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
Introduction & motivation

Literature

Methodology

Mode choice case study

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.
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
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).
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
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).
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
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
Specification of integrated choice and latent class model

**Choice model:**
\[ U_{in}^s = V(X_{in}, X_n, \beta^s) + \varepsilon_{in}^s, \quad \text{with } \varepsilon_{in}^s \sim EV(0, 1) \]

**Latent class model:**

- **Class-membership function:**
  \[ F_{ns} = f(X_n, \gamma^s) + \xi_{ns}, \quad \text{where } \xi_{ns} \sim EV(0, 1) \]

- **Measurement indicators:**
  \[ \tilde{G}_{lk,n}^s = g(X_n; \alpha_k^s) + \nu_{kn}^s, \quad \text{with } \nu_{kn}^s \sim \text{Logistic}(0, 1) \]
  
  \[ G_{lk,n}^s = \begin{cases} 
  1 & \text{if } -\infty < \tilde{G}_{lk,n}^s \leq \tau_{1,k}^s \\
  2 & \text{if } \tau_{1,k}^s < \tilde{G}_{lk,n}^s \leq \tau_{2,k}^s \\
  3 & \text{if } \tau_{2,k}^s < \tilde{G}_{lk,n}^s \leq \tau_{3,k}^s \\
  4 & \text{if } \tau_{3,k}^s < \tilde{G}_{lk,n}^s \leq \tau_{4,k}^s \\
  5 & \text{if } \tau_{4,k}^s < \tilde{G}_{lk,n}^s \leq +\infty 
\end{cases} \]
Specification of integrated choice and latent class model

**Choice model:**

\[ U_{in}^s = V(X_{in}, X_n, \beta^s) + \xi_{in}^s, \quad \text{with} \quad \xi_{in}^s \sim EV(0, 1) \]

**Latent class model:**

- Class-membership function:

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

- Measurement indicators:

\[ \tilde{G}_{lk,n}^s = g(X_n, \alpha_k^s) + v_{kn}^s, \quad \text{with} \quad v_{kn}^s \sim \text{Logistic}(0, 1) \]

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

Class-specific parameters
Specification of integrated choice and latent class model

**Choice model:**

\[ U_{in}^s = V(X_{in}, X_n, \beta^s) + \epsilon_{in}^s, \quad \text{with } \epsilon_{in}^s \sim EV(0, 1) \]

**Latent class model:**

- **Class-membership function:**
  \[ F_{ns} = f(X_n, \gamma^s) + \xi_{ns}, \quad \text{where } \xi_{ns} \sim EV(0, 1) \]

- **Measurement indicators:**
  \[ \tilde{G}_{lk,n}^s = g(X_n, \alpha_k^s) + \nu_{kn}^s, \quad \text{with } \nu_{kn}^s \sim \text{Logistic}(0, 1) \]

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

Socio-economic information of the individual
Specification of integrated choice and latent class model

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

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

\[
y_{in} = \begin{cases} 1 \text{ if } U_{in} = \max_j U_{jn} \\ 0 \text{ otherwise} \end{cases}
\]
Specification of integrated choice and latent class model

Likelihood function given by:

\[ L = \prod_{n=1}^{N} f(y_{in}, I_n | X_{in}; \alpha, \beta, \lambda) \]

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

\[ f(y_{in}, I_n | X_{in}; \alpha, \beta, \lambda) = \int_{X^*_n} P(y_{in} | X_{in}, X_n^*; \beta) \cdot \prod_{k} P(I_{kn} | X_{in}, X_n^*; \alpha) \cdot f(X_n^* | X_n; \lambda) dX_n^* \]
Specification of integrated choice and latent class model

Likelihood function given by:

\[ L = \prod_{n=1}^{N} f(y_{in}, I_n | X_{in}; \alpha, \beta, \lambda) \]

with

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

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

Analogy with integrated choice and latent variable model

Class-specific parameters in measurement equation
Specification of integrated choice and latent class model

Likelihood function given by:

\[ L = \prod_{n=1}^{N} f(y_{in}, I_n | X_{in}; \alpha, \beta, \lambda) \]

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
Specification of integrated choice and latent class model

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

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

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

Analogy with integrated choice and latent variable model

Class-specific parameters in measurement equation

Socio-economic variables as explanatory of response to indicator
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
Development of latent class choice model:

Analysis of the transportation mode choices for individuals segmented according to dependent class and 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
MODE CHOICE CASE STUDY

Latent class model

Class-specific choice model

Explanatory variables

Travel cost
Travel time
Distance
Trip purpose
Student
Number of children
Number of cars
Number of bikes
French part vs German part
Urban vs rural

Utility

Choice

Private motorized modes (PMM), Public transportation (PT), Soft modes (SM)

Latent classes

Independent
Dependent

Explanatory variables

High income
Family
Single

\[ \mathcal{L}_{LCCM} = \prod_{n} \left\{ P_n(i|\text{class 1}) \cdot P_n(\text{class 1}) + P_n(i|\text{class 2}) \cdot P_n(\text{class 2}) \right\} \]
MODE CHOICE CASE STUDY

Class-specific choice model

Latent class model

Explanatory variables
- Travel cost
- Travel time
- Distance
- Trip purpose
- Student
- Number of children
- Number of cars
- Number of bikes
- French part vs German part
- Urban vs rural

Utility

Choice

Private motorized modes (PMM), Public transportation (PT), Soft modes (SM)

Latent classes
- Independent
- Dependent

Indicators
- Hard to take PT when I travel with my children.
- With my car, I can go where I want when I want.
- I would like to spend more time with my family and friends.

Explanatory variables
- High income
- Family
- Single

\[
S_{LCCM2} = \prod_{n} \{ P_n(i|\text{class 1}) \cdot \pi_{11} \cdot \pi_{21} \cdot \pi_{31} \cdot P_n(\text{class 1}) + P_n(i|\text{class 2}) \cdot \pi_{12} \cdot \pi_{22} \cdot \pi_{32} \cdot P_n(\text{class 2}) \}
\]
MODE CHOICE CASE STUDY

LCCM 3

Class-specific choice model

Latent class model

Explanatory variables
- Travel cost
- Travel time
- Distance
- Trip purpose
- Student
- Number of children
- Number of cars
- Number of bikes
- French part vs German part
- Urban vs rural

Utility

Choice

Indicators
- Hard to take PT when I travel with my children.
- With my car, I can go where I want when I want.
- I would like to spend more time with my family and friends.

Latent classes
- Independent
- Dependent

Explanatory variables
- High income
- Family
- Single

Explanatory variables
- Children
- Number of cars
- Full time job
- Couples + children
- Couples - children
- Single parents

\[
\mathcal{L}_{\text{LCCM3}} = \prod_{n} \{P_{n}(i|\text{class 1}) \cdot P_{n}(I1|\text{class 1}) \cdot P_{n}(I2|\text{class 1}) \cdot P_{n}(I3|\text{class 1}) \cdot P_{n}(\text{class 1}) \\
+ P_{n}(i|\text{class 2}) \cdot P_{n}(I1|\text{class 2}) \cdot P_{n}(I2|\text{class 2}) \cdot P_{n}(I3|\text{class 2}) \cdot P_{n}(\text{class 2}) \} 
\]
### Estimation results for the class-membership model

<table>
<thead>
<tr>
<th>Parameters</th>
<th>LCCM 1</th>
<th>LCCM 2</th>
<th>LCCM 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>estimate</td>
<td>t-test</td>
<td>estimate</td>
</tr>
<tr>
<td>$ASC_{class}$</td>
<td>-0.215</td>
<td>-0.86**</td>
<td>-0.629</td>
</tr>
<tr>
<td>$\gamma_{family}$</td>
<td>0.136</td>
<td>0.51**</td>
<td>3.92</td>
</tr>
<tr>
<td>$\gamma_{income}$</td>
<td>0.693</td>
<td>2.76</td>
<td>0.460</td>
</tr>
<tr>
<td>$\gamma_{single}$</td>
<td>0.408</td>
<td>1.34**</td>
<td>0.704</td>
</tr>
</tbody>
</table>

- **Increase of the significance** of the parameters of the latent class model.
MODE CHOICE CASE STUDY

Estimation results for the class-membership model

<table>
<thead>
<tr>
<th>Parameters</th>
<th>LCCM 1</th>
<th>LCCM 2</th>
<th>LCCM 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>estimate</td>
<td>t-test</td>
<td>estimate</td>
</tr>
<tr>
<td>$ASC_{class}$</td>
<td>-0.215</td>
<td>-0.86**</td>
<td>-0.629</td>
</tr>
<tr>
<td>$\gamma_{family}$</td>
<td>0.136</td>
<td>0.51**</td>
<td>3.92</td>
</tr>
<tr>
<td>$\gamma_{income}$</td>
<td>0.693</td>
<td>2.76</td>
<td>0.460</td>
</tr>
<tr>
<td>$\gamma_{single}$</td>
<td>0.408</td>
<td>1.34**</td>
<td>0.704</td>
</tr>
</tbody>
</table>

- Increase of the significance of the parameters of the latent class model.
- **Income** parameter has become more important.
Model application: computation of VOT

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>LCCM 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class independent</td>
<td>3.06</td>
<td></td>
<td>3.72</td>
<td></td>
</tr>
<tr>
<td>Class dependent</td>
<td>52.63</td>
<td></td>
<td>17.53</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>28.97</td>
<td></td>
<td>10.94</td>
<td></td>
</tr>
<tr>
<td>LCCM 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class independent</td>
<td>35.78</td>
<td></td>
<td>15.38</td>
<td></td>
</tr>
<tr>
<td>Class dependent</td>
<td>22.05</td>
<td></td>
<td>8.84</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>29.53</td>
<td></td>
<td>12.40</td>
<td></td>
</tr>
<tr>
<td>LCCM 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class independent</td>
<td>63.27</td>
<td></td>
<td>16.21</td>
<td></td>
</tr>
<tr>
<td>Class dependent</td>
<td>34.16</td>
<td></td>
<td>5.99</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>36.94</td>
<td></td>
<td>18.40</td>
<td></td>
</tr>
</tbody>
</table>

- 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!