

DEMAND MODELS FOR TRANSPORTATION MODES

A FOCUS ON THE MEASUREMENT OF LATENT CONSTRUCTS AFFECTING DECISIONS

Aurélie Glerum
Ricardo Hurtubia
My Hang Nguyen
Bilge Atasoy
Michel Bierlaire

TLA/ToL joint seminar
KTH Royal Institute of Technology

Friday 12th October 2012

Introduction & motivation

Methodology

The data

- Vehicle choice case study
- Mode choice case study

Incorporation of measurements into HCM

- Vehicle choice case study (ICLV example)
- Mode choice case study (ICLC example)

Conclusion

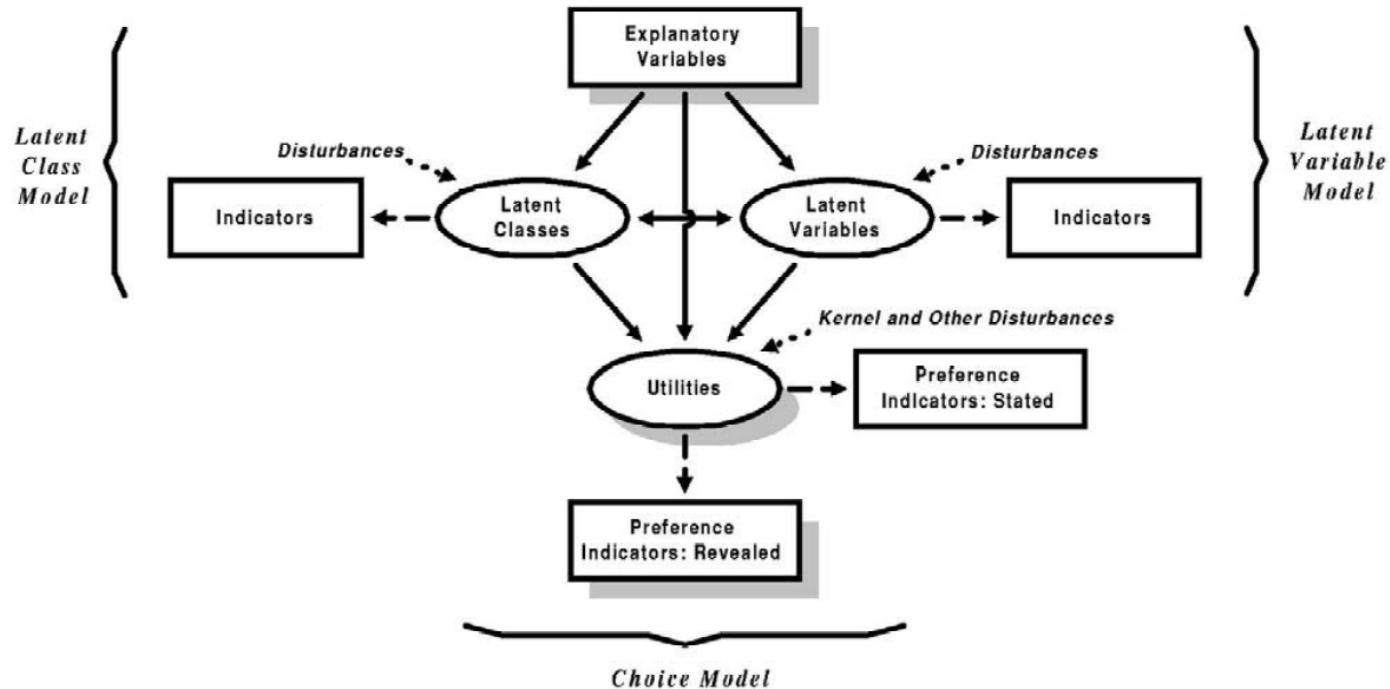
Recent developments in demand modeling for transportation

- **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.



- 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.

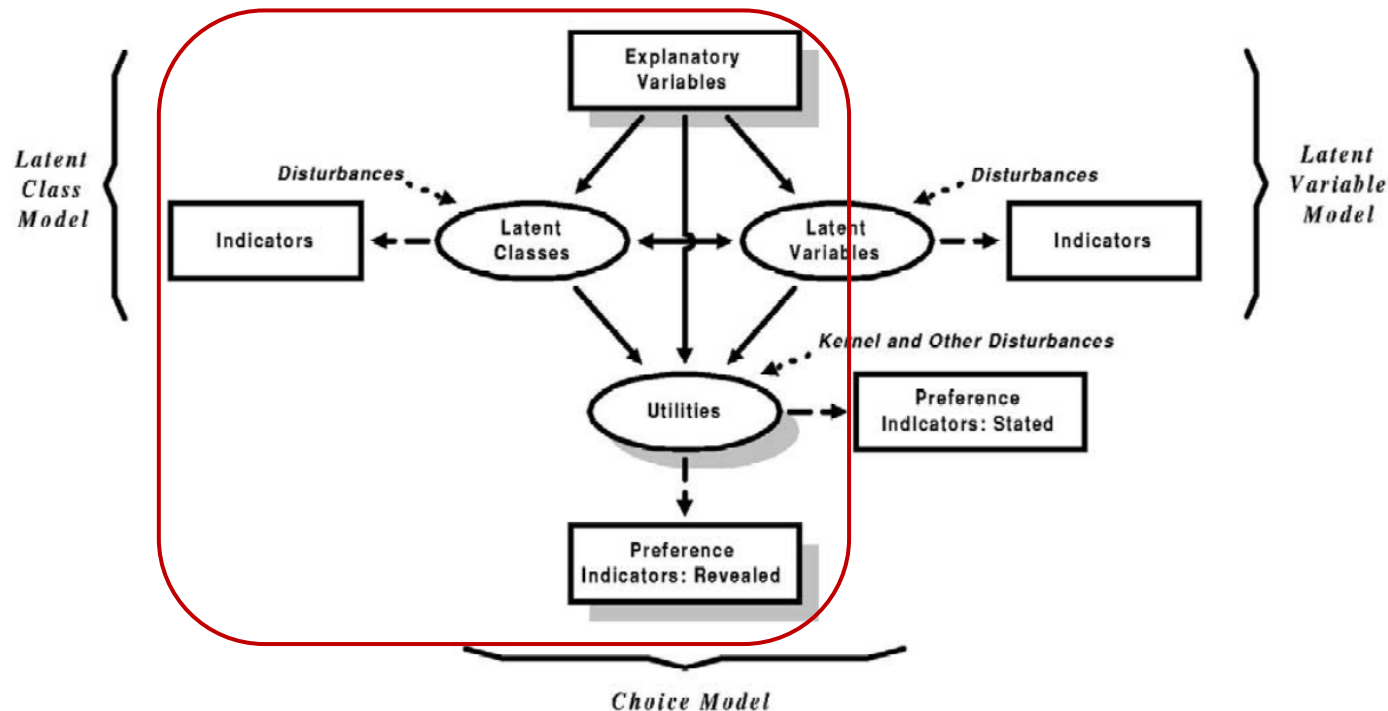
Hybrid choice model (HCM): DCM with latent constructs.



Hybrid choice model (HCM): DCM with latent constructs.

Latent construct can be... either a **latent class model**

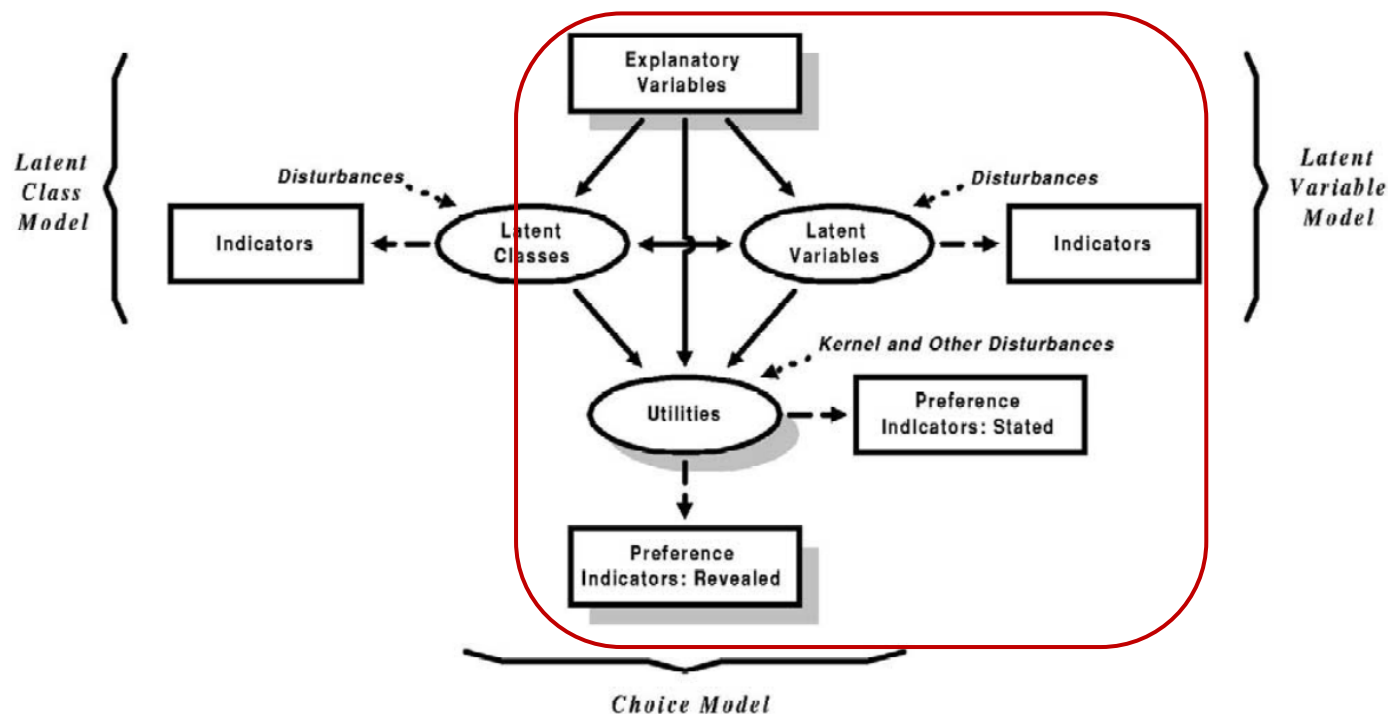
- Unobservable construct is **discrete**
- Useful for **segmentation** according to lifestyle



Hybrid choice model (HCM): DCM with latent constructs.

Latent construct can be... or a **latent variable model**

- Unobservable construct is **continuous**
- Useful to analyze the impact of changes in prices across individuals → **pricing**



Important issues in the use of HCMs:

1. **Measurement of latent variable / latent class**

⇒ How to obtain the most realistic and accurate measure of an attitude / perception / lifestyle?

Opinion statements: usual way in the literature

2. **Integration of the measurement into the choice model**

⇒ How to incorporate this information in the choice modeling framework?

Important issues in the use of HCMs:

1. **Measurement of latent variable / latent class**

⇒ How to obtain the most realistic and accurate measure of an attitude / perception / lifestyle?

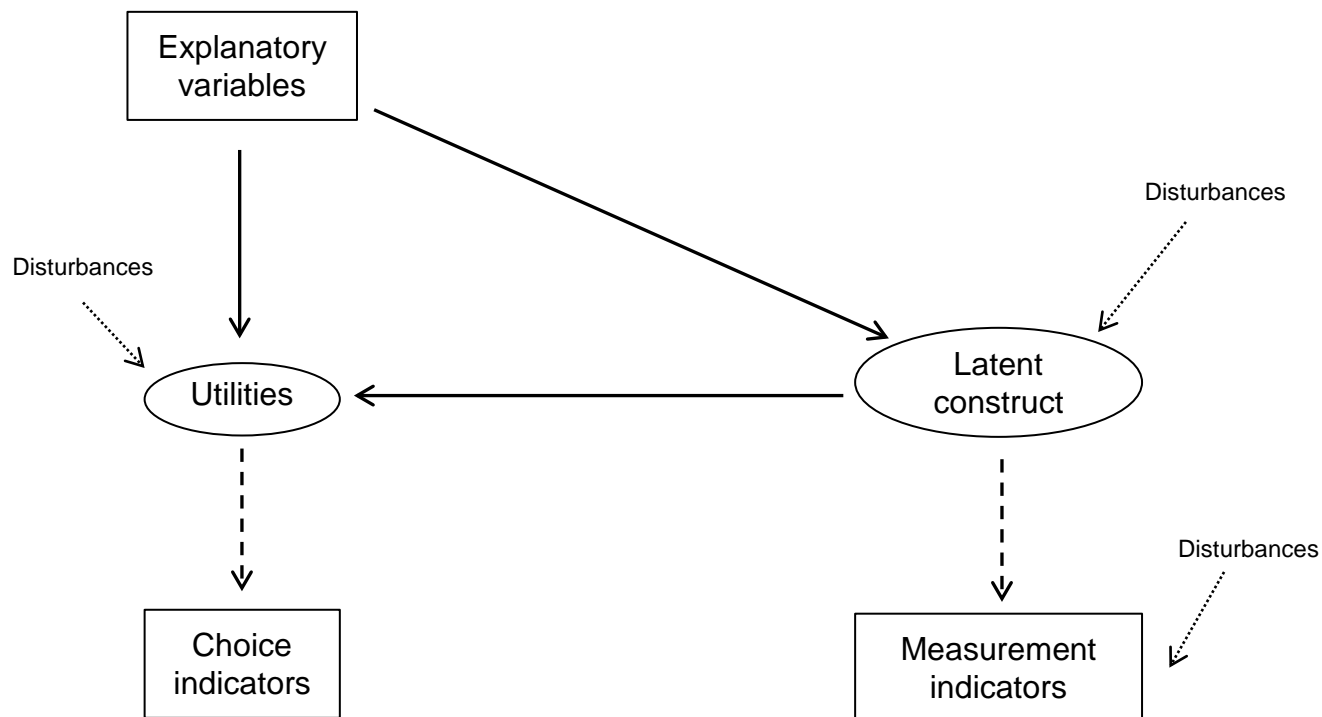
Opinion statements: usual way in the literature

2. **Integration of the measurement into the choice model**

⇒ How to incorporate this information in the choice modeling framework?

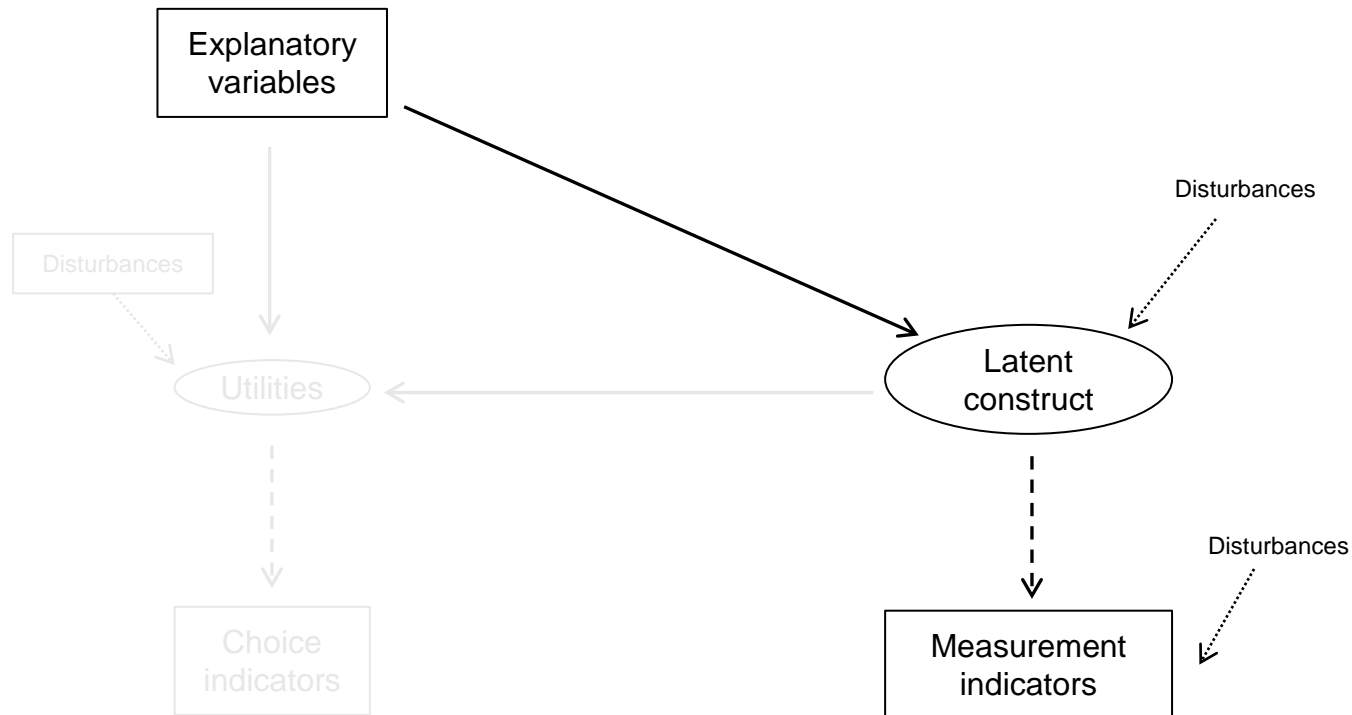
Focus of this research: measurement model

Integration of the measurement into the choice model



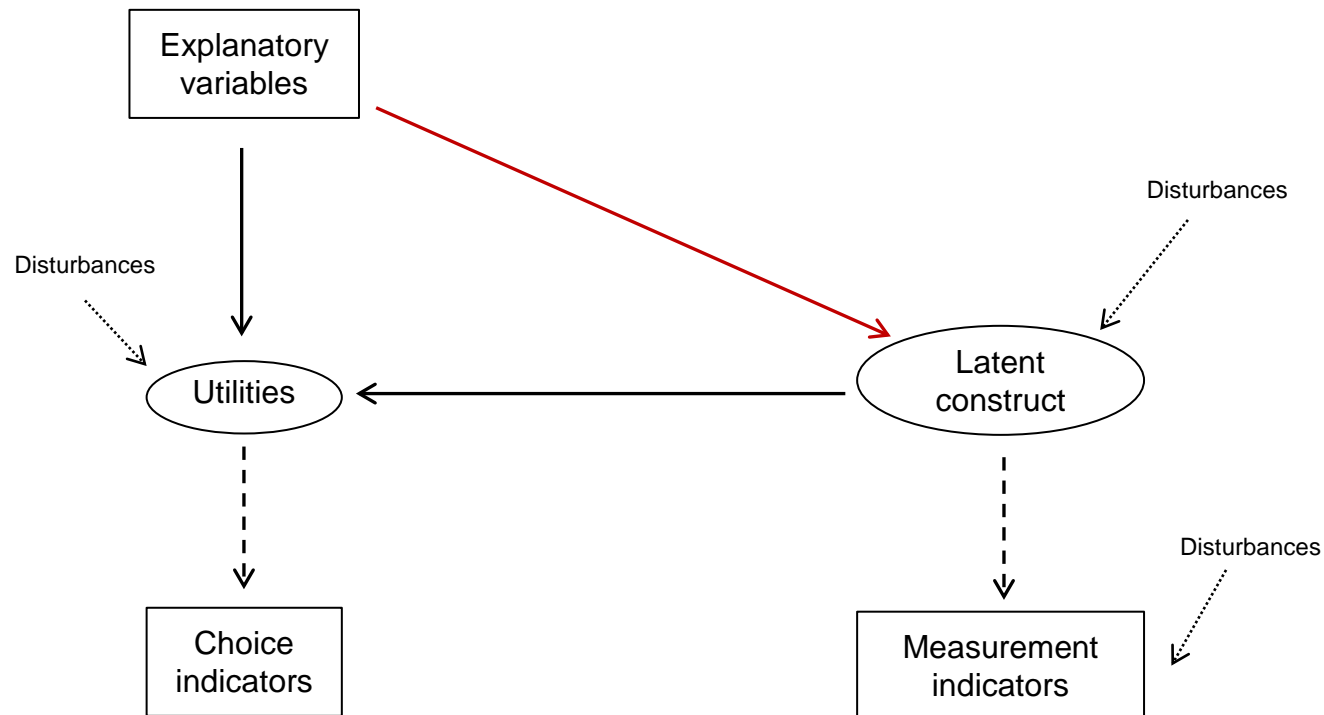
Integration of the measurement into the choice model:

- **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).



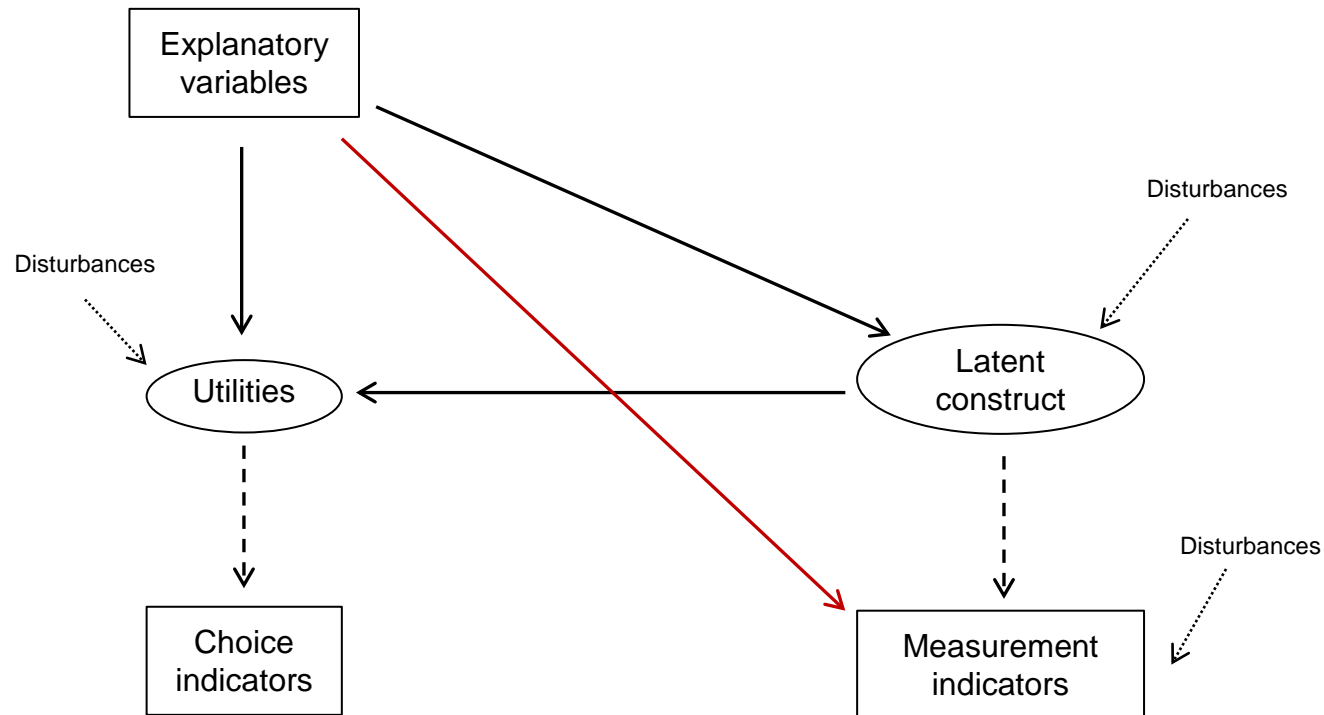
Integration of the measurement into the choice model

- In transportation applications:
 - **Heterogeneity of latent construct (e.g. attitude) captured among population**
 - But: also need to capture heterogeneity in reporting indicators of latent construct



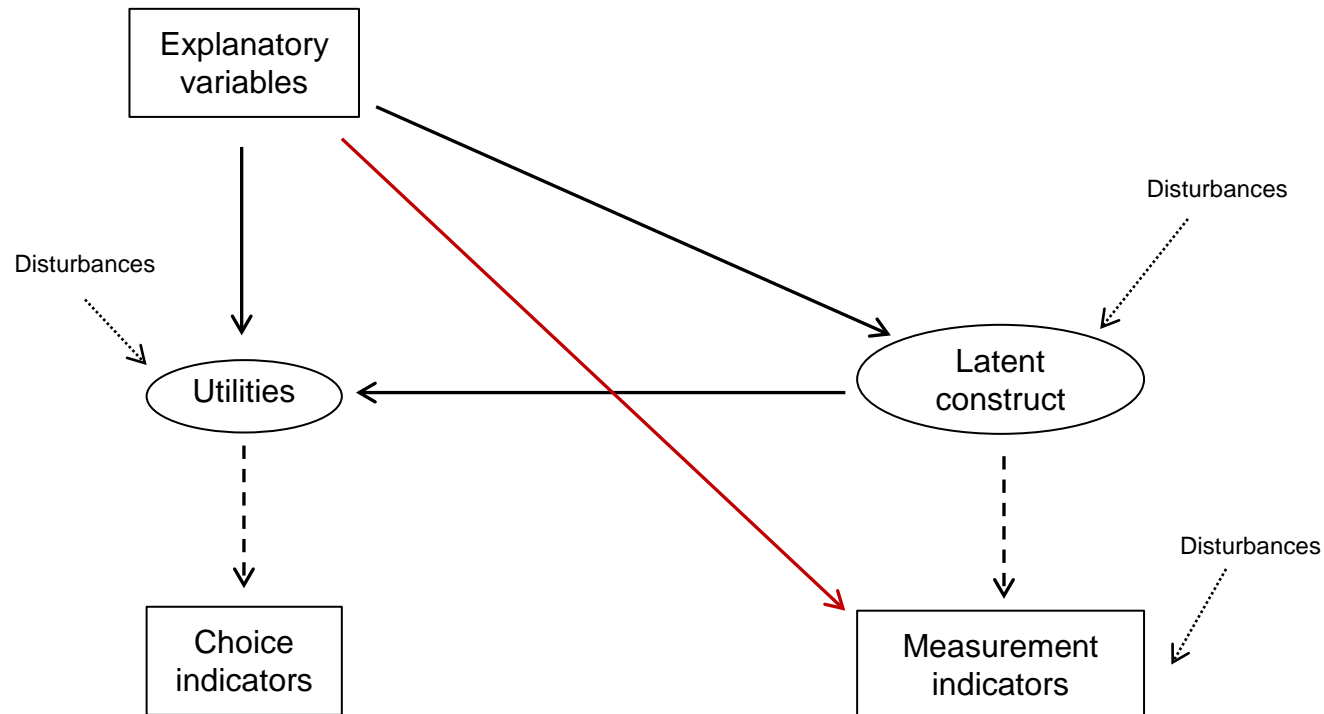
Integration of the measurement into the choice model

- In transportation applications:
 - Heterogeneity of latent construct (e.g. attitude) captured among population
 - But: also need to **capture heterogeneity in reporting indicators of latent construct**



Integration of the measurement into the choice model

- In transportation applications:
 - Heterogeneity of latent construct (e.g. attitude) captured among population
 - But: also need to **capture heterogeneity in reporting indicators of latent construct**
- Focus of this presentation



Model specification

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

Integrated choice and latent variable model

$$f(y_{in}, I_n | X_{in}; \alpha, \beta, \lambda, \sigma_\omega) = \int_{X_n^*} P(y_{in} | X_{in}, X_n^*; \beta)^{y_{in}} \cdot f(I_n | X_{in}, X_n^*; \alpha) \cdot f(X_n^* | X_n; \lambda, \sigma_\omega) dX_n^*$$

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

Integrated choice and latent class model

$$P(y_{in}, I_n | X_{in}; \alpha, \beta, \lambda, \sigma_\omega) = \left\{ \sum_{s \in S} P(y_{in} | X_{in}, s; \beta) \cdot P(I_n | X_{in}, s; \alpha) \cdot P(s | X_n; \lambda, \sigma_\omega) \right\}^{y_{in}}$$

Model specification

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

Integrated choice and latent variable model

$$f(y_{in}, I_n | X_{in}; \alpha, \beta, \lambda, \sigma_\omega) = \int_{X_n^*} P(y_{in} | X_{in}, X_n^*; \beta)^{y_{in}} \cdot f(I_n | X_{in}, X_n^*; \alpha) \cdot f(X_n^* | X_n; \lambda, \sigma_\omega) dX_n^*$$

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

Integrated choice and latent class model

$$P(y_{in}, I_n | X_{in}; \alpha, \beta, \lambda, \sigma_\omega) = \left\{ \sum_{s \in S} P(y_{in} | X_{in}, s; \beta) \cdot P(I_n | X_{in}, s; \alpha) \cdot P(s | X_n; \lambda, \sigma_\omega) \right\}^{y_{in}}$$

Few examples that incorporate socio-economic information into the measurement model

Two case studies:

1. **Integrated choice and latent variable model (ICLV):** analysis of the impact of pro-convenience attitude on choice of car.

Car purchase choice case study

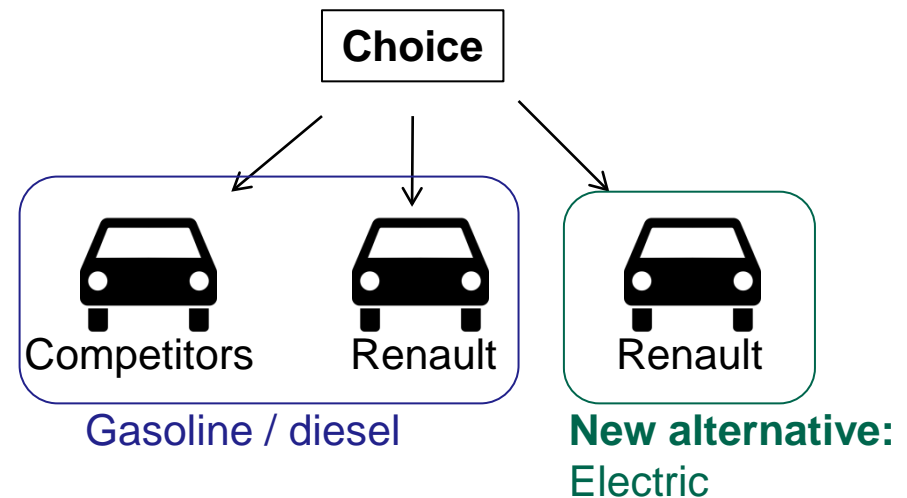
2. **Integrated choice and latent class model (ICLC):** analysis of the transportation mode choices for individuals segmented according to dependent / independent classes.

Mode choice case study

VEHICLE CHOICE CASE STUDY

Stated preferences (SP) survey:

- **Car purchase choice study**
- Conducted in Switzerland in 2011 among individuals who bought a new car recently or intend to buy one soon.
- Conducted with Renault Suisse SA.
- Customized choice situations
- **693 questionnaires** obtained



VEHICLE CHOICE CASE STUDY

Opinion statements related to five themes

- Environmental concern *An electric car is a 100% ecological solution.*
- Attitude towards new technologies *A control screen is essential in my use of a car.*
- Perception of the reliability of an electric vehicle *Electric cars are not as secure as gasoline cars.*
- Perception of leasing *Leasing is an optimal contract which allows me to change car frequently.*
- Attitude towards design *Design is a secondary element when purchasing a car, which is above all a practical transport mode.*

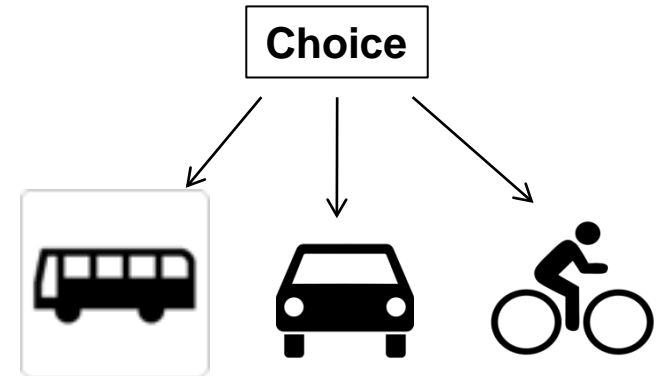
Ratings

- Total disagreement (1)
- Disagreement (2)
- Neutral opinion (3)
- Agreement (4)
- Total agreement (5)
- I don't know (6)

MODE CHOICE CASE STUDY

Revealed preferences (RP) survey

- **Mode choice study**
- Conducted between 2009-2010 in low-density areas of Switzerland
- Conducted with PostBus (major bus company in Switzerland, operates in low-density areas)
- Info on **all trips performed by inhabitants in one day**:
 - Transport mode
 - Trip duration
 - Cost of trip
 - Activity at destination
 - Etc.
- **1763 valid questionnaires** collected



MODE CHOICE CASE STUDY

Opinion statements related to four themes

- **Environment** *The price of gasoline should be increased in order to reduce traffic congestion and air pollution.*
- **Mobility** *Taking the bus helps making a town more comfortable and welcoming.*
- **Residential choice** *Accessibility and mobility conditions are important in the choice of an accommodation.*
- **Lifestyle** *I always plan my activities a long time in advance.*

Ratings

- Total disagreement (1)
- Disagreement (2)
- Neutral opinion (3)
- Agreement (4)
- Total agreement (5)
- I don't know (6)

Role of indicators of latent construct:

- Measure a latent variable
- Enhance a latent class model

Issue: biases in the measurement of indicators due to **heterogeneity of response behavior**

By introducing socio-economic information into the measurement component of the HCM, the bias is reduced.

Two examples:

- **Car choice case study (ICLM):** capture exaggeration effects in responses to indicators.
- **Transportation mode choice case study (ICLC):** capture bias in responses to indicators due to various socio-economic characteristics.

Motivation for integration of explanatory factors of measurement indicators:

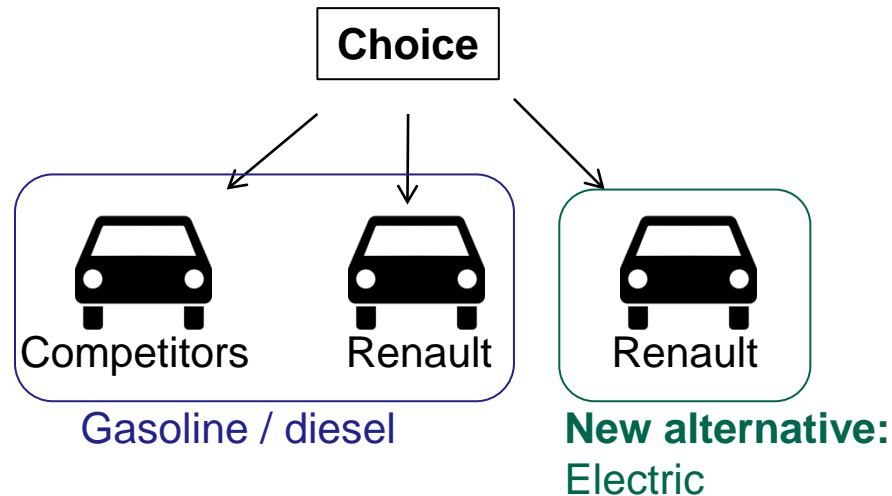
- **Dispersion effects:**
 - Exaggeration effects in experiments on survey design in social science literature (Schuman and Presser, 1996)
 - Some individuals tend to report responses at extremities of scale of agreement though their commitment to the opinion statement is not strong.
- **Socio-economic characteristics** might explain different response behaviors



Need to account for heterogeneity of response behavior

1. **Integrated choice and latent variable model (ICLV):** analysis of the impact of pro-convenience attitude on choice of car.

Vehicle choice case study

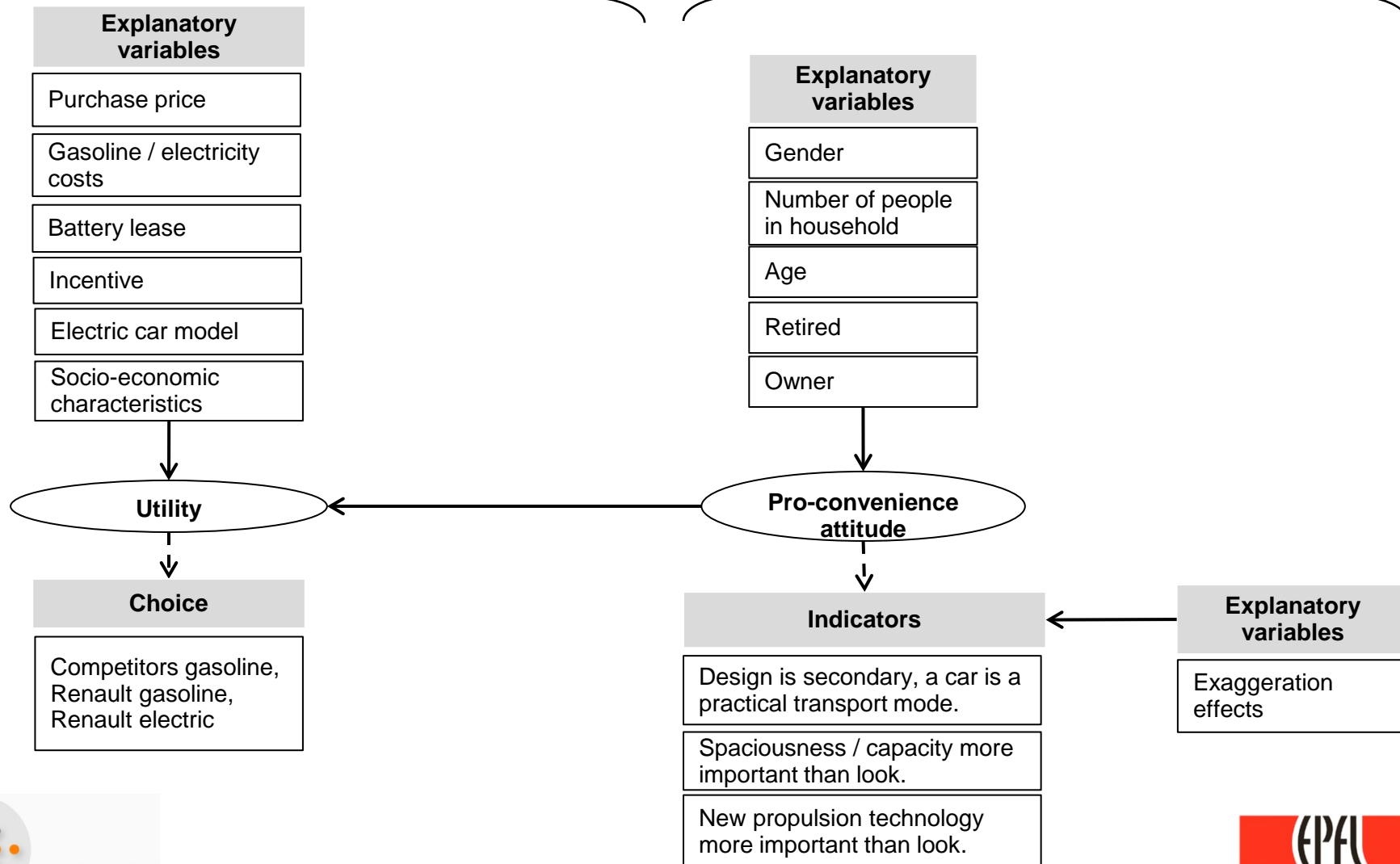


INCORPORATION OF MEASUREMENTS INTO HCM²⁴

VEHICLE CHOICE CASE STUDY (ICLV EXAMPLE)

Discrete choice model

Latent variable model



Definition of index:

- Definition of *degree of extremity*

$$E_n = \sum_{r=1}^R J_{rn} \quad \text{with} \quad J_{rn} = \begin{cases} 1 & \text{if } I_{rn} = 1 \text{ or } I_{rn} = 5 \\ 0 & \text{otherwise} \end{cases}$$

- E_n : number of occurrences of 'total disagreement' and 'total agreement' for individual n over all R opinion questions of the survey

Definition of scale parameter:

- Measurement model:

$$I_n^* = m(X_n^*; \alpha) + v_n$$

$$v_n \sim \text{Logistic}(0, \sigma_{v_n})$$

- Scale that captures heterogeneity in response behavior:

$$\sigma_{v_n} = I_{E_n < \theta} \cdot 1 + (1 - I_{E_n < \theta}) \cdot \sigma_{v_{Ext}}(E_n)$$

$$= I_{E_n < \theta} \cdot 1 + (1 - I_{E_n < \theta}) \cdot E_n \cdot \gamma$$

Definition of scale parameter:

- Measurement model:

$$I_n^* = m(X_n^*; \alpha) + v_n$$

$$v_n \sim \text{Logistic}(0, \sigma_{v_n})$$

- Scale that captures heterogeneity in response behavior:

$$\sigma_{v_n} = I_{E_n < \theta} \cdot 1 + (1 - I_{E_n < \theta}) \cdot \sigma_{v_{Ext}}(E_n)$$

$$= I_{E_n < \theta} \cdot 1 + (1 - I_{E_n < \theta}) \cdot E_n \cdot \gamma$$

Define **threshold θ** above which individuals show extreme behavior
Statistical analyses show that highest fit for **$\theta = 7$** .

Definition of scale parameter:

- Measurement model:

$$I_n^* = m(X_n^*; \alpha) + v_n$$

$$v_n \sim \text{Logistic}(0, \sigma_{v_n})$$

- Scale that captures heterogeneity in response behavior:

$$\sigma_{v_n} = I_{E_n < \theta} \cdot 1 + (1 - I_{E_n < \theta}) \cdot \sigma_{v_{Ext}}(E_n) \quad \text{Group-specific scale}$$

$$= I_{E_n < \theta} \cdot 1 + (1 - I_{E_n < \theta}) \cdot E_n \cdot \gamma$$

Definition of scale parameter:

- Measurement model:

$$I_n^* = m(X_n^*; \alpha) + v_n$$

$$v_n \sim \text{Logistic}(0, \sigma_{v_n})$$

- Scale that captures heterogeneity in response behavior:

$$\begin{aligned}\sigma_{v_n} &= I_{E_n < \theta} \cdot 1 + (1 - I_{E_n < \theta}) \cdot \sigma_{v_{Ext}}(E_n) \\ &= I_{E_n < \theta} \cdot 1 + (1 - I_{E_n < \theta}) \cdot \boxed{E_n \cdot \gamma}\end{aligned}$$

Progressive scale:

- The higher the degree of extremity, the higher the scale.
- γ parameter to estimate

INCORPORATION OF MEASUREMENTS INTO HCM³⁰

VEHICLE CHOICE CASE STUDY (ICLV EXAMPLE)

Results for the latent variable model

<i>Structural equation</i>			<i>Measurement equation</i>		
Name	Value	<i>t</i> -test	Name	Value	<i>t</i> -test
β_{Mean}	-6.03	-17.32	τ_1	-9.23	-33.72
β_{Male}	-0.256	-1.54**	γ	0.203	29.62
β_{NbPeople}	0.362	5.46	δ_1	4.76	32.36
β_{Age}	0.0166	5.55	δ_2	2.15	40.76
β_{Retired}	1.40	5.31	δ_3	3.45	41.46
$\beta_{\text{Homeowner}}$	0.673	4.31	α_2	0.552	31.53
σ_ω	3.21	28.04	α_3	0.574	22.61

Simultaneous estimation of the HCM using the extended version of Biogeme (Bierlaire and Fétiarison, 2009)

Results from the latent variable model

<i>Structural equation</i>			<i>Measurement equation</i>		
Name	Value	<i>t</i> -test	Name	Value	<i>t</i> -test
β_{Mean}	-6.03	-17.32	τ_1	-9.23	-33.72
β_{Male}	-0.256	-1.54**	γ	0.203	29.62
β_{NbPeople}	0.362	5.46	δ_1	4.76	32.36
β_{Age}	0.0166	5.55	δ_2	2.15	40.76
β_{Retired}	1.40	5.31	δ_3	3.45	41.46
$\beta_{\text{Homeowner}}$	0.673	4.31	α_2	0.552	31.53
σ_ω	3.21	28.04	α_3	0.574	22.61

$$\sigma_{v_n} = I_{E_n < \theta} \cdot 1 + (1 - I_{E_n < \theta}) \cdot E_n \cdot \gamma$$

Results from the latent variable model

<i>Structural equation</i>			<i>Measurement equation</i>		
Name	Value	<i>t</i> -test	Name	Value	<i>t</i> -test
β_{Mean}	-6.03	-17.32	τ_1	-9.23	-33.72
β_{Male}	-0.256	-1.54**	γ	0.203	29.62
β_{NbPeople}	0.362	5.46	δ_1	4.76	32.36
β_{Age}	0.0166	5.55	δ_2	2.15	40.76
β_{Retired}	1.40	5.31	δ_3	3.45	41.46
$\beta_{\text{Homeowner}}$	0.673	4.31	α_2	0.552	31.53
σ_ω	3.21	28.04	α_3	0.574	22.61

$$\sigma_{v_n} = I_{E_n < \theta} \cdot 1 + (1 - I_{E_n < \theta}) \cdot E_n \cdot \gamma$$

We observe dispersion effects, since for the 'extreme' group we have:

$$\sigma_{v_n} = 7 \cdot \gamma = 1.42$$

VEHICLE CHOICE CASE STUDY (ICLV EXAMPLE)

Results from the choice model

Name	Value	t-test	Name	Value	t-test
<i>Parameters in linear terms</i>			<i>Parameters in linear terms (ctd)</i>		
ASC_{CG}	-2.54	-4.23	$\beta_{Battery}$	-4.73	-1.63**
ASC_{RG}	-1.78	-2.98	$\beta_{French_{CG}}$	0.347	2.77
$\beta_{UseCostGasoline}$	-0.0706	-2.10	$\beta_{French_{RG}}$	0.109	0.91**
$\beta_{UseCostElecHigh_{Fluence}}$	-0.282	-2.35	$\beta_{Age_{CG}}$	0.0206	4.37
$\beta_{UseCostElecHigh_{Zoé}}$	-0.818	-5.13	$\beta_{Age_{RG}}$	0.00487	1.09**
$\beta_{UseCostElecMed_{Zoé}}$	-0.483	-3.11	$\beta_{TG12_{CG}}$	1.66	4.35
$\beta_{IncentiveHigh}$	0.748	5.80	$\beta_{TG12_{RG}}$	0.681	1.80*
$\beta_{IncentiveMed}$	0.0630	0.47**	$\beta_{TG3_{CG}}$	-0.984	-1.33**
$\beta_{IncentiveLow}$	-0.0150	-0.11**	$\beta_{TG3_{RG}}$	1.29	3.10
$\beta_{PT_{CG,TG1245}}$	-0.251	-1.86*	<i>Parameters in non-linear terms</i>		
$\beta_{PT_{RG,TG1245}}$	-0.596	-4.03	$\beta_{price_{CG}}$	-4.15	-6.05
$\beta_{PT_{CG,TG3}}$	-2.10	-2.88	$\beta_{price_{RG,TG1245}}$	-1.97	-6.36
$\beta_{PT_{RG,TG3}}$	-1.01	-4.63	$\beta_{price_{RG,TG3}}$	-0.843	-3.51
$\beta_{NbCars_{CG}}$	-0.269	-3.65	$\beta_{price_{RE,TG12}}$	-1.01	-7.05
$\beta_{NbCars_{RG}}$	-0.361	-5.48	$\beta_{price_{RE,TG3}}$	-0.843	-3.51
$\beta_{Income_{CG}}$	-0.272	-2.33	$\beta_{price_{RE,TG45}}$	-0.766	-4.62
$\beta_{Income_{RG}}$	-0.281	-2.64	β_X^*	-0.0527	-4.81

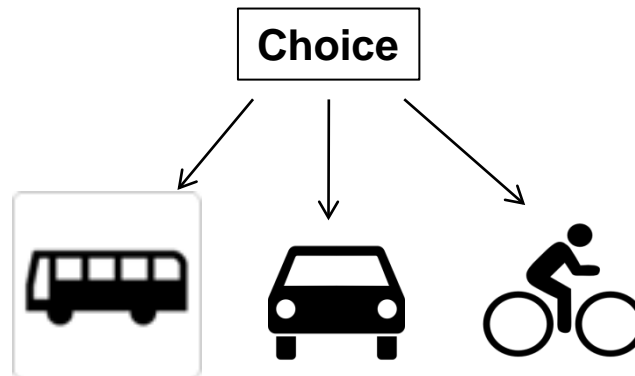
Pro-convenience attitude significantly affects car choice.

Improvement of fit over model without dispersion effects

Model	Q	$\mathcal{L}(0)$	$\mathcal{L}(\hat{\mu})$	$\bar{\rho}^2$
Without dispersion	46	-16'746	-14'030	0.16
With dispersion	47	-13'687	-18'083	0.24

2. **Integrated choice and latent class model (ICLC):** analysis of the transportation mode choices for individuals segmented according to dependent / independent classes.

Mode choice case study

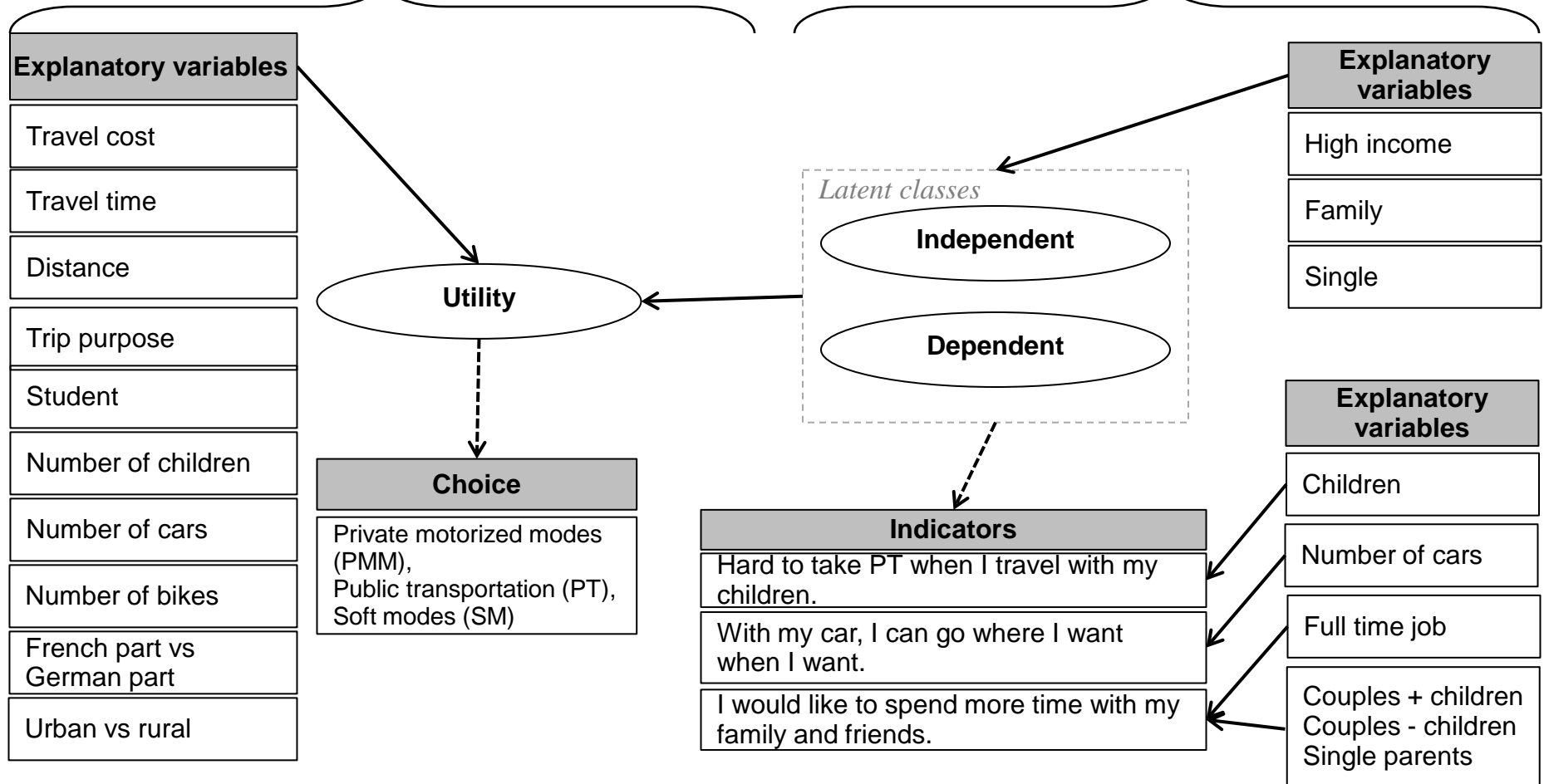


INCORPORATION OF MEASUREMENTS INTO HCM ³⁶

MODE CHOICE CASE STUDY (ICLC EXAMPLE)

Class-specific choice model

Latent class model



Class-specific measurement equations:

$$\tilde{I}_{k,n}^s = m(X_n; \lambda^s) + \xi_{k,n}^s \text{ with } \xi_{k,n}^s \sim \text{Logistic}(0, 1)$$

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

Class-specific measurement equations:

Class-specific parameters

$$\tilde{I}_{k,n}^s = m(X_n; \lambda^s) + \xi_{k,n}^s \text{ with } \xi_{k,n}^s \sim \text{Logistic}(0, 1)$$

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

Class-specific measurement equations:

Class-specific parameters

$$\tilde{I}_{k,n}^s = m(\boxed{X_n}; \lambda^s) + \xi_{k,n}^s \text{ with } \xi_{k,n}^s \sim \text{Logistic}(0, 1)$$

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

Socio-economic information as explanatory variables of response to indicators

Estimation results for ICLC

Parameters	No indicators		Atasoy et al., 2012		Extended ICLC	
	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.

Estimation results for LCCM

Parameters	No indicators		Atasoy et al., 2012		Extended ICLC	
	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**.

INCORPORATION OF MEASUREMENTS INTO HCM⁴²

MODE CHOICE CASE STUDY (ICLC EXAMPLE)

Model application: computation of VOT

ICLC		VOT [CHF/hour]	PMM	VOT [CHF/hour]	PT
No indicators	Class independent	3.06		3.72	
	Class dependent	52.63		17.53	
	Overall	28.97		10.94	
Atasoy et al., 2012	Class independent	35.78		15.38	
	Class dependent	22.05		8.84	
	Overall	29.53		12.40	
Extended ICLC	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.

Main findings:

- **Heterogeneity of response behavior** exists and can be captured by individual-specific information in measurement model
- **Evidence for the importance of accounting for it:**
 - **ICLV model of car choice:**
 - Significant scale parameter
 - Increases as degree of extremity increases
 - **ICLC model of mode choice:**
 - Socio-economic characteristics affect response to opinion questions significantly
 - Parameters of the class membership utility increase in significance
 - VOT are comparable with existing studies

Thanks!