Hess

2019 Colombian Congress of Transit and Transport
Cartagena
26 June 2019
Outline

• Travel behaviour modelling and data
• Hypothetical data vs real data
• Real data: GPS
• Real data: mobile phone records
• Virtual reality data
• Avoiding fake data
• Conclusions
Travel behaviour modelling and data
Need understanding and prediction of demand
How does it work?

Valuation of individual components and overall valuation

Forecasting of choices/demand in specific scenarios
Hypothetical data vs real data
Need data, either real or hypothetical
Background

• Initially, travel behaviour analysis relied on “real” data
• Actual trips made by actual people
• Expensive, often with missing data, and difficult/impossible to look at new scenarios
• Increasingly replaced by stated preference data, i.e. hypothetical choices
• We have become very lazy
Can you make this choice for a hypothetical journey?

Below are 4 different travel options for your 2 day trip from your home to Boston. Assume that none of the options require a transfer or connection.

If the options below are the only options available for your trip, which would you prefer?

Highlighted information may have changed.

<table>
<thead>
<tr>
<th>Option 1: Train</th>
<th>Option 2: Personal Car</th>
<th>Option 3: Air</th>
<th>Option 4: Bus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time driving to station &amp; time at station: 0 hr 15 min</td>
<td>Time in car: 3 hr 40 min</td>
<td>Time driving to airport, check-in &amp; security: 1 hr 30 min</td>
<td>Time driving to station &amp; time at station: 0 hr 15 min</td>
</tr>
<tr>
<td>On-board travel time: 3 hr 30 min</td>
<td></td>
<td>Time in plane: 1 hr 3 min</td>
<td>On-board travel time: 3 hr 22 min</td>
</tr>
<tr>
<td>Destination station to final destination: 0 hr 38 min</td>
<td></td>
<td>Airport to final destination: 0 hr 31 min</td>
<td>Destination station to final destination: 0 hr 30 min</td>
</tr>
<tr>
<td>Total Travel Time: 4 hr 23 min</td>
<td>Total Travel Time: 3 hr 40 min</td>
<td>Total Travel Time: 3 hr 24 min</td>
<td>Total Travel Time: 4 hr 7 min</td>
</tr>
<tr>
<td>Parking fees for total trip: $68.00</td>
<td></td>
<td>One-way gas costs: $32.00</td>
<td></td>
</tr>
<tr>
<td>One-way cost per person: $112.00</td>
<td>Implied one-way cost per person (1/2 of parking fees + one-way gas costs): $34.00</td>
<td>One-way cost per person: $125.00</td>
<td>One-way cost per person: $60.00</td>
</tr>
<tr>
<td>One-way cost for entire party of 2: $224</td>
<td>Implied one-way cost for entire party of 2: $68.00</td>
<td>One-way cost for entire party of 2: $250</td>
<td>One-way cost for entire party of 2: $120</td>
</tr>
</tbody>
</table>
How many of you would choose Uber Elevate?
Easier to make this choice in a hypothetical setting

Imagine the options below were the only options available for a similar *[purpose]* trip in the future, even if they are not currently available to you. Which option would you most prefer?

<table>
<thead>
<tr>
<th></th>
<th>Per Passenger Cost</th>
<th>Time Comparison (Minutes)</th>
<th>Total Time</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>uberPOOL</strong></td>
<td>$85</td>
<td>75 min</td>
<td>75 minutes</td>
</tr>
<tr>
<td>(save by sharing ride with others)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>uberX</strong></td>
<td>$90</td>
<td>70 min</td>
<td>70 minutes</td>
</tr>
<tr>
<td><strong>uberAIR</strong></td>
<td>$140</td>
<td>10 min + 15 min + 5 min</td>
<td>30 minutes</td>
</tr>
</tbody>
</table>
Background

- Good transport models need good data
- Hypothetical (SP) data good for some things
  - if we can, we should use real data
- Real world travel surveys are expensive
  - conducted rarely
  - small samples
- Great promise of big data
  - thus far largely used only for visualisation and validation
  - we use it for model development
- Increasing interest in virtual reality
Real data: GPS
GPS data is now collected in vast amounts
Two different types of data

• GPS data collected for travel behaviour modelling
• GPS data collected as a by-product (e.g. route guidance)
• Both are missing important information on choice sets
• The latter is also missing information about the traveller
• Focus on this as it shows the most general steps needed
TagMyDay data from Pisa, Italy

- Collected in Pisa, Italy, spring-summer 2014
- Trips tracked with smartphone app
- One example of many such surveys not conducted by or for transport modellers
- Mode availability missing for most people
- No information on mode consideration at the trip level
Spatial distribution
The sample

- People who initially registered: approx. 800
- 130 people annotated all their trips (n. trips = 8325)
- 54 people annotated all their trips and completed both surveys (n. trips = 4854)
- After data cleaning, final sample of 102 people, 5149 trips
Trip characteristics

**N. trips by mode**
- **bus**: 
- **car**: 
- **scooter**: 
- **bike**: 
- **walk**:

**N. trips by purpose**
- **Home**, 1717, 33%
- **Leisure**, 390, 8%
- **Study**, 363, 7%
- **Shopping**, 172, 3%
- **Services**, 283, 6%
- **Groceries**, 311, 6%
- **Work**, 676, 13%
- **Eating out**, 256, 5%
- **Social**, 675, 13%
- **Pick up-Drop off**, 306, 6%

**Average distance by mode**
- **walk**: 
- **bike**: 
- **scooter**: 
- **car**: 
- **bus**:

**Trips by Job Type**
- **Student**, 56%
- **Employed**, 28%
- **Unemployed**, 13%
- **Retired**, 3%
Trip duration and distance was inferred using Google Maps, reproducing the trip by each mode at the recorded date and time.

Trip costs:
- Bus – dependent on distance and urban/inter-urban nature of the trip. Computed using the transportation companies fare tables.
- Car – dependent on average speed, computed using the cost curve provided by the WebTAG tool of DFT
- Scooter - dependent on average speed, computed the Copert emission models and other data.
Data processing – elevation and weather

• Elevation data from Google API Elevation. Gradients and total ascending/descending computed for each trip.

• Weather archive used to retrieve daily information about rain, fog, temperature, humidity.

• Heat index computed for each day:

\[
HI = -42.379 + 2.04901523 \times T + 10.14333127 \times RH - 0.22475541 \times T \times RH - 0.00683783 \times T^2 - 0.05481717 \times RH^2 + 0.00122874 \times T^2 \times RH + 0.00085282 \times T \times RH^2 - 0.00000199 \times T^2 \times RH^2
\]

where \( T \) is temperature in degrees F and \( RH \) is relative humidity in percent. \( HI \) is the heat index expressed as an apparent temperature in degrees F.
Data processing – mode availability

Mode availability reported by approx. 50% of the sample.

Bus availability derived from network data on a hourly basis

Walk considered to be available to everyone
Availability ≠ Consideration
Dealing with missing availability/consideration

• Availability component (person specific) – car, scooter, bike
  • inferred probabilistically, calibrated on small sample with observed availability

• Consideration component (person-trip specific) – walk, bike
  • probabilistic function of distance, purpose and weather

• Choice model component (mode choice)
A latent class model with two levels of classes

• Mode availability calibrated on data for respondents who provided it
  \[ L_n = \sum_{s=1}^{S} \pi_{s,n} \prod_{t=1}^{T} \sum_{c=1}^{C} \theta_{c,n,t} P_{c,s,n,t} \]

• Then treated probabilistically for all

• Mode consideration is function of trip characteristics, including weather

\[ \theta_{1,n,t} = \kappa_{biket, n,t} \cdot \kappa_{walkt, n,t} \]

\[ \theta_{2,n,t} = \kappa_{biket, n,t} \cdot (1 - \kappa_{walkt, n,t}) \]

\[ \kappa_{m,n,t} = \frac{1}{1 + e^{-r_{m,n,t}}} \]

Consideration probability coefficients

<table>
<thead>
<tr>
<th>Distance</th>
<th>Walk</th>
<th>5.874</th>
<th>3.08</th>
</tr>
</thead>
<tbody>
<tr>
<td>Square of distance</td>
<td>Walk</td>
<td>-1.098</td>
<td>-3.18</td>
</tr>
<tr>
<td>Purpose=groceries</td>
<td>Walk</td>
<td>5.064</td>
<td>1.35</td>
</tr>
<tr>
<td>Distance</td>
<td>Bicycle</td>
<td>-0.060</td>
<td>-0.32</td>
</tr>
<tr>
<td>Heat index</td>
<td>Bicycle</td>
<td>0.027</td>
<td>2.09</td>
</tr>
</tbody>
</table>

Utility coefficients

<table>
<thead>
<tr>
<th>Mode</th>
<th>Estimate</th>
<th>Rob t-rat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bicycle</td>
<td>-1.931</td>
<td>-0.68</td>
</tr>
<tr>
<td>Scooter</td>
<td>-4.930</td>
<td>-1.37</td>
</tr>
<tr>
<td>Car</td>
<td>1.244</td>
<td>0.43</td>
</tr>
<tr>
<td>Bus</td>
<td>-2.552</td>
<td>-0.90</td>
</tr>
<tr>
<td>Walk</td>
<td>-3.453</td>
<td>-11.67</td>
</tr>
<tr>
<td>Bicycle</td>
<td>-1.052</td>
<td>-4.40</td>
</tr>
<tr>
<td>Scooter</td>
<td>-0.263</td>
<td>0.94</td>
</tr>
<tr>
<td>Car</td>
<td>0.198</td>
<td>4.60</td>
</tr>
<tr>
<td>Bicycle</td>
<td>0.695</td>
<td></td>
</tr>
<tr>
<td>Scooter</td>
<td>-0.263</td>
<td>-6.67</td>
</tr>
<tr>
<td>Car</td>
<td>-0.278</td>
<td>-6.43</td>
</tr>
<tr>
<td>Bus</td>
<td>-0.137</td>
<td>-5.30</td>
</tr>
<tr>
<td>Scooter</td>
<td>-1.058</td>
<td>-3.41</td>
</tr>
<tr>
<td>Walk</td>
<td>0.707</td>
<td>2.40</td>
</tr>
<tr>
<td>Bicycle</td>
<td>1.126</td>
<td>2.91</td>
</tr>
<tr>
<td>Scooter</td>
<td>1.076</td>
<td>2.37</td>
</tr>
<tr>
<td>Car</td>
<td>0.997</td>
<td>2.39</td>
</tr>
<tr>
<td>Bus</td>
<td>-0.137</td>
<td>-5.30</td>
</tr>
<tr>
<td>Scooter</td>
<td>0.445</td>
<td>1.38</td>
</tr>
<tr>
<td>Walk</td>
<td>0.707</td>
<td>2.40</td>
</tr>
<tr>
<td>Bicycle</td>
<td>1.126</td>
<td>2.91</td>
</tr>
<tr>
<td>Scooter</td>
<td>1.076</td>
<td>2.37</td>
</tr>
<tr>
<td>Car</td>
<td>0.997</td>
<td>2.39</td>
</tr>
<tr>
<td>Bus</td>
<td>-0.137</td>
<td>-5.30</td>
</tr>
<tr>
<td>Scooter</td>
<td>0.445</td>
<td>1.38</td>
</tr>
<tr>
<td>Walk</td>
<td>0.707</td>
<td>2.40</td>
</tr>
<tr>
<td>Bicycle</td>
<td>1.126</td>
<td>2.91</td>
</tr>
<tr>
<td>Scooter</td>
<td>1.076</td>
<td>2.37</td>
</tr>
<tr>
<td>Car</td>
<td>0.997</td>
<td>2.39</td>
</tr>
<tr>
<td>Bus</td>
<td>-0.137</td>
<td>-5.30</td>
</tr>
<tr>
<td>Scooter</td>
<td>0.445</td>
<td>1.38</td>
</tr>
</tbody>
</table>
Incorporation of latent availabilities and consideration improves model fit

Model highlights relevance of weather, trip purpose and socio-demographics

Produces reasonable values of travel time: 14.9€/h scooter, 15.8 €/h car, 7.7 €/h bus

Not bad for data collected with no interest in travel demand modelling!
Real data: mobile phone records
Two types of mobile phone data

- Event-driven (e.g. CDR and cell phone handover data)
  ➔ Discontinuous trajectories - Readily available

<table>
<thead>
<tr>
<th>Anonymous User ID</th>
<th>Date</th>
<th>Time</th>
<th>Duration</th>
<th>Tower Longitude</th>
<th>Tower latitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABH03JACKAAAgfBALA</td>
<td>20120622</td>
<td>11:41:49</td>
<td>25</td>
<td>23.9339</td>
<td>90.2931</td>
</tr>
<tr>
<td>ABH03JAC8AAAbZfAHW</td>
<td>20120622</td>
<td>13:43:25</td>
<td>13</td>
<td>23.7931</td>
<td>90.2603</td>
</tr>
<tr>
<td>ABH03JAC4AAAcvbABB</td>
<td>20120622</td>
<td>13:27:39</td>
<td>8</td>
<td>23.7761</td>
<td>90.4261</td>
</tr>
<tr>
<td>ABH03JAC9AAAbWFAVV</td>
<td>20120622</td>
<td>15:27:27</td>
<td>51</td>
<td>23.7097</td>
<td>90.4036</td>
</tr>
<tr>
<td>ABH03JABkAAHvEkAQX</td>
<td>20120622</td>
<td>18:32:38</td>
<td>50</td>
<td>23.7386</td>
<td>90.4494</td>
</tr>
</tbody>
</table>

- Network-driven (e.g. GSM and location area update data)
  ➔ Continuous trajectories – Not readily available

<table>
<thead>
<tr>
<th>Anonymous User ID</th>
<th>GSM Cell ID</th>
<th>Unix timestamp</th>
<th>Time zone</th>
</tr>
</thead>
<tbody>
<tr>
<td>8851</td>
<td>712</td>
<td>1251762486</td>
<td>-7200</td>
</tr>
<tr>
<td>8851</td>
<td>712</td>
<td>1251762546</td>
<td>-7200</td>
</tr>
<tr>
<td>8851</td>
<td>836</td>
<td>1251762606</td>
<td>-7200</td>
</tr>
<tr>
<td>8851</td>
<td>836</td>
<td>1251762663</td>
<td>-7200</td>
</tr>
</tbody>
</table>
Previous studies using CDRs

- Road Usage Pattern
- Business Catchment Patterns
- Commute and Migration Pattern
- Route Identification
Travel demand models

- Econometric models of travel demand
  - Travel Demand = f(Demographics, Attributes)
- Traditional household travel surveys - expensive, lower update frequencies, fatigue effects, and prone to reporting errors
- Aim: Eliminate the need to fill travel diaries using GSM data
- Only ask for demographic information and mobile phone data access permission
- Application of such models requires detailed demographic information e.g. from census data - Lower update frequencies (e.g. once in 5-10 years)
- Predict the demographics using phone usage behaviour extracted from CDR data
Trip generation modelling using mobile phone data

- GSM data records continuous movements
  ➔ But no socio-demographic information
- Subsample with socio-demographic information
  - Relate call behavior from CDR data (call detail record) to socio-demographic information
- Can then infer socio-demographic data for all
- Tested using Nokia Mobile Data Challenge dataset from Lausanne, Switzerland
  - 50.8 million GSM records and 0.5 million CDRs
The Nokia MDC data

- Collected in Switzerland in 2011
- Call logs, GSM locations and demographic information from 158 users
Example of demographic prediction on basis of CDR data
With more calls made between 6AM and 8AM, individual is most likely to be male, working, and older than 21

Results for trip generation model
Model using inferred demographics gives results in line with model using observed demographics
➔ But of course it can be used for the full sample, where socio-demographics missing for most
Validation on subsets
Route choice modelling using CDR data

- Unlike GSM data, CDR data readily available
- We use Orange data from Senegal (2013)
- 9 million unique users
- Use monthly subsets of approximately 150,000 people making 40 million calls
- But CDR data is not continuous
- Route can be only inferred if user makes phone calls during their journey
- Possible that several routes exist that include the locations of these calls
Unique (70%) and unclear (30%) choices

Example - Northern 1 (Unique)
Dakar – 71- Bakel

Example - Northern 1/Northern 2
Dakar – 122- Bakel

<table>
<thead>
<tr>
<th>Anonymised User ID with monthly identifier (e.g. January)</th>
<th>CDR Trajectory</th>
<th>Route/ Broad sub-group</th>
</tr>
</thead>
<tbody>
<tr>
<td>131891.01</td>
<td>Dakar-11-C1-Bakel</td>
<td>Northern 1</td>
</tr>
<tr>
<td>131891.01</td>
<td>Dakar-C1-Bakel</td>
<td>Northern 1</td>
</tr>
<tr>
<td>132801.01</td>
<td>Dakar-122-Bakel</td>
<td>Northern (Northern 1/Northern 2)</td>
</tr>
<tr>
<td>132801.01</td>
<td>Dakar-28-C2-123-Bakel</td>
<td>Northern 2</td>
</tr>
<tr>
<td>132801.01</td>
<td>Dakar-C1-123-Bakel</td>
<td>Northern 1</td>
</tr>
</tbody>
</table>
Extraction of trajectories between regions of interest

- We first order the data according to user IDs and timestamps
- This gives the monthly mobility trajectory of each unique user
- We then filter-out sub-trajectories between regions of interest (duration < 24 hrs)
Dakar-Bakel as an example
Northern route
Central route

KEY

- Nth route arrx
- Ctr route arrx
- Sth1 route arrx
- Sth2 route arrx
- Shared Nth - Ctr
- Shared Ctr – Sth1
- Shared Sth1 – Sth2
- Shared Nth-Ctr-Sth1
- Dakar
- Bakel
Central route (scenario – 2)
South-1 route
South-2 route

KEY

- Nth route arrx
- Ctr route arrx
- Sth1 route arrx
- Sth2 route arrx
- Shared Nth - Ctr
- Shared Ctr – Sth1
- Shared Sth1 – Sth2
- Shared Nth-Ctr-Sth1
- Dakar
- Bakel
Proportionally distributed between South-1 & 2 routes in a random manner

KEY
- Yellow: Nth route arrx
- Grey: Ctr route arrx
- Light purple: Sth1 route arrx
- Pink: Sth2 route arrx
- Black: Shared Nth - Ctr
- White: Shared Ctr – Sth1
- Dark grey: Shared Sth1 – Sth2
- Green: Shared Nth-Ctr-Sth1
- Red: Dakar
- Green: Bakel
Modelling framework

Cases with unique choice

\[ P_{nr} = \frac{\exp(V_{nr})}{\sum_{r^* \in C_{n}} \exp(V_{nr^*})} \]

Cases with unclear choice

**Broad Choice Framework** of Brownstone and Li (2017)

\[ P_{nb} = \sum_{r \in S_n} P_n(r) \]

Accounting for route overlap

**C-Logit model**

\[ P_{nr} = \frac{\exp(V_{nr} + CF_{nr})}{\sum_{r^* \in C_n} \exp(V_{nr^*} + CF_{nr^*})} \]

\[ CF_{nr} = \beta_0 \ln \left( \sum_{r^* \in C_n} \left( \frac{L_{nr^*}}{\sqrt{L_{nr^*} L_{r^*}}} \right)^y \right) \]

**Path size logit model**

\[ P_{nr} = \frac{\exp(V_{nr} + \ln PS_{nr})}{\sum_{r^* \in C_n} \exp(V_{nr^*} + \ln PS_{nr^*})} \]

\[ PS_{nr} = \sum_{a \in \Gamma_r} \left( \frac{1}{L_r} \right) \frac{1}{N_{ar}} \]
## Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>MNL</th>
<th>C-Logit</th>
<th>Path Size Logit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Parameter</td>
<td>t-stat</td>
<td>Parameter</td>
</tr>
<tr>
<td>Natural log. Travel time (Hrs)</td>
<td>-4.2221</td>
<td>-11.60</td>
<td>-4.1644</td>
</tr>
<tr>
<td>Natural log. Cost (USD)</td>
<td>-1.6668</td>
<td>-13.78</td>
<td>-1.5987</td>
</tr>
<tr>
<td>Natural log. Av dist between towns</td>
<td>0.5081</td>
<td>2.54</td>
<td>0.1705</td>
</tr>
<tr>
<td>CF Beta</td>
<td></td>
<td>-</td>
<td>-0.0063</td>
</tr>
<tr>
<td>CF Gamma</td>
<td></td>
<td>-</td>
<td>0.6563</td>
</tr>
<tr>
<td>Path size term</td>
<td></td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

### Values in USD/hr and 2013 prices

<table>
<thead>
<tr>
<th>Model</th>
<th>Values in USD/hr and 2013 prices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VTT current study</td>
</tr>
<tr>
<td>MNL</td>
<td>1.0822</td>
</tr>
<tr>
<td>C-logit model</td>
<td>1.0524</td>
</tr>
<tr>
<td>Path size logit model</td>
<td>0.6846</td>
</tr>
</tbody>
</table>

*Note: Values in USD/hr and 2013 prices are sourced from different studies.*
Virtual reality data
Why use VR?

• Some scenarios are difficult to capture in reality
• Huge ethics issues too
• But giving “dangerous” choices in a pure SP context is not very realistic
• Lack of immersion
• Longstanding interest in driving simulators, etc
• Increasing move to VR for immersive data collection
Our current work relies on VR and EEG
VR experimental procedure

• 24 simulations of risky road scenarios for cyclists

• 3 behavioural responses (acceleration, braking, freewheeling)

• Also stated assessment of riskiness of scenarios and willingness to cycle (1-7 scale)
Example of pavement scenario
Proposed model framework

- Experienced risk
- Stated risk
- Willingness to cycle
- Cycling behaviour
- Traffic attributes
- EEG
Correlations between stated variables

• Inverse relationship between risk and willingness to cycle (1)
• Positive correlation between scenario riskiness and stated risk (2)
• Negative relation between scenario riskiness and willingness to cycle (3)

<table>
<thead>
<tr>
<th></th>
<th>Stated risk</th>
<th>Willingness to cycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Willingness to cycle</td>
<td>-0.55</td>
<td></td>
</tr>
<tr>
<td>Scenario riskiness</td>
<td>0.17</td>
<td>-0.15</td>
</tr>
</tbody>
</table>
## DCM example: MNL model pavement

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Rob.std.err</th>
<th>Rob.t.ratio(0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>currently accelerating</td>
<td>ASC for accelerating</td>
<td>0.2593</td>
<td>0.095</td>
</tr>
<tr>
<td></td>
<td>ASC for braking</td>
<td>-2.4971</td>
<td>0.2606</td>
</tr>
<tr>
<td></td>
<td>ASC for freewheeling</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>currently braking</td>
<td>ASC for accelerating</td>
<td>2.826</td>
<td>0.151</td>
</tr>
<tr>
<td></td>
<td>ASC for braking</td>
<td>3.7918</td>
<td>0.132</td>
</tr>
<tr>
<td></td>
<td>ASC for freewheeling</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>currently freewheeling</td>
<td>ASC for accelerating</td>
<td>0.1443</td>
<td>0.0767</td>
</tr>
<tr>
<td></td>
<td>ASC for braking</td>
<td>-3.1188</td>
<td>0.1757</td>
</tr>
<tr>
<td></td>
<td>ASC for freewheeling</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>shifts for 3D</td>
<td>ASC for accelerating</td>
<td>0.0124</td>
<td>0.0292</td>
</tr>
<tr>
<td></td>
<td>ASC for braking</td>
<td>0.1577</td>
<td>0.0593</td>
</tr>
<tr>
<td></td>
<td>ASC for freewheeling</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>pedestrians within 3 metres in front</td>
<td>gain in utility for accelerating</td>
<td>-0.0058</td>
<td>0.0009</td>
</tr>
<tr>
<td></td>
<td>gain in utility for braking</td>
<td>-0.0101</td>
<td>0.0022</td>
</tr>
<tr>
<td></td>
<td>gain in utility for freewheeling</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>pedestrians within 3 metres behind</td>
<td>gain in utility for accelerating</td>
<td>-0.001</td>
<td>0.0012</td>
</tr>
<tr>
<td></td>
<td>gain in utility for braking</td>
<td>0.0035</td>
<td>0.0025</td>
</tr>
<tr>
<td></td>
<td>gain in utility for freewheeling</td>
<td>0</td>
<td>-</td>
</tr>
</tbody>
</table>
Dynamic EEG and behaviour

Alpha and theta waves and cycling behaviour for a single participant in one scenario
Increased realism: real bike, VR cycling
Avoiding fake data
Not all survey companies are honest

• Increasingly collect data over the internet
• When we don’t programme this ourselves, we rely on survey companies
• Three bad experiences
  • Web robot programmed by Chinese survey company for vehicle choice study
  • Exact 50-50 split in data on route choice in an online survey in the US
  • Interviewers add answers for incomplete face-to-face interviews, or just fill in entire survey forms for some “respondents”
Conclusions
Summary & conclusions

• It is possible to estimate advanced travel demand models on mobile phone data and GPS data not collected for this purpose

• Of course there are problems with such data
  • But they are “real” problems as opposed to “hypothetical” problems as in SP

• Also increasing interest in virtual reality
  • But should be careful not to think this is “real” behaviour

• SP remains an important part of our toolkit
  • But we should no longer just assume that SP is always the answer
Thank you!