A latent variable exponential family modeling approach to estimate suppressed demand effects for increasing car travel costs

Basil Schmid

IVT
ETH Zurich

6th hEART Symposium
Haifa, September 14, 2017
Post-Car World: A multi-stage travel survey

- Motivation: Understanding travel behavior in a hypothetical world where privately owned cars are substituted by various forms of shared mobility
- Investigation of pricing mechanisms as a driving force to achieve behavioral reactions
  - Main focus: Transition towards (and not actual state of) such a (Pre-)Post-Car World
- One week travel diary and mobility tool data (stage I) as empirical basis for behavioral experiments (stage II & III)
  - Data collection: Canton of Zurich, 2015 - 2016
  - Average response rate: 55%, N = 220 households
Adaptations in daily scheduling

- How would respondents change their daily travel in the **short-run**, given the increase in travel costs?
- Personalized stated adaptation interviews with mode-specific total RP travel cost $R_{tc,n}$

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Car</td>
<td>$R_{tc,n} \cdot 1.5 + 0.4$</td>
<td>$R_{tc,n} \cdot 2 + 0.8$</td>
<td>$R_{tc,n} \cdot 4 + 1.4$</td>
<td>$R_{tc,n} \cdot 8 + 2$</td>
</tr>
<tr>
<td>Moto</td>
<td>$R_{tc,n} \cdot 1.5 + 0.2$</td>
<td>$R_{tc,n} \cdot 12 + 0.4$</td>
<td>$R_{tc,n} \cdot 4 + 0.7$</td>
<td>$R_{tc,n} \cdot 8 + 1$</td>
</tr>
<tr>
<td>PT</td>
<td>$R_{tc,n} \cdot 1.1$</td>
<td>$R_{tc,n} \cdot 1.2$</td>
<td>$R_{tc,n} \cdot 1.3$</td>
<td>$R_{tc,n} \cdot 1.5$</td>
</tr>
<tr>
<td>CS</td>
<td>$R_{tc,n} \cdot 1.1$</td>
<td>$R_{tc,n} \cdot 1.2$</td>
<td>$R_{tc,n} \cdot 1.3$</td>
<td>$R_{tc,n} \cdot 1.5$</td>
</tr>
<tr>
<td>CP</td>
<td>$R_{tc,n} \cdot 1.5$</td>
<td>$R_{tc,n} \cdot 2$</td>
<td>$R_{tc,n} \cdot 4$</td>
<td>$R_{tc,n} \cdot 8$</td>
</tr>
</tbody>
</table>

- Experimental framing:
  - Road tolls, fuel and congestion taxes
  - Future policy developments to reduce MIV usage
  - Promotion of shared mobility (PT, CS, CP)
Adaptations in daily scheduling

Durchschnittlicher OEV-Takt: 3 min.
Zeit zum nächsten Carsharing Fahrzeug: 3min
Zeit zum nächsten Carpooling Fahrzeug: 3min

<table>
<thead>
<tr>
<th>Aktivität:</th>
<th>Zu Hause</th>
<th>Einkaufen</th>
<th>Fr. Bedarf</th>
<th>Arbeit/Ausbildung</th>
<th>Dienstlich</th>
<th>Zu Hause</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ort der Aktivität:</td>
<td>Zu Hause</td>
<td>Tomac3</td>
<td>Sihlfeldstrasse 53</td>
<td>Seebahnstrasse 8</td>
<td>Plantaweg 21</td>
<td>Zu Hause</td>
</tr>
<tr>
<td>Straße:</td>
<td>Nordstrasse 21</td>
<td>Zuerich</td>
<td>Zuerich</td>
<td>Zuerich</td>
<td>Zuerich</td>
<td>Nordstrasse 21</td>
</tr>
<tr>
<td>Stadt:</td>
<td>Zuerich</td>
<td>08:17</td>
<td>08:24</td>
<td>08:24</td>
<td>08:24</td>
<td>Zuerich</td>
</tr>
<tr>
<td>Ankunftszeit:</td>
<td>00:00</td>
<td>08:05</td>
<td>01:55</td>
<td>11:31</td>
<td>00:44</td>
<td></td>
</tr>
<tr>
<td>Dauer der Aktivität:</td>
<td>08:05</td>
<td>08:22</td>
<td>10:19</td>
<td>13:11</td>
<td>15:18</td>
<td></td>
</tr>
<tr>
<td>Abfahrtszeit:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zu Fuß</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auto (Fahrer)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auto (Mitfahrer)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Velo</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OEV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carpooling (Mitfahrer)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carsharing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motorrad</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>zurückgelegte Distanz:</td>
<td>2.78</td>
<td>0.88</td>
<td>134.19</td>
<td>134.10</td>
<td>2.43</td>
<td></td>
</tr>
<tr>
<td>Reisezeit:</td>
<td>00:12</td>
<td>00:02</td>
<td>01:12</td>
<td>01:23</td>
<td>00:13</td>
<td></td>
</tr>
<tr>
<td>Reisekosten:</td>
<td>0.00</td>
<td>0.00</td>
<td>36.23</td>
<td>36.21</td>
<td>2.20</td>
<td></td>
</tr>
</tbody>
</table>

**Summe Reisekosten (in CHF): 79.04**
Adaptations in daily scheduling

Focus of today:

- Suppressed demand effects for MIV (car driver, car passenger, motorbike) usage: What is the effect on daily mileage driven, given the increase in travel costs?
- "Aggregate" response function (given low sample size) using highly disaggregate data (activity-based perspective)
- Assumption: Cost minimizing behavior, given underlying (unobserved) preferences for daily plan
- "Two-step approach" for modeling (unobserved) heterogeneity
Environmental sensitivity / car loving traits ...

envi1: Higher fuel prices should subsidize public transport
envi2: Daily life without car is impossible
envi3: Car driving is bad for the environment
envi4: I could imagine to give up car usage completely
envi5: Zurich without cars is inconceivable
envi6: Environmental problems get too much attention
envi7: The never-ending discussions about the greenhouse effect is exaggerated
envi8: Fuel prices should increase to reduce pollution of the environment
... and socio-demographic characteristics
Data

- \( N = 162 \) respondents, 810 initial choice scenarios
- Dependent variable: Distance traveled by MIV \( y_{n,t} \equiv km_{n,t} \) after adaptation in \textbf{current} scenario
  - Highly right-skewed data with some zeros (respondents might choose not to use MIV anymore)
  - Pseudo-balanced panel: After drop-out, respondents are excluded \((\rightarrow 735 \text{ actual choice observations})\)
- Main explanatory variable: Average MIV travel cost per km \( x_{n,t} \equiv \log(CHF_{n,t-1}) \) after adaptation in \textbf{previous} scenario
Adaptation patterns in distance traveled

Traveled distance by MIV [km]

[Diagram showing data distribution for different categories B, S1, S2, S3, S4, with box plots and individual data points indicated]
Change in MIV travel cost

Average MIV km cost [CHF/km]

B S1 S2 S3 S4
Modeling framework: GLM

- Log-linear OLS model is **inconsistent**
  - $E[\log(\eta_{n,t}|X_{n,t})] \neq 0$ if CEF is exponential ($\eta_{n,t}$ is LN) and presence of heteroscedasticity (*Jensen’s inequality*)
  - Incompatible with mass point at zero
- Exponential family modeling approach using *pseudo* maximum likelihood techniques (Gourieroux et al., 1984)

\[
f(Y_{n,t}|X_{n,t}, z_n, \Lambda) = \exp \left( \frac{Y_{n,t} f(X_{n,t}, z_n, \Lambda) - b(f(X_{n,t}, z_n, \Lambda))}{a(\phi)} + c(\phi, Y_{n,t}) \right)
\]

→ FOC score vector: GLM **consistent** as long as CEF is correctly specified (Santos-Silva and Tenreyro, 2006)
- Poisson: $E[Y_{n,t}|X_{n,t}, z_n] = \exp(f(X_{n,t}, z_n, \Lambda))$
- Heterosced.: $E[Y_{n,t}|X_{n,t}, z_n] = \text{Var}[Y_{n,t}|X_{n,t}, z_n] = \lambda_{n,t}$
- Globally concave, simple and fast in convergence
Modeling framework: Panel structure

- Large variety in respondents’ characteristics and their daily plans (unobserved heterogeneity)
- Starting point: Poisson regression for a continuous, non-negative dependent variable with mixed effects (Hausman test: $H_0$ plausible $\rightarrow$ RE more efficient)
- Hausman et al. (1984): Equidispersion assumption further relaxed by the RE specification $\text{Var}[Y_{n,t}|X_{n,t}] = \lambda_{n,t} + \theta \lambda_{n,t}^2$
- Huber/White sandwich estimator for SEs (Arellano, 1987)
Modeling framework: Log-linear index

\[
\lambda_{1,n,t} = \epsilon_n \cdot \exp \left( \alpha + \beta_{\text{COST}} \cdot \log(CHF_{n,t-1}) \cdot \left( \frac{\text{dist}_{n,0}}{\text{dist}} \right)^{\omega_{\text{DIST}}} \right)
\]

\[
\lambda_{2,n,t} = \epsilon_n \cdot \exp \left( \alpha + \alpha_{\text{INC}} \cdot \text{inc}_n + \alpha_{\text{ENVI}} \cdot \text{envi}_n + \right.

\quad \left( \beta_{\text{COST}} + \beta_{\text{INC}} \cdot \text{inc}_n + \beta_{\text{ENVI}} \cdot \text{envi}_n \right) \cdot \log(CHF_{n,t-1}) \cdot \left( \frac{\text{dist}_{n,0}}{\text{dist}} \right)^{\omega_{\text{DIST}}} \right)
\]

\[
\lambda_{3,n,t} = \epsilon_n \cdot \exp \left( \alpha - \exp(\beta_{\text{COST}} + \psi_n) \cdot \log(CHF_{n,t-1}) \cdot \left( \frac{\text{dist}_{n,0}}{\text{dist}} \right)^{\omega_{\text{DIST}}} \right)
\]

\[
\lambda_{4,n,t} = \epsilon_n \cdot \exp \left( \alpha + \alpha_{\text{INC}} \cdot \text{inc}_n + \alpha_{\text{ENVI}} \cdot \text{envi}_n

\quad - \exp(\beta_{\text{COST}} + \beta_{\text{INC}} \cdot \text{inc}_n + \beta_{\text{ENVI}} \cdot \text{envi}_n + \psi_n) \cdot \log(CHF_{n,t-1}) \cdot \left( \frac{\text{dist}_{n,0}}{\text{dist}} \right)^{\omega_{\text{DIST}}} \right)
\]
Modeling framework: Estimation (1)

- **Analytical solution (random intercept):** Assuming that $\epsilon_n \sim \Gamma(1, \theta)$ and $y_{n,t}$ is distributed Poisson with mean $\lambda_{s,n,t} \equiv \lambda_{s,n,t}/\epsilon_n$, the likelihood of observing the sequence $Y_{n,t}$ given $X_{n,t}$ and $z_n$ of respondent $n$ is given by

$$L_L(Y_{n,t} | X_{n,t}, z_n, \Lambda) = \log \Gamma \left( \frac{1}{\theta} + \sum_{t=1}^{T_n} y_{n,t} \right) - \sum_{t=1}^{T_n} \log \Gamma \left( 1 + y_{n,t} \right) - \log \Gamma \left( \frac{1}{\theta} \right) + \frac{1}{\theta} \cdot \log(u_n) + \log(1 - u_n) \sum_{t=1}^{T_n} y_{n,t} +$$

$$\sum_{t=1}^{T_n} y_{n,t} \cdot \log \left( \lambda_{s,n,t} \right) - \left( \sum_{t=1}^{T_n} y_{n,t} \right) \log \left( \sum_{t=1}^{T_n} \lambda_{s,n,t} \right)$$
Modeling framework: Estimation (2)

- Simulation (random coefficient or LV): The expected likelihood \( \mathcal{L}_n^*(.) \) over all possible values of \( \psi_n \) or \( LV_n \) is given by the integral of the exponent of the log-likelihood function over the distribution of \( \psi_n \) or \( LV_n \)

\[
\mathcal{L}_n^*(Y_{n,t}, l_{w,n}|X_{n,t}, z_n, \Omega) = \int_{\psi_n, LV_n} \exp(\mathcal{L}_n(Y_{n,t}|X_{n,t}, z_n, \Lambda, \psi_n)) \ u(l_{w,n}|LV_n, \tau_{l_w}, \sigma_{l_w}) \\
\times h(\psi_n|R) \ g(LV_n|z_n, \rho_z, \eta_{LV_z}) \ \ d\psi_n \ dLV_n
\]

\[
\widetilde{\mathcal{L}}_n^*(Y_{n,t}, l_{w,n}|X_{n,t}, z_n, \Omega) = \frac{1}{R} \sum_{r=1}^{R} \exp(\mathcal{L}_n(Y_{n,t}|X_{n,t}, z_n, \Lambda, \psi_n)) \ u(l_{w,n}|LV_n, \tau_{l_w}, \sigma_{l_w})
\]

\[
\max \ \widetilde{\mathcal{L}}(\Omega) = \sum_{n=1}^{N} \log \left( \widetilde{\mathcal{L}}_n^*(Y_{n,t}|X_{n,t}, z_n, \Omega) \right)
\]

→ Posterior analysis of cost elasticity
## Estimation results

<table>
<thead>
<tr>
<th></th>
<th>REP Coef./SE</th>
<th>REPS Coef./SE</th>
<th>LVREP Coef./SE</th>
<th>MEP Coef./SE</th>
<th>MEPS Coef./SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>3.20***</td>
<td>3.15***</td>
<td>3.06***</td>
<td>3.08***</td>
<td>3.05***</td>
</tr>
<tr>
<td>$\alpha_{INC}$</td>
<td>–</td>
<td>0.17</td>
<td>0.16</td>
<td>–</td>
<td>0.16</td>
</tr>
<tr>
<td>$\alpha_{ENVI}$</td>
<td>–</td>
<td>–0.13***</td>
<td>–0.62***</td>
<td>–</td>
<td>–0.11**</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.65***</td>
<td>0.59***</td>
<td>0.51***</td>
<td>1.32***</td>
<td>1.27***</td>
</tr>
<tr>
<td>$\beta_{COST}$</td>
<td>–0.43***</td>
<td>–0.44***</td>
<td>–0.87***</td>
<td>–0.72***</td>
<td>–0.70***</td>
</tr>
<tr>
<td>$\omega_{DIST}$</td>
<td>0.43***</td>
<td>0.47***</td>
<td>0.58***</td>
<td>0.56***</td>
<td>0.58***</td>
</tr>
<tr>
<td>$\beta_{INC}$</td>
<td>–</td>
<td>0.03</td>
<td>–0.08</td>
<td>–</td>
<td>–0.28**</td>
</tr>
<tr>
<td>$\beta_{ENVI}$</td>
<td>–</td>
<td>–0.05***</td>
<td>0.65***</td>
<td>–</td>
<td>0.08</td>
</tr>
<tr>
<td>$\sigma_{COST}$</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1.09***</td>
<td>1.06***</td>
</tr>
<tr>
<td># param.</td>
<td>4</td>
<td>8</td>
<td>30</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td># respond.</td>
<td>162</td>
<td>162</td>
<td>162</td>
<td>162</td>
<td>162</td>
</tr>
<tr>
<td># obs.</td>
<td>735</td>
<td>735</td>
<td>735</td>
<td>735</td>
<td>735</td>
</tr>
<tr>
<td># draws</td>
<td>–</td>
<td>–</td>
<td>2000</td>
<td>2000</td>
<td>2000</td>
</tr>
<tr>
<td>$LL_{final}^*$</td>
<td>–7029</td>
<td>–6911</td>
<td>–6621</td>
<td>–6047</td>
<td>–6039</td>
</tr>
<tr>
<td>AICc</td>
<td>14066</td>
<td>13840</td>
<td>13154</td>
<td>12104</td>
<td>12097</td>
</tr>
</tbody>
</table>

Robust standard errors: *** : $p < 0.01$, ** : $p < 0.05$, * : $p < 0.1$

Note: LV model coefficients not reported in the table.
Results: Distribution of cost elasticities

[Box plot showing the distribution of cost elasticities for different categories: REP, REPS, MEP, MEPS, LVREP. The plot indicates the posterior distribution of cost elasticities in percentage.]

-2.5  -2  -1.5  -1  -0.5  0
Results: Distance dependency

![Graph showing cost elasticity vs. distance for different categories: REP, REPS, LVREP, MEP, MEPS. The graph demonstrates a clear negative trend as distance increases.]

- Cost elasticity [%] vs. Distance [km]
- Categories: REP, REPS, LVREP, MEP, MEPS
- The graph shows a consistent decrease in cost elasticity with increasing distance for all categories.
Conclusions

- Median elasticity: If MIV travel costs increase by 1%, distance decreases by $\approx 0.3$ to 0.4% (re-weighted by MZMV distances)
- Remaining issues: Potential endogeneity of $dist_{n,0}$
- Strong, positive distance dependency
- Relatively high elasticities compared to related literature; usually between $-0.1$ (SR) and $-0.4$ (LR)
  - Sampling bias / low sample size
  - Survey design (daily travel, activity-based approach, etc.)
  - Very high variation in travel cost
- Respondents with pro-environmental traits travel less and show a stronger adaptation behavior
Questions?