In-Store or Online Shopping? A Hybrid Choice Approach

Basil Schmid

IVT
ETH Zurich

April 21, 2016

Institut für Verkehrsplanung und Transportsysteme
Institute for Transport Planning and Systems

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Shopping in a Post-Car World

- Main hypothesis:
  - Choice and taste heterogeneity strongly determined by attitudes towards online shopping
- How sensitive are individuals towards different attributes related to their choice btw. online vs. in-store shopping?
- How do income and attitudes affect attribute sensitivities?
- What is the distribution of attitudes, and which socio-demographic characteristics are affecting them?

$\Longrightarrow$ Post-Car World: First alternative-specific Hybrid Choice model in this research field

Data (220 households; 339 participants)

Variable	Value	MZ2010 [%]	PCW15 [%]
Household income	Not reported	24.1	5.7
	$<$ 12'000 CHF	61.0	27.6
	$\geq$ 12'000 CHF	18.4	61.8
Household type	Single-person household	31.6	18.7
	Couple without kids	33.0	25.2
	Couple with kids	26.6	48.0
	Single-parent household	5.8	4.5
	Living community	3.1	3.7
Education	Low	21.0	14.7
	Medium	54.9	22.3
	High	24.1	63.0

- Post-Car World: A multi-stage travel survey
- Sample selection bias: "American Dream" households overrepresented

Experimental conditions

- Coherent choice situations:
  - home based round trip for in-store alternative
  - no social motives; buying goods is the one and only purpose
  - groceries and durable goods experiment: "Daily or weekly grocery shopping" and "multimedia, HiFi and electronic (household) appliances"
  - quality of the goods is assumed to be identical between the two shopping channels
  - in-store alternative without private cars (Post-Car World)

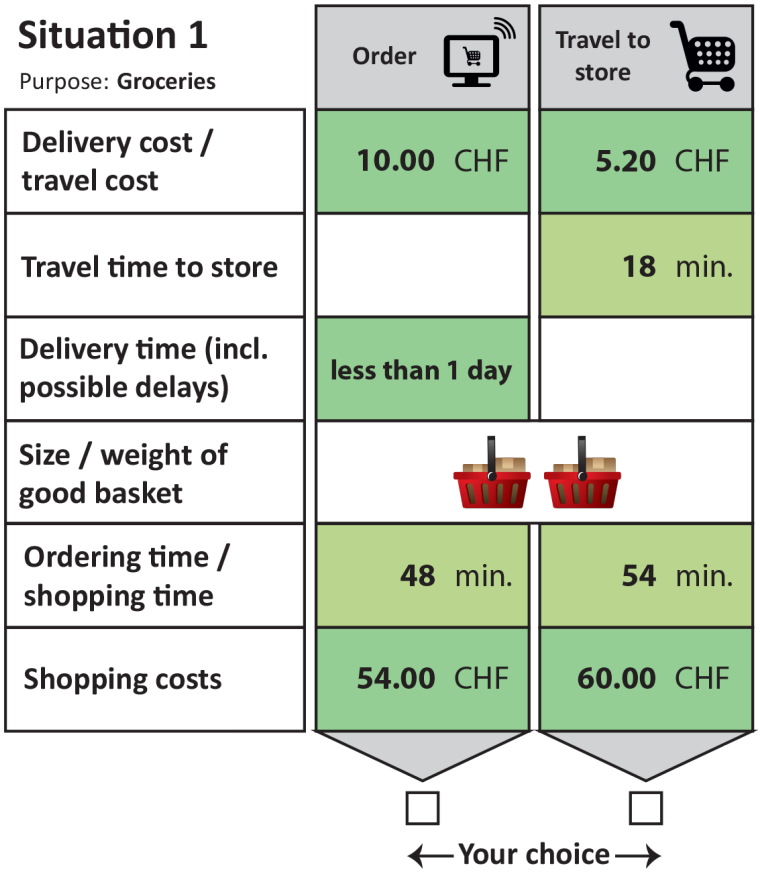
Experimental Design: Attribute levels

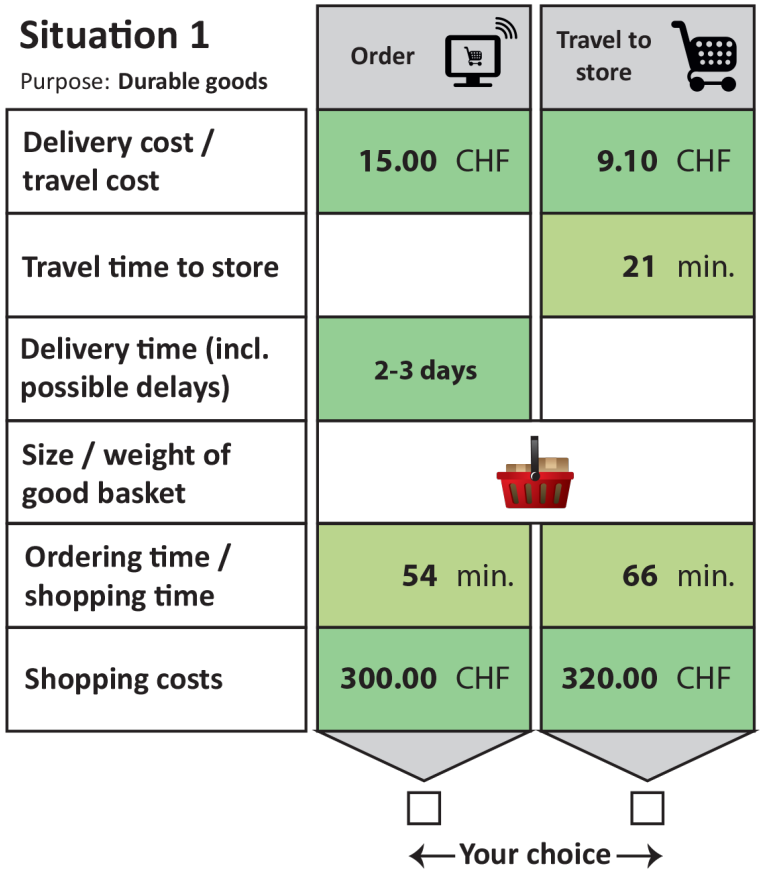
Attributes	Online	In-store	Levels	μ	σ	ν
Shopping cost [CHF]	$\surd$	–	$-10\%, -5\%, 0\%$	235.2	190.4	0.8
Shopping cost [CHF]	–	$\surd$	$-5\%, 0\%, +5\%$	248.0	200.7	0.8
Time for shop. [min]	$\surd$	–	$-20\%, -10\%, +5\%$	38.1	16.2	1.3
Time for shop. [min]	–	$\surd$	$-10\%, 0\%, +10\%$	41.8	17.9	1.4
Delivery cost and duty	$\surd$	–	0, 5, 10, 15 CHF	7.6	5.6	0.0
Travel cost [CHF]	–	$\surd$	$-20\%, +10\%, +40\%$	5.2	3.5	3.1
Delivery time groceries	$\surd$	–	$<$ 1 day, 1-2 days, $>$ 2 days	–	–	–
Delivery time durables	$\surd$	–	2-3 days, 4-7 days, $>$ 1 week	–	–	–
Travel time [min]	–	$\surd$	$-30\%, 0\%, +30\%, \geq$ 3 min	24.4	17.5	2.4
Size/weight of the good basket	$\surd$	$\surd$	Low, medium, high (same for both alternatives)	–	–	–

μ = mean, σ = standard deviation, ν = skewness; for attribute values in the choice experiment

- D-Efficient design; 3 blocks with 8 choice sets
- Participants were assigned to the "groceries" (38 %) or "durable goods" (62 %) experiment based on reported shopping trips

Example choice situations





Attitudes towards online shopping

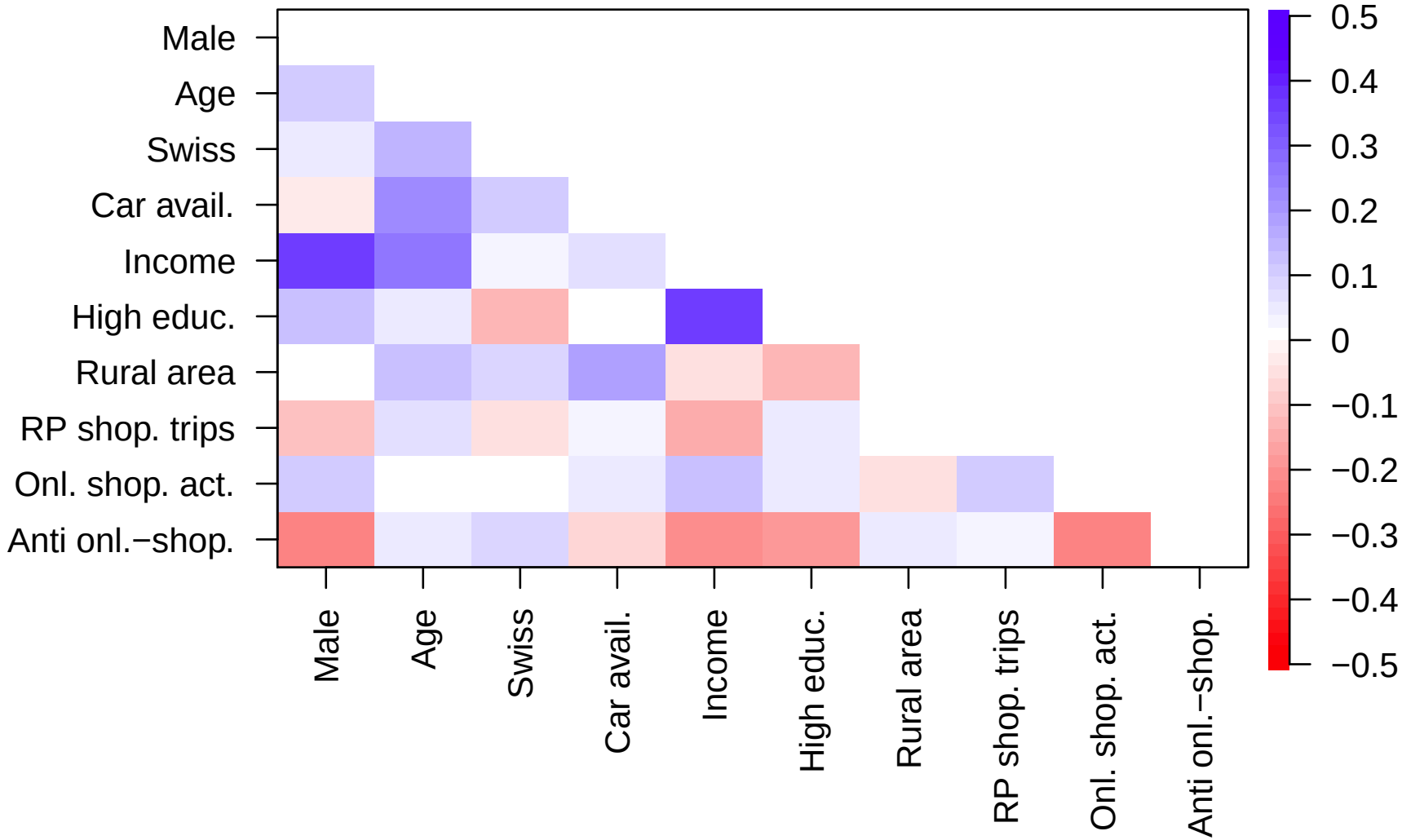
Questionnaire item	Factor loading
sh1: I often order products in the internet	–0.69
sh2: Online shopping is associated with risks	+0.48
sh3: Credit card fraud is one the reasons why I don't like online shopping	+0.69
sh4: The internet has more cons than pros	+0.54
sh5: A disadvantage of online shopping is that I cannot physically examine the products	+0.29
sh6: Online shopping facilitates the comparison of prices	–0.54
sh7: The risk of receiving a wrong product is one the main reasons why I don't like online shopping	+0.65

Estimation method: Maximum likelihood
Rotation method: Orthogonal varimax
Variance explained: 31.5 %. Cronbach's Alpha: 0.75
Kaiser-Meyer-Olkin measure of sampling adequacy: 0.80
Likelihood-ratio test: 1 factor vs. saturated: $p < 0.00$
Number of subjects: 339. Subject-to-Item ratio: 48.4

Attitudes and socio-economic characteristics

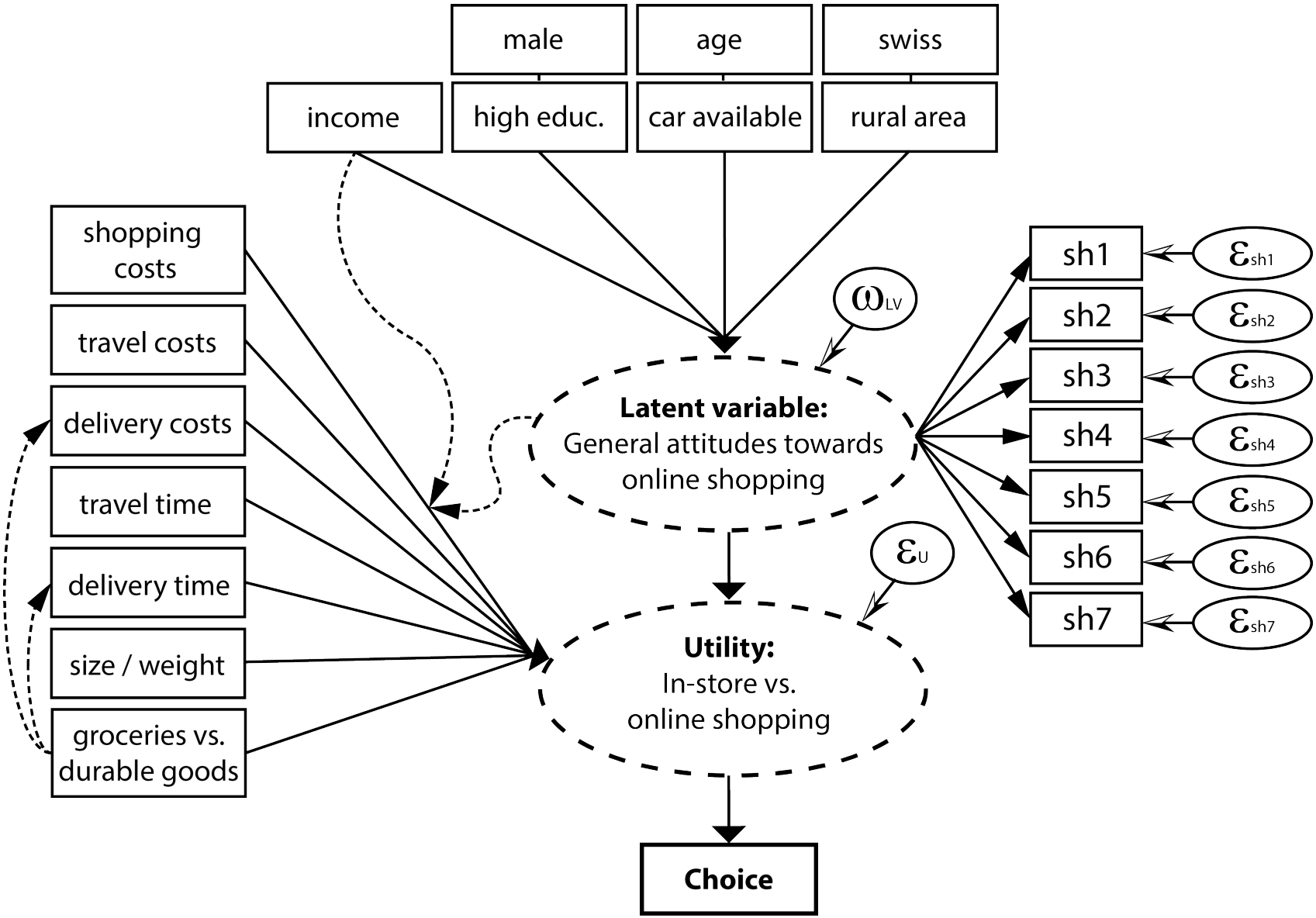
?

Attitudes and socio-economic characteristics



Attitudes and socio-economic characteristics

Modeling Framework: Hybrid choice



Modeling Framework: Structural model

Utility equation for shopping channel i with choice attributes X_{i_n} and the latent online shopping variable LV_n:

$$\begin{aligned} U_{O_n} = \beta_{C_O} + \beta_{O_n} \cdot X_{O_n} + \beta_{sc,O} * sc_O * \left(\frac{inc}{\overline{inc}}\right)^{\lambda_{inc}} + \\ \mu_{LV} \cdot (LV_n - \overline{LV_n}) + \\ \mu_{sc,LV} \cdot sc_{O_n} \cdot (LV_n - \overline{LV_n}) + \epsilon_{O_n} \end{aligned} \tag{1}$$

$$\begin{aligned} U_{IS_n} = \beta_{IS} \cdot X_{IS_n} + \beta_{sc,IS} * sc_{IS} * \left(\frac{inc}{\overline{inc}}\right)^{\lambda_{inc}} + \\ \mu_{sc,LV} \cdot sc_{IS_n} \cdot (LV_n - \overline{LV_n}) + \epsilon_{IS_n} \end{aligned} \tag{2}$$

Latent variable equation with socio-economic characteristics X_n:

$$LV_n = \overline{LV_n} + \kappa_X \cdot X_n + \omega_{LV_n} \tag{3}$$

Modeling Framework: Structural model

Relative importance of choice attribute X_{i_n} compared to shopping costs as a function of income and the latent variable LV_n:

$$f(inc_n, LV_n) = \frac{\beta_{X_{i_n}}}{\beta_{sc} \cdot \left(\frac{inc_n}{\overline{inc_n}}\right)^{\lambda_{inc}} + \mu_{sc,LV} \cdot (LV_n - \overline{LV_n})} \tag{4}$$

- If $\lambda_{inc} < 0$ and $\mu_{cost,LV} > 0$: Shopping cost sensitivity increases with lower income and a more positive attitude towards online shopping
- For the "average" respondent, the equation collapses to

$$f(\overline{inc}, \overline{LV_n}) = \frac{\beta_{X_{i_n}}}{\beta_{sc}} \tag{5}$$

Modeling Framework: Measurement model

Latent variable measurement equations with responses to the 7 online shopping items I_{sh}:

$$I_{sh_n} = \overline{I_{sh}} + \tau_{LV_{I_{sh}}} \cdot LV_n + \epsilon_{I_{sh_n}} \tag{6}$$

Choice equation: Choice of individual n for shopping channel i by maximizing utility U_i:

$$if \quad U_{O,n} > U_{IS,n} : choice_{i,n} = \begin{cases} \text{Online shopping} \\ else \text{ In-store shopping} \end{cases} \tag{7}$$

β_j, μ_j, λ_{inc}, $\overline{LV_n}$, κ_j, $\sigma_{\omega_{LV}}$, $\overline{I_{sh}}$, τ_{sh} and $\sigma_{I_{sh}}$ are the parameters to be estimated (45 in total)

Estimation

Likelihood of individual n choosing alternative i is the joint probability of observing the choice and the 7 online shopping items I_{sh_n}, given choice attributes and socio-economic characteristics $X_{i,n}$:

$$Likelihood = \int_{\omega_{LV_n}} P(choice_{i,n}|X_{i,n}, \omega_{LV_n}) \prod_{sh=1}^{7} f_{sh_n}(I_{sh_n}, \omega_{LV_n}) \phi(\omega_{LV}) d\omega_{LV_n} \tag{8}$$

$$\omega_{LV} \sim N(0, \sigma_{\omega_{LV}}) \tag{9}$$

$$P(ch_{i,n}|X_{i,n}, \omega_{LV_n}) = \frac{exp(U(X_{i,n}))}{\sum_j^2 exp(U(X_{j,n}))} \tag{10}$$

$$f_{sh_n}(I_{sh_n}, \omega_{LV_n}) = \frac{1}{\sigma_{I_{sh}}} \phi \left(\frac{I_{sh_n} - \overline{I_{sh}} - \tau_{I_{sh}} \cdot LV_n}{\sigma_{I_{sh}}} \right) \tag{11}$$

Maximum likelihood estimation with PythonBiogeme version 2.4 on Euler (HPCC, 2 cores, runtime 45 min., 139 iterations)

Estimation results: Choice models

Variable	Base model	Factor model	Hybrid model
Shopping cost	-0.021^{***}	-0.024^{***}	-0.025^{***}
Income elasticity of shopping cost	0.041	-0.034	-0.054
"Anti-onl.-shop." factor/LV x shop. cost	–	0.007^{***}	0.019^{***}
Travel time (IS)	-0.022^{***}	-0.024^{***}	-0.025^{***}
Travel cost (IS)	-0.036^{**}	-0.035^{**}	-0.037^{**}
Medium delivery time (ONL)	-0.110	-0.142	-0.152
Med. delivery time x durables (ONL)	-0.182	-0.172	-0.172
High delivery time (ONL)	-0.813^{***}	-0.873^{***}	-0.894^{***}
High delivery time x durables (ONL)	0.256	0.243	0.243
Delivery cost (ONL)	-0.093^{***}	-0.099^{***}	-0.101^{***}
Delivery cost x durables (ONL)	0.057^{***}	0.055^{***}	0.055^{***}
ASC (ONL)	-1.550^{***}	-1.540^{***}	-1.570^{***}
Purpose durables (ONL)	0.529^{**}	0.425^{*}	0.448^{*}
Medium size (ONL)	1.050^{***}	1.100^{***}	1.130^{***}
Large size (ONL)	2.250^{***}	2.410^{***}	2.460^{***}
"Anti-online-shopping" factor/LV (ONL)	–	-0.466^{***}	-1.210^{***}
# estimated parameters	14	16	45
Choice observations (participants)		2698 (339)	
Log-likelihood null	-1870.1	-1870.1	-66075.5
Log-likelihood model	-1485.5	-1393.5	-23098.9
McFadden ρ^2	0.21	0.26	0.65
Iterations	16	29	139

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Estimation results: LV model

Variable	Dep. variable: LV_n
$\overline{LV_n}$	2.160^{***}
Age	-0.012^{**}
$\text{Age}^2/100$	0.019^{***}
Car availability	-0.125^{***}
High education	-0.111^{***}
Income	-0.084^{***}
Rural	0.119^{***}
Male	-0.247^{***}
Swiss	0.106^{***}
$\sigma_{\omega_{LV}}$	0.469^{***}

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

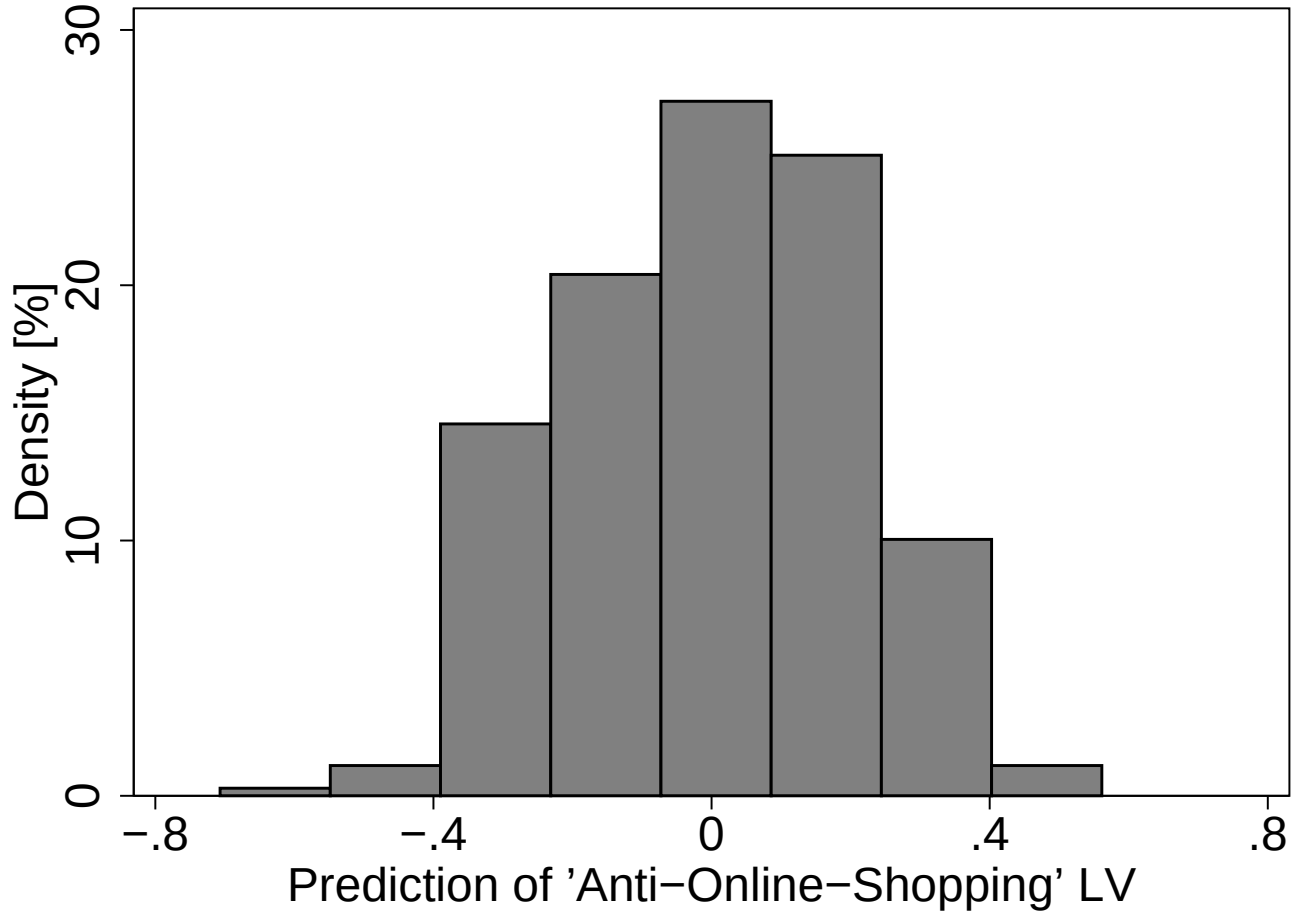
- Female and Swiss non-car users with low education and income living in rural residential locations have the most negative attitudes towards online shopping
- Maximal pro-online shopping attitudes with 32 years of age

Value of time for shopping trips/delivery

Coefficient ratios	Base model	Factor model	Hybrid model
VTTS shopping trips (travel cost) [CHF/h]	36.46	41.02	39.89
VTTS shopping trips (shop. cost) [CHF/h]	62.86	61.01	60.00
VDTS medium delivery time groceries* [CHF/t.u.]	1.19	1.43	1.50
VDTS high delivery time groceries [CHF/t.u.]	8.76	8.80	8.85
VDTS medium delivery time durables [CHF/t.u.]	8.09	7.12	7.01
VDTS high delivery time durables [CHF/t.u.]	15.43	14.29	14.12
VDTS groceries [CHF/day]	6.1	6.2	6.2
VDTS durable goods [CHF/day]	2.6	2.4	2.4
Travel cost / shopping cost [-]	1.72	1.49	1.50
Delivery cost groceries / shopping cost [-]	4.42	4.17	4.11
Delivery cost durables / shopping cost [-]	1.72	1.85	1.87

- Current study: Value of travel time savings (VTTS) of 40 CHF/h; about 50 % higher if considering shopping instead of travel costs as reference (values in brackets)
- VDTS: For easier interpretation, linear interpolation of delivery time

Prediction of latent variable



- $\widetilde{LV_n} = LV_n - \overline{LV_n}$ is approximately normally distributed with mean 0

Prediction of latent variable

Total market shares	Online 51%	In-store 49%
Shopping cost	−2.83 (2.96)	−3.12 (2.98)
Shopping cost (max. $\widehat{LV_n}$)	−1.61 (1.69)	−1.78 (1.70)
Shopping cost (min. $\widehat{LV_n}$)	−4.37 (4.57)	−4.82 (4.61)
Travel time	–	−0.31 (0.29)
Travel cost	–	−0.10 (0.09)
Delivery cost groceries	−0.37 (0.39)	–
Delivery cost durables	−0.17 (0.18)	–

*: Not significant at the 5% level.

- Attribute sensitivities relative to shopping costs are increasing for higher anti-online shopping attitudes $\Longrightarrow$ price-insensitive trade-off behavior; in-store dominant channel

Conclusions

- Behavioral richness and estimation efficiency increase substantially when including latent variables
- Structural model reveals distribution of LV in the population based on fundamental socio-demographic characteristics
- VTTS vs. VDTS: Large potential of online shopping given the relatively high value of travel time savings for shopping trips
- Pro-online shopping attitudes lead to a sign. increase in shopping cost sensitivity $\Longrightarrow$ larger choice set when considering both online and in-store shopping as possible shopping channels
- 1 CHF $\neq$ 1 CHF: Delivery costs are perceived as more negative than travel and shopping costs (avoidability hypothesis) $\Longrightarrow$ online retailers better incorporate delivery costs in shopping prices

Problems

- Panel structure: Was not able to account for it when working on Euler (same code, same data, but error). bioDraws('...') , bioNormalDraws('...','id')
- Other distributions?
- $\widetilde{LV_n} = LV_n - \overline{LV_n}$: Evaluation of interaction effects at their means? Problem: $\widetilde{LV_n} \neq 0$
- How to account for scale heterogeneity (SMNL) in PythonBiogeme?
- Including more latent variables (risk aversion, love of variety, etc.): What's the right estimation approach in PythonBiogeme? (eirass_panel.py)

Questions?

Project website:

http://postcarworld.epfl.ch/