ENHANCED MEASUREMENT EQUATIONS FOR LATENT CLASS CHOICE MODELS

Aurélie Glerum
Ricardo Hurtubia
My Hang Nguyen
Michel Bierlaire

Eighth Workshop on Discrete Choice Models

Friday 31st May 2013

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

OUTLINE

Introduction & motivation

Literature

Methodology

Mode choice case study

Conclusion

INTRODUCTION & MOTIVATION

Recent developments in demand modeling for transportation

- Hybrid choice model (HCM) framework (Walker, 2001; Ben-Akiva et al., 2002)
  Comprehensive framework that allows to incorporate unobservable factors as explanatory variables of choice.

Discrete choice model (DCM)	+	Latent variable model (LVM) or Latent class model (LCM)

- Choice of transportation mode, car, etc.
  - Influenced by economic factors:
    - Price
    - Trip duration
    - Etc.
  - Often also involve more subjective factors:
    - Attitudes
    - Perceptions
    - Lifestyles
    - Habits
- HCM framework incorporates these subjective factors.

INTRODUCTION & MOTIVATION

Important issues in the use of HCMs:

- Well-established that soft features have an important effect on choice.
- The best way to integrate them in choice models is not obvious.
- Many applications using psychometrics as measures of attitudes in the integrated choice and latent variable model framework, fewer for latent class choice models.

Aim of this research:

- Enhance latent class choice models with psychometric indicators

LITERATURE

Integration of psychometrics in latent class choice models:

- Measurement indicators usually used for LVM, but relatively few applications using them in LCM.
  - Specification of the class-membership model of a LC by means of an intermediate LV (Gopinath, 1995).
  - Estimation of class-specific probabilities to respond to each indicator (Collins and Lanza, 2010; Atasoy et al., forthcoming).

LITERATURE

Integration of psychometrics in latent class choice models:

- Measurement indicators usually used for LVM, but relatively few applications using them in LCM.

    - Specification of the class-membership model of a LC by means of an intermediate LV (Gopinath, 1995).

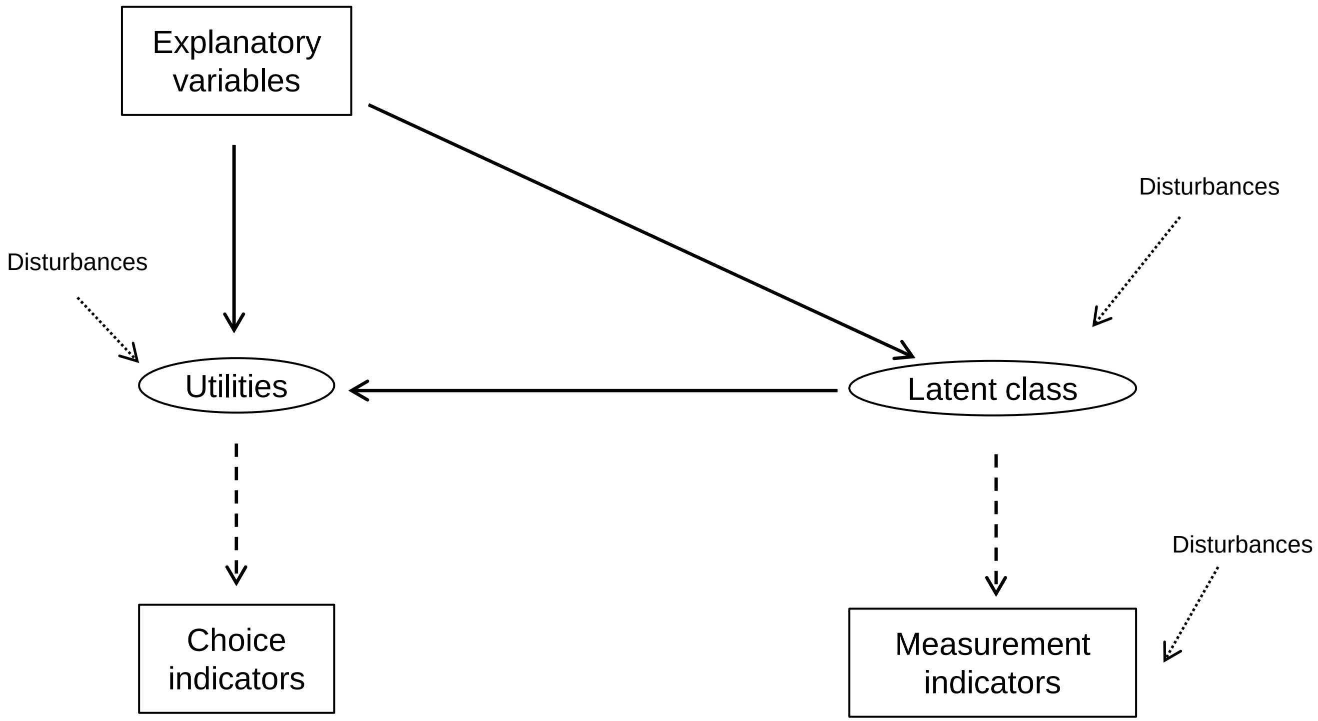
      AIM ⇨ Simpler characterisation where direct link between indicators and latent classes.

    - Estimation of class-specific probabilities to respond to each indicator (Collins and Lanza, 2010; Atasoy et al., forthcoming).

      AIM ⇨ Add structural information in the measurement equation

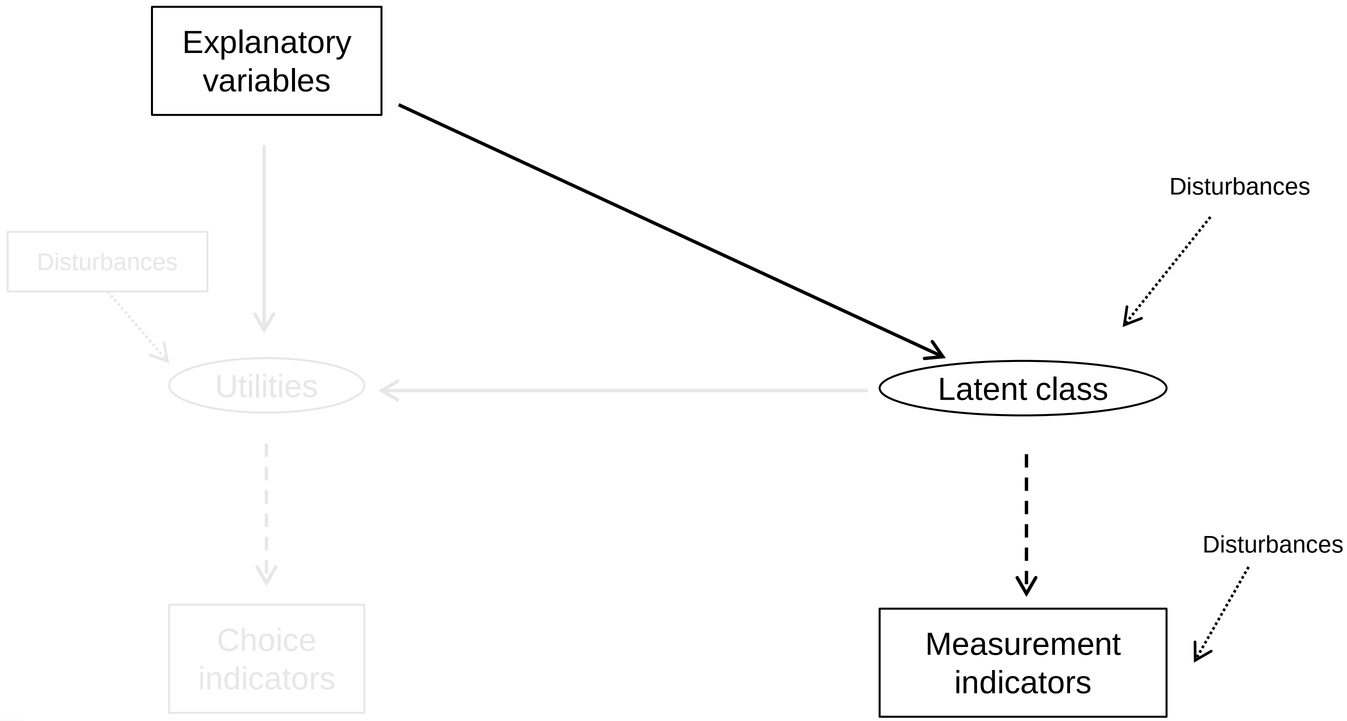
METHODOLOGY

Integration of psychometrics in latent class choice models:



Integration of psychometrics in latent class choice models:

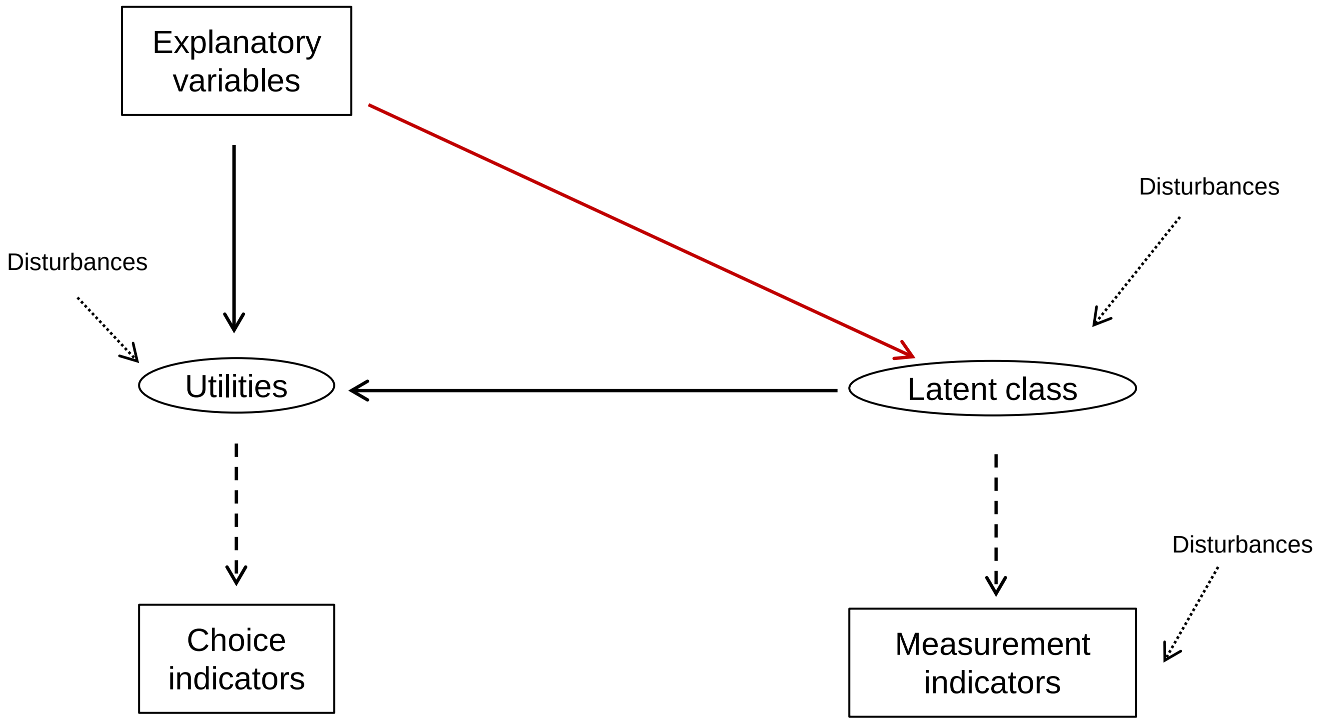
- Structural equation model (SEM) framework used to characterize latent construct and relate it to its measurement indicators
  (e.g. Bollen, 1989; Hancock and Mueller, 2006; Bartholomew et al., 2011).



METHODOLOGY

Integration of psychometrics in latent class choice models:

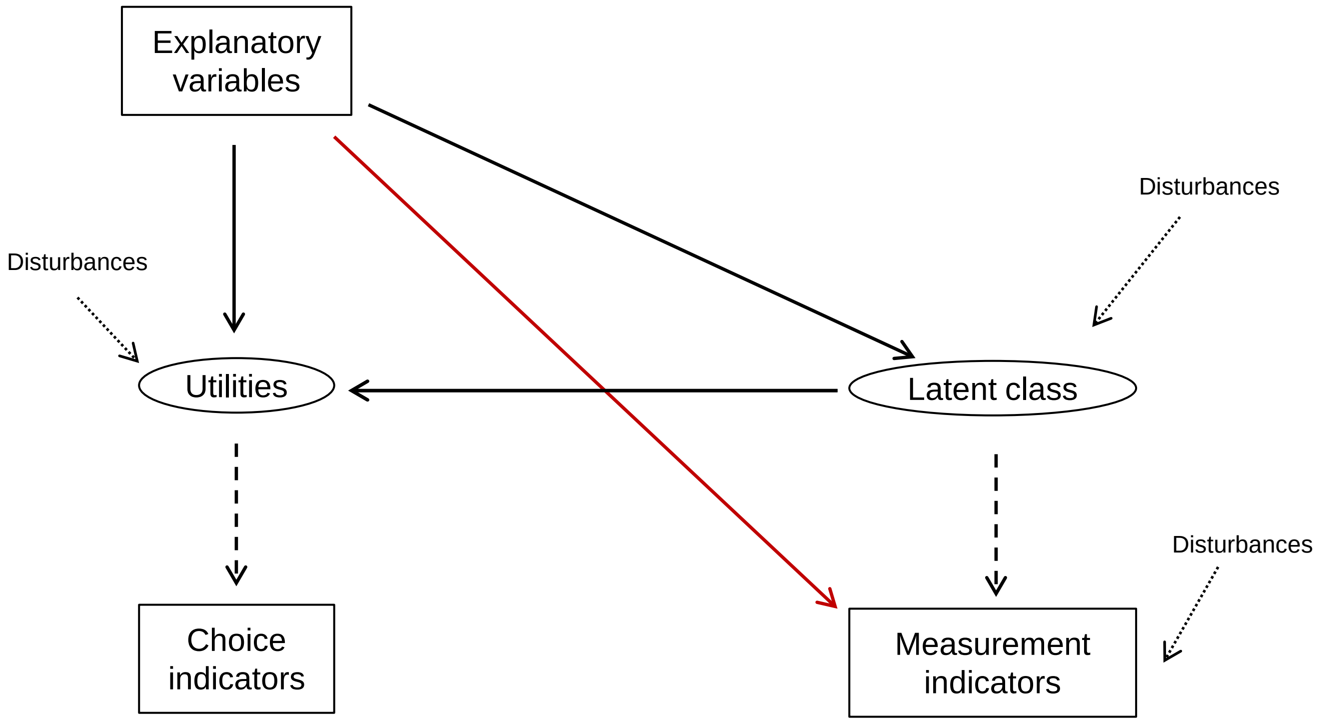
- Heterogeneity of membership to latent classes captured among population
- But also need to capture heterogeneity in reporting indicators of latent class



METHODOLOGY

Integration of psychometrics in latent class choice models:

- Heterogeneity of membership to latent classes captured among population
- But also need to capture heterogeneity in reporting indicators of latent class



METHODOLOGY

Specification of integrated choice and latent class model

Choice model:

$$U_{in}^s = V(X_{in}, X_n, \beta^s) + \varepsilon_{in}^s, \qquad \text{with } \varepsilon_{in}^s \sim EV(0,1)$$

Latent class model:

Class-membership function:

$$F_{ns} = f(X_n, \gamma^s) + \xi_{ns}, \qquad \text{where } \xi_{ns} \sim EV(0,1)$$

Measurement indicators:

$$\tilde{G}_{I_k,n}^s = g(X_n; \alpha_k^s) + \nu_{kn}^s,$$

$$\text{with } \nu_{kn}^s \sim \text{Logistic}(0,1)$$

$$G_{I_k,n}^s = \begin{cases} 1 & \text{if } -\infty < \tilde{G}_{I_k,n}^s \leq \tau_{1,k}^s \\ 2 & \text{if } \tau_{1,k}^s < \tilde{G}_{I_k,n}^s \leq \tau_{2,k}^s \\ 3 & \text{if } \tau_{2,k}^s < \tilde{G}_{I_k,n}^s \leq \tau_{3,k}^s \\ 4 & \text{if } \tau_{3,k}^s < \tilde{G}_{I_k,n}^s \leq \tau_{4,k}^s \\ 5 & \text{if } \tau_{4,k}^s < \tilde{G}_{I_k,n}^s \leq +\infty \end{cases}$$

METHODOLOGY

Specification of integrated choice and latent class model

Choice model:

$$U_{in}^s = V(X_{in}, X_n, \beta^s) + \varepsilon_{in}^s, \qquad \text{with } \varepsilon_{in}^s \sim EV(0,1)$$

Latent class model:

Class-membership function:

$$F_{ns} = f(X_n, \gamma^s) + \xi_{ns}, \qquad \text{where } \xi_{ns} \sim EV(0,1)$$

Measurement indicators:

$$\tilde{G}_{I_k,n}^s = g(X_n, \alpha_k^s) + \nu_{kn}^s,$$

$$\text{with } \nu_{kn}^s \sim \text{Logistic}(0,1)$$

Class-specific parameters

$$G_{I_k,n}^s = \begin{cases} 1 & \text{if } -\infty < \tilde{G}_{I_k,n}^s \le \tau_{1,k}^s \\ 2 & \text{if } \tau_{1,k}^s < \tilde{G}_{I_k,n}^s \le \tau_{2,k}^s \\ 3 & \text{if } \tau_{2,k}^s < \tilde{G}_{I_k,n}^s \le \tau_{3,k}^s \\ 4 & \text{if } \tau_{3,k}^s < \tilde{G}_{I_k,n}^s \le \tau_{4,k}^s \\ 5 & \text{if } \tau_{4,k}^s < \tilde{G}_{I_k,n}^s \le +\infty \end{cases}$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

METHODOLOGY

Specification of integrated choice and latent class model

Choice model:

$$U_{in}^{s} = V(X_{in}, X_n, \beta^{s}) + \varepsilon_{in}^{s}, \qquad \text{with } \varepsilon_{in}^{s} \sim EV(0,1)$$

Latent class model:

Class-membership function:

$$F_{ns} = f(X_n, \gamma^{s}) + \xi_{ns}, \qquad \text{where } \xi_{ns} \sim EV(0,1)$$

Measurement indicators:

$$\tilde{G}_{I_k,n}^{s} = g(X_n; \alpha_k^{s}) + \nu_{kn}^{s},$$

$$\text{with } \nu_{kn}^{s} \sim \text{Logistic}(0,1)$$

Socio-economic information of the individual

$$G_{I_k,n}^{s} = \begin{cases} 1 & \text{if } -\infty < \tilde{G}_{I_k,n}^{s} \leq \tau_{1,k}^{s} \\ 2 & \text{if } \tau_{1,k}^{s} < \tilde{G}_{I_k,n}^{s} \leq \tau_{2,k}^{s} \\ 3 & \text{if } \tau_{2,k}^{s} < \tilde{G}_{I_k,n}^{s} \leq \tau_{3,k}^{s} \\ 4 & \text{if } \tau_{3,k}^{s} < \tilde{G}_{I_k,n}^{s} \leq \tau_{4,k}^{s} \\ 5 & \text{if } \tau_{4,k}^{s} < \tilde{G}_{I_k,n}^{s} \leq +\infty \end{cases}$$

Specification of integrated choice and latent class model

Likelihood function given by: $L = \prod_{n=1}^{N} f(y_{in}, I_n \mid X_{in}; \alpha, \beta, \lambda)$ with

$$f(y_{in}, I_n | X_{in}, s; \alpha, \beta, \lambda) = \sum_s \left\{ P_n(y_{in} | X_{in}, s; \beta) \prod_k P(I_{kn} | X_{in}, s; \alpha_s) \right\} P_n(s | X_n, s; \lambda)$$

$$y_{in} = \begin{cases} 1 \text{ if } U_{in} = \max_j U_{jn} \\ 0 \text{ otherwise} \end{cases}$$

METHODOLOGY

Specification of integrated choice and latent class model

Likelihood function given by: $L = \prod_{n=1}^{N} f(y_{in}, I_n \mid X_{in}; \alpha, \beta, \lambda)$ with

$$f(y_{in}, I_n | X_{in}, s; \alpha, \beta, \lambda) = \sum_{s} \left\{ P_n(y_{in} | X_{in}, s; \beta) \prod_{k} P(I_{kn} | X_{in}, s; \alpha_s) \right\} P_n(s | X_n, s; \lambda)$$

$$y_{in} = \begin{cases} 1 \text{ if } U_{in} = \max_j U_{jn} \\ 0 \text{ otherwise} \end{cases}$$

⇨ **Analogy with integrated choice and latent variable model**

$$f(y_{in}, I_n \mid X_{in}; \alpha, \beta, \lambda) = \int_{X_n^*} P(y_{in} \mid X_{in}, X_n^*; \beta) \cdot \prod_{k} P(I_{kn} \mid X_{in}, X_n^*; \alpha) \cdot f(X_n^* \mid X_n; \lambda) dX_n^*$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Specification of integrated choice and latent class model

Likelihood function given by: $L = \prod_{n=1}^{N} f(y_{in}, I_n \mid X_{in}; \alpha, \beta, \lambda)$ with

$$f(y_{in}, I_n | X_{in}, s; \alpha, \beta, \lambda) = \sum_s \left\{ P_n(y_{in} | X_{in}, s; \beta) \prod_k P(I_{kn} | X_{in}, s; \alpha_s) \right\} P_n(s | X_n, s; \lambda)$$

$$y_{in} = \begin{cases} 1 \text{ if } U_{in} = \max_j U_{jn} \\ 0 \text{ otherwise} \end{cases}$$

⇨ **Analogy with integrated choice and latent variable model**

⇨ **Class-specific parameters in measurement equation**

METHODOLOGY

Specification of integrated choice and latent class model

Likelihood function given by: $L = \prod_{n=1}^{N} f(y_{in}, I_n \mid X_{in}; \alpha, \beta, \lambda)$ with

$$f(y_{in}, I_n | X_{in}, s; \alpha, \beta, \lambda) = \sum_{s} \left\{ P_n(y_{in} | X_{in}, s; \beta) \prod_{k} P(I_{kn} | X_{in}, s; \alpha_s) \right\} P_n(s | X_n, s; \lambda)$$

$$y_{in} = \begin{cases} 1 \text{ if } U_{in} = \max_j U_{jn} \\ 0 \text{ otherwise} \end{cases}$$

⇨ **Analogy with integrated choice and latent variable model**

⇨ **Class-specific parameters in measurement equation**

⇨ **Socio-economic variables as explanatory of response to indicator**

METHODOLOGY

Specification of integrated choice and latent class model

Likelihood function given by: $L = \prod_{n=1}^{N} f(y_{in}, I_n \mid X_{in}; \alpha, \beta, \lambda)$ with

$$f(y_{in}, I_n|X_{in}, s; \alpha, \beta, \lambda) = \sum_{s} \left\{ P_n(y_{in}|X_{in}, s; \beta) \prod_{k} P(I_{kn}|X_{in}, s; \alpha_s) \right\} P_n(s|X_n, s; \lambda)$$

$$y_{in} = \begin{cases} 1 \text{ if } U_{in} = \max_j U_{jn} \\ 0 \text{ otherwise} \end{cases}$$

⇨ **Analogy with integrated choice and latent variable model**

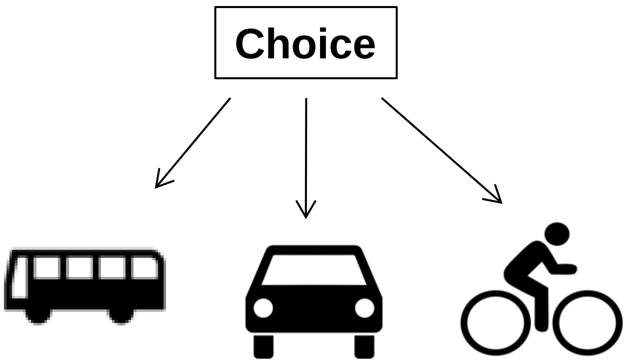
⇨ **Class-specific parameters in measurement equation**

⇨ **Socio-economic variables as explanatory of response to indicator**

MODE CHOICE CASE STUDY

Revealed preferences (RP) survey

- Transportation mode choice study
- Conducted between 2009-2010 in low-density areas of Switzerland
- Conducted with bus operator PostBus
- Info on all round trips performed by inhabitants in one day:
    - Transport mode
    - Trip duration
    - Cost of trip
    - Activity at destination
    - Etc.
- Psychometric indicators
- 1763 valid questionnaires collected



MODE CHOICE CASE STUDY

Development of latent class choice model:

Analysis of the transportation mode choices for individuals segmented according to ⇨ dependent class
⇨ independent class

Comparison of 3 models:

LCCM 1: without psychometric indicators

LCCM 2: with psychometric indicators, estimation of class-specific item-response probabilities (Atasoy et al., 2012)

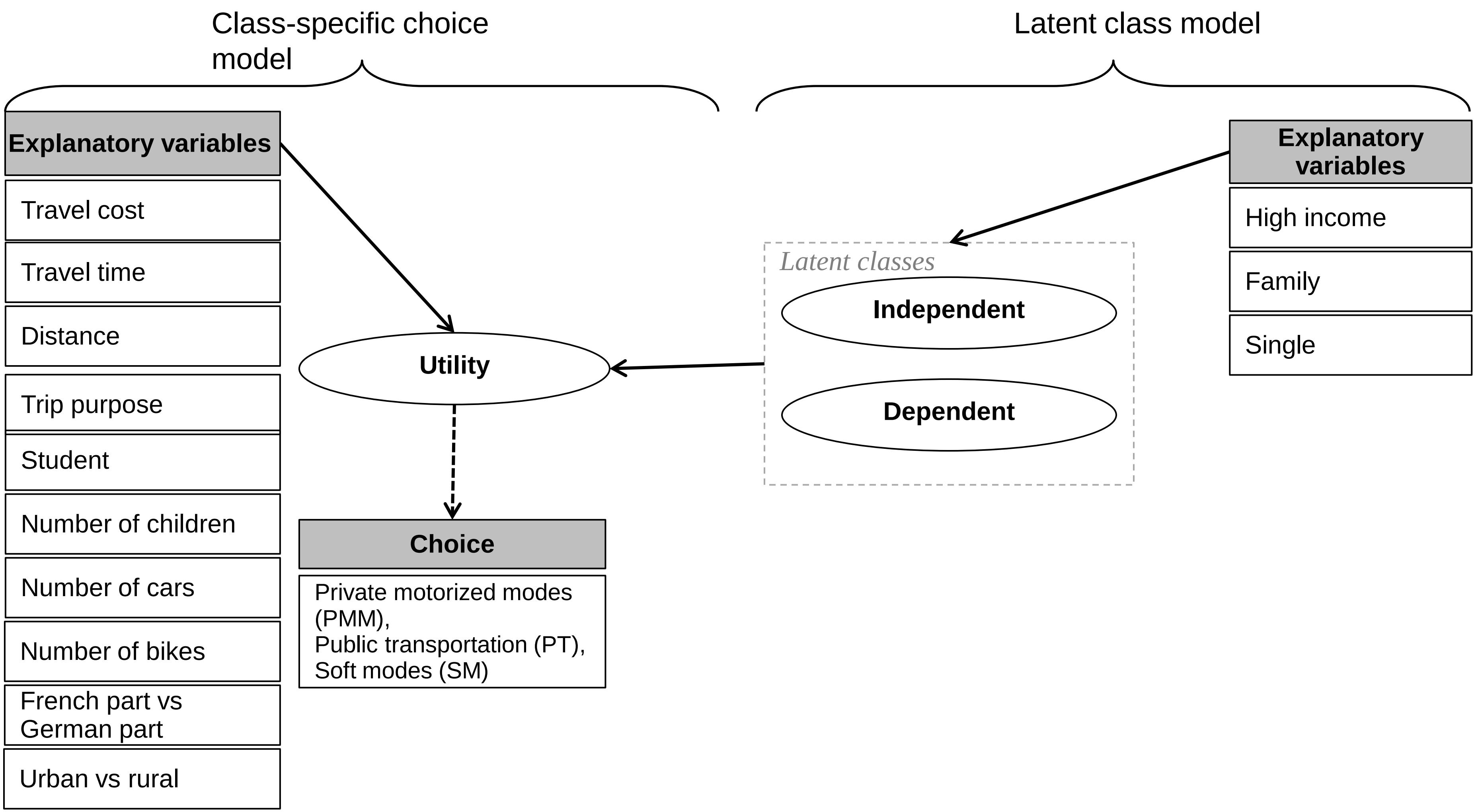
LCCM 3: with psychometric indicators, related to socio-economic characteristics of the respondent by structural measurement relationships

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

MODE CHOICE CASE STUDY

LCCM 1



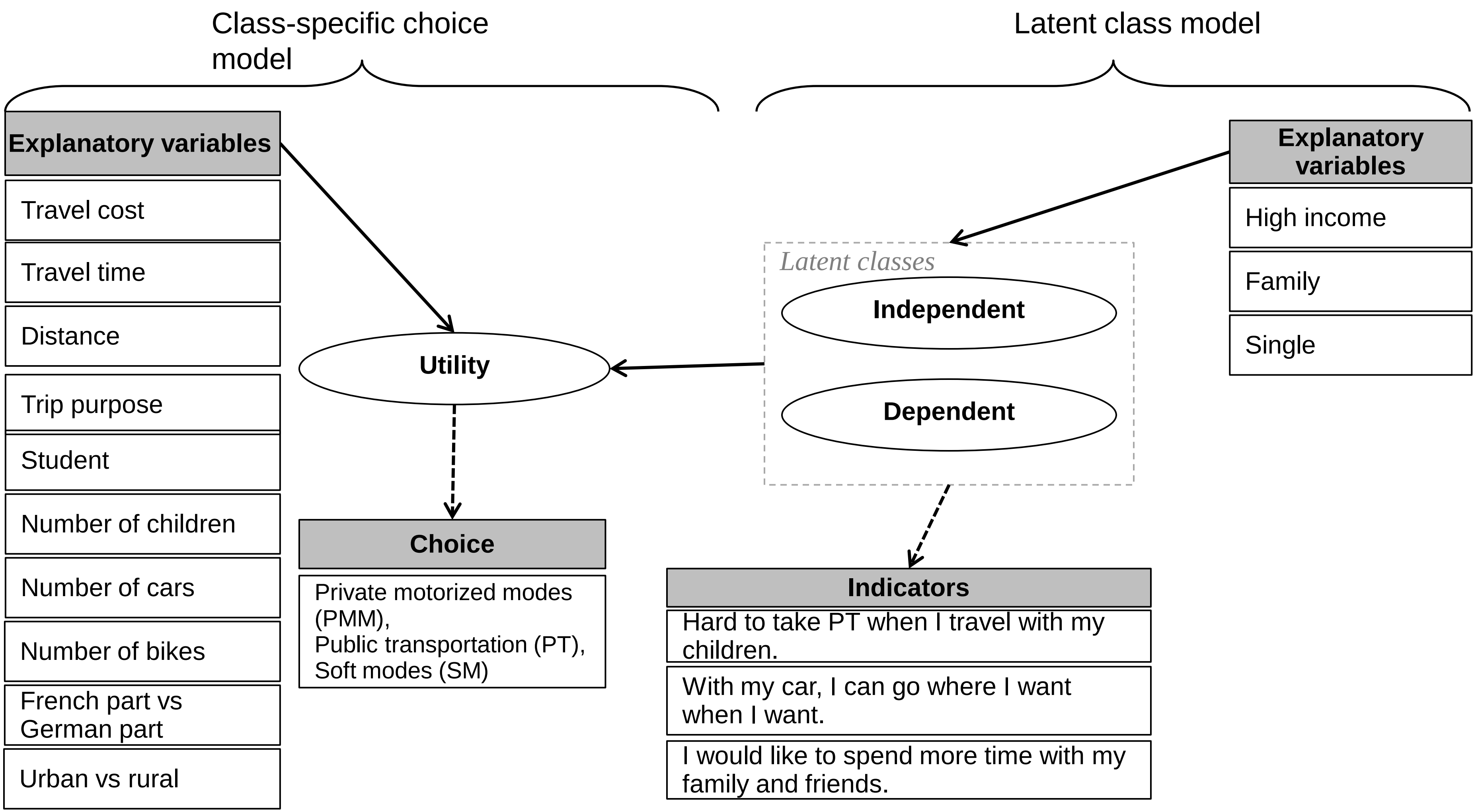
$$\mathscr{L}_{\text{LCM1}} = \prod_n \{P_n(i|\text{class } 1) \cdot P_n(\text{class } 1) + P_n(i|\text{class } 2) \cdot P_n(\text{class } 2)\}$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

MODE CHOICE CASE STUDY

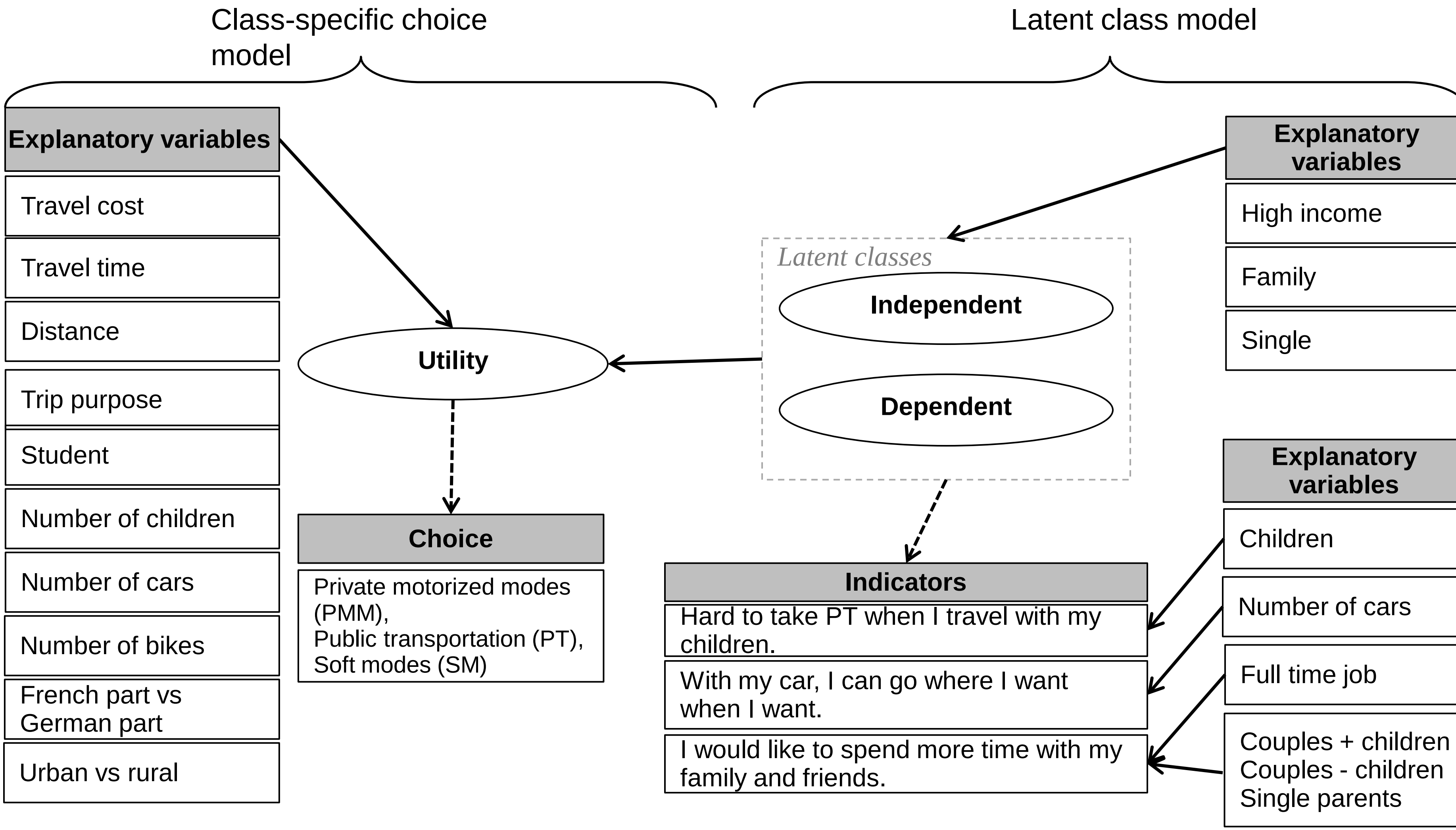
LCCM 2



$$\mathcal{L}_{\mathrm{LCM2}} = \prod_n \{P_n(i|\text{class } 1)\cdot\pi_{11}\cdot\pi_{21}\cdot\pi_{31}\cdot P_n(\text{class } 1) + P_n(i|\text{class } 2)\cdot\pi_{12}\cdot\pi_{22}\cdot\pi_{32}\cdot P_n(\text{class } 2)\}$$

MODE CHOICE CASE STUDY

LCCM 3



$$\mathscr{L}_{\text{LCM3}} = \prod_n \{P_n(i|\text{class } 1) \cdot P_n(I1|\text{class } 1) \cdot P_n(I2|\text{class } 1) \cdot P_n(I3|\text{class } 1) \cdot P_n(\text{class } 1)$$

$$+P_n(i|\text{class } 2) \cdot P_n(I1|\text{class } 2) \cdot P_n(I2|\text{class } 2) \cdot P_n(I3|\text{class } 2) \cdot P_n(\text{class } 2)\}$$

MODE CHOICE CASE STUDY

Estimation results for the class-membership model

	LCCM 1		LCCM 2		LCCM 3	
Parameters	estimate	t-test	estimate	t-test	estimate	t-test
ASC_{class}	-0.215	-0.86**	-0.629	-3.25	-0.589	-3.39
γ_{family}	0.136	0.51**	3.92	4.84	0.967	5.41
γ_{income}	0.693	2.76	0.460	2.22	0.684	4.50
γ_{single}	0.408	1.34**	0.704	3.57	0.743	3.33

- Increase of the significance of the parameters of the latent class model.

MODE CHOICE CASE STUDY

Estimation results for the class-membership model

	LCCM 1		LCCM 2		LCCM 3	
Parameters	estimate	t-test	estimate	t-test	estimate	t-test
ASC_{class}	-0.215	-0.86**	-0.629	-3.25	-0.589	-3.39
γ_{family}	0.136	0.51**	3.92	4.84	0.967	5.41
γ_{income}	0.693	2.76	0.460	2.22	0.684	4.50
γ_{single}	0.408	1.34**	0.704	3.57	0.743	3.33

- Increase of the significance of the parameters of the latent class model.
- Income parameter has become more important.

MODE CHOICE CASE STUDY

Model application: computation of VOT

		VOT PMM [CHF/hour]	VOT PT [CHF/hour]
LCCM 1	Class independent	3.06	3.72
	Class dependent	52.63	17.53
	Overall	28.97	10.94
LCCM 2	Class independent	35.78	15.38
	Class dependent	22.05	8.84
	Overall	29.53	12.40
LCCM 3	Class independent	63.27	16.21
	Class dependent	34.16	5.99
	Overall	36.94	18.40

- VOTs comparable with literature on transport economics (Jara-Diaz, 2007), where VOT can be compared to wage rate.
- Individuals in the independent class have higher incomes (> 8000 CHF), hence a higher value of time.

CONCLUSION

Main findings:

- Specified **LCCM with psychometric indicators** as **measurements of LC**
- Account for **heterogeneity of response behavior** that can be captured by individual-specific information in measurement model of LCM
- **Importance of accounting for it in LCCM**
    - Socio-economic characteristics affect response to opinion questions significantly
    - Parameters of the class membership utility increase in significance
    - VOT are comparable with existing studies
- **Residual analysis needed to further validate the model**

Thanks!