

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

**Author(s):**
Schmid, Basil

**Publication Date:**
2017-09-14

**Permanent Link:**
https://doi.org/10.3929/ethz-b-000263468 →

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

Basil Schmid

IVT
ETH Zurich

Institut für Verkehrsplanung und Transportsysteme
Institute for Transport Planning and Systems

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

$6^{th}$ 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 $R_{tc,n}$

| Mode | Sc. 1 [in CHF] | Sc. 2 [in CHF] | Sc. 3 [in CHF] | Sc. 4 [in CHF] |
|---|---|---|---|---|
| Car | $R_{tc,n} \cdot 1.5 + 0.4$ | $R_{tc,n} \cdot 2 + 0.8$ | $R_{tc,n} \cdot 4 + 1.4$ | $R_{tc,n} \cdot 8 + 2$ |
| Moto | $R_{tc,n} \cdot 1.5 + 0.2$ | $R_{tc,n} \cdot 12 + 0.4$ | $R_{tc,n} \cdot 4 + 0.7$ | $R_{tc,n} * 8 + 1$ |
| PT | $R_{tc,n} \cdot 1.1$ | $R_{tc,n} \cdot 1.2$ | $R_{tc,n} \cdot 1.3$ | $R_{tc,n} \cdot 1.5$ |
| CS | $R_{tc,n} \cdot 1.1$ | $R_{tc,n} \cdot 1.2$ | $R_{tc,n} \cdot 1.3$ | $R_{tc,n} \cdot 1.5$ |
| CP | $R_{tc,n} \cdot 1.5$ | $R_{tc,n} \cdot 2$ | $R_{tc,n} \cdot 4$ | $R_{tc,n} \cdot 8$ |

- 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 naechsten Carsharing Fahrzeug: 3min**

**Zeit zum naechsten Carpooling Fahrzeug: 3min**

| | | | | | |
|---|---|---|---|---|---|
| Aktivitaet: | Zu Hause | Einkauf lfr. Bedar | Arbeit/Ausbildun | Dienstlich | Zu Hause |
| Ort der Aktivitaet: | Zu Hause | Tomac3 | Arbeit/Ausbildun | Dienstlich5 | Zu Hause |
| Strasse: | Nordstrasse 21 | Sihlfeldstrasse 53 | Seebahnstrasse 8 | Plantaweg 21 | Nordstrasse 21 |
| Stadt: | Zuerich | Zuerich | Zuerich | Chur | Zuerich |
| Ankunftszeit: | 00:00 | 08:17 | 08:24 | 11:31 | 14:34 |
| Laenge der Aktivitaet: | 08:05 | 00:05 | 01:55 | 01:40 | 00:44 |
| Abfahrtszeit: | 08:05 | 08:22 | 10:19 | 13:11 | 15:18 |
| Zu Fuss | ○ | ○ | ○ | ○ | ○ |
| Auto(Fahrer) | ○ | ○ | ● | ● | ○ |
| Auto(Mitfahrer) | ○ | ○ | ○ | ○ | ○ |
| Velo | ● | ● | ○ | ○ | ○ |
| OEV | ○ | ○ | ○ | ○ | ● |
| Carpooling(Mitfahrer) | ○ | ○ | ○ | ○ | ○ |
| Carsharing | ○ | ○ | ○ | ○ | ○ |
| Motorrad | ○ | ○ | ○ | ○ | ○ |
| Zurueckgelegte Distanz: | 2.78 | 0.88 | 134.19 | 134.10 | 2.43 |
| Reisezeit: | 00:12 | 00:02 | 01:12 | 01:23 | 00:13 |
| Reisekosten | 0.00 | 0.00 | 36.23 | 36.21 | 2.20 |
| | Entfernen | Entfernen | Entfernen | Entfernen | Entfernen |

**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

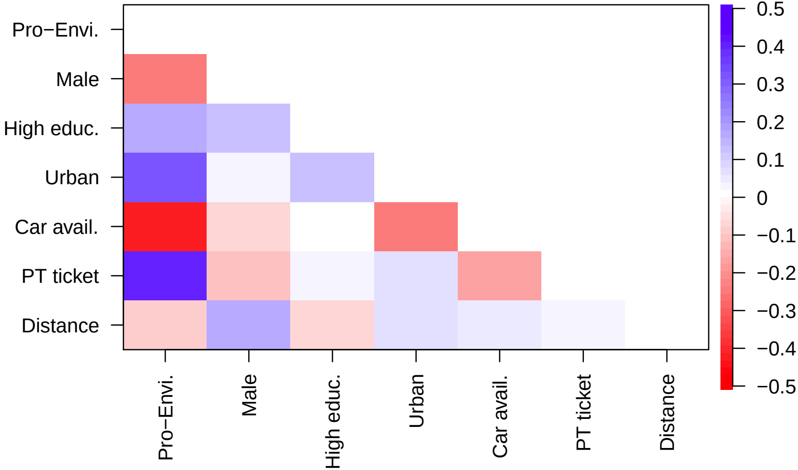
**envi5:** Zurich without cars is inconceivable

**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

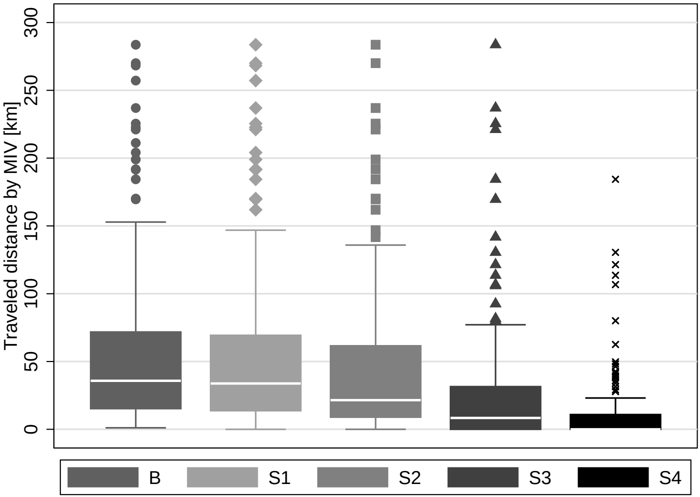
# … and socio-demographic characteristics



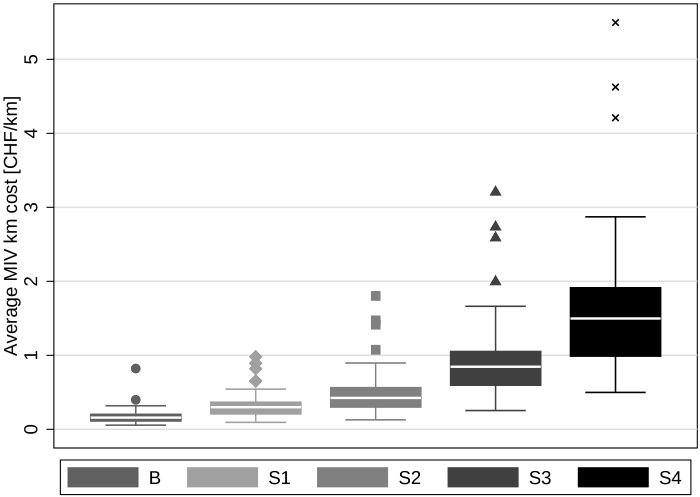
# Data

- N = 162 respondents, 810 initial choice scenarios
- Dependent variable: Distance traveled by MIV
  $y_{n,t} \equiv km_{n,t}$ after adaptation in **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 actual choice observations)
- Main explanatory variable: Average MIV travel cost per km
  $x_{n,t} \equiv \log(CHF_{n,t-1})$ after adaptation in **previous** scenario

# Adaptation patterns in distance traveled



# Change in MIV travel cost



# 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.: $E[Y_{n,t}|X_{n,t}, z_n] = \mathrm{Var}[Y_{n,t}|X_{n,t}, z_n] = \lambda_{n,t}$
  - 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 $\text{Var}[Y_{n,t}|X_{n,t}] = \lambda_{n,t} + \theta\lambda^2_{n,t}$
- Huber/White sandwich estimator for SEs (Arellano, 1987)

# Modeling framework: Log-linear index

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

$$\lambda_{2,n,t} = \epsilon_n \cdot \exp\Bigg(\alpha + \alpha_{INC} \cdot inc_n + \alpha_{ENVI} \cdot envi_n + (\beta_{COST} + \beta_{INC} \cdot inc_n + \beta_{ENVI} \cdot envi_n) \cdot \log(CHF_{n,t-1}) \cdot \left(\frac{dist_{n,0}}{\overline{dist}}\right)^{\omega_{DIST}}\Bigg)$$

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

$$\lambda_{4,n,t} = \epsilon_n \cdot \exp\Bigg(\alpha + \alpha_{INC} \cdot inc_n + \alpha_{ENVI} \cdot envi_n - \exp(\beta_{COST} + \beta_{INC} \cdot inc_n + \beta_{ENVI} \cdot envi_n + \psi_n) \cdot \log(CHF_{n,t-1}) \cdot \left(\frac{dist_{n,0}}{\overline{dist}}\right)^{\omega_{DIST}}\Bigg)$$

# 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 $\widetilde{\lambda_{s,n,t}} \equiv \lambda_{s,n,t}/\epsilon_n$, the likelihood of observing the sequence $Y_{n,t}$ given $X_{n,t}$ and $z_n$ of respondent $n$ is given by

$$\mathcal{L}\mathcal{L}_n(Y_{n,t}|X_{n,t}, z_n, \Lambda) = \log \Gamma\left(1/\theta + \sum_{t=1}^{T_n} y_{n,t}\right) - \sum_{t=1}^{T_n} \log \Gamma\left(1 + y_{n,t}\right) -$$

$$\log \Gamma(1/\theta) + 1/\theta \cdot \log(u_n) + \log(1 - u_n) \sum_{t=1}^{T_n} y_{n,t} +$$

$$\sum_{t=1}^{T_n} y_{n,t} \cdot \log\left(\widetilde{\lambda_{s,n,t}}\right) - \left(\sum_{t=1}^{T_n} y_{n,t}\right) \log\left(\sum_{t=1}^{T_n} \widetilde{\lambda_{s,n,t}}\right)$$

# 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 $LV_n$

$$\mathcal{L}_n^*(Y_{n,t}, I_{w,n}|X_{n,t}, z_n, \Omega) = \int_{\psi_n, LV_n} \exp\left(\mathcal{LL}_n(Y_{n,t}|X_{n,t}, z_n, \Lambda, \psi_n)\right) u(I_{w,n}|LV_n, \tau_{I_w}, \sigma_{I_w}) \times h(\psi_n|R)\, g(LV_n|z_n, \rho_z, \eta_{LV_z})\, d\psi_n\, dLV_n$$

$$\widetilde{\mathcal{L}_n^*}(Y_{n,t}, I_{w,n}|X_{n,t}, z_n, \Omega) = \frac{1}{R}\sum_{r=1}^{R} \exp\left(\mathcal{LL}_n(Y_{n,t}|X_{n,t}, z_n, \Lambda, \psi_n)\right) u(I_{w,n}|LV_n, \tau_{I_w}, \sigma_{I_w})$$

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

$\rightarrow$ Posterior analysis of cost elasticity

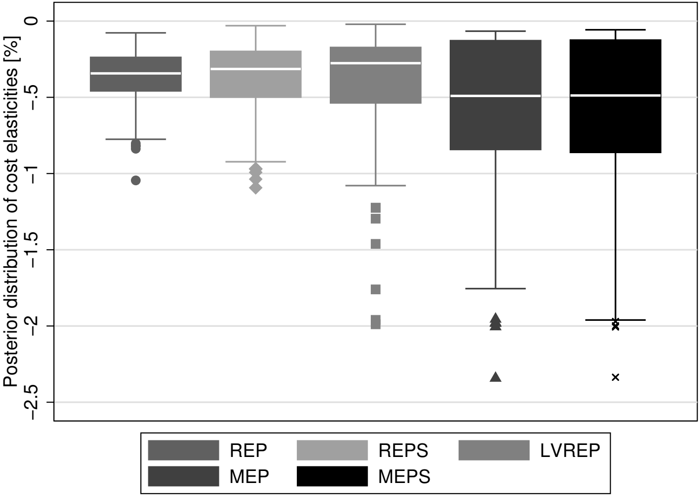
# Estimation results

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

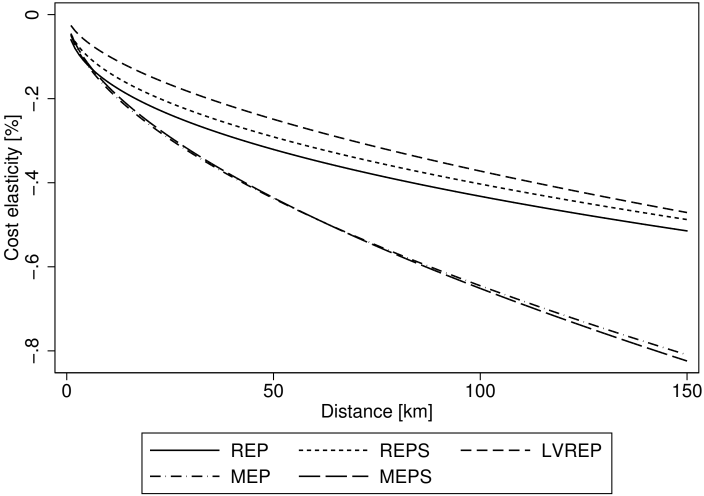
Robust standard errors: $^{***}: p < 0.01$, $^{**}: p < 0.05$, $^{*}: p < 0.1$

Note: LV model coefficients not reported in the table.

# Results: Distribution of cost elasticities



# Results: Distance dependency



# 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?