On estimation of Hybrid Choice Models

Denis Bolduc and Ricardo Alvarez-Daziano
Département d’économique, Université Laval

May 11, 2010
Objective

- Improve **standard discrete choice** models by including attitudes and perceptions
- Enhance on the behavioral representation of the decision process
- Solve econometric challenges in Discrete Choice Modeling
  - Ex: Tackle the problem of measurement errors in variables in a natural way

**The Hybrid Choice Model (HCM)**

It explicitly incorporates psychological factors affecting decision making, with the goal of enhancing the behavioral representation of the choice process.

Leads to more flexibility and realism.
The Standard Discrete Choice Model

$$U_n = X_n\beta + \nu_n,$$

$$y_{in} = \begin{cases} 
1 & \text{if } U_{in} \geq U_{jn}, \forall j \in C_n, j \neq i \\
0 & \text{otherwise.}
\end{cases}$$

- $U_n$ vector of utilities
- $C_n$ the choice set of agent $n$
- $X_n$ a vector of explanatory (attributes) variables
- $\beta$ vector of coefficients
- $\nu_n$ the error term

- Hypothesis: Choice decisions are affected not only by the attributes but also by an unobservable construct
- Unobservable construct: \( Z \)

\[
U_n = X_n \beta + \Gamma \cdot Z_n + \nu_n
\]

\[
Z_n = w_n b + \zeta_n
\]

\[
I_n = \alpha + \Lambda \cdot Z_n + \varepsilon_n
\]

- \( Z \) is explained by the characteristics of the agent
- \( Z \) is identified through the agent perceptions and attitudes toward the issue
HCM: Latent Variables and Discrete Choice

The basic HCM

Integrated Choice and Latent Variable (ICLV) model: improved representation model of the choice process that involves dealing with unobserved (latent) psychometric variables
The Equations

**Structural equations**

\[
\begin{align*}
    z_n^* &= \Pi z_n^* + Bw_n + \zeta_n = (I_L - \Pi)^{-1}Bw_n + (I_L - \Pi)^{-1}\zeta_n, \\
    \zeta_n &\sim N(0, \Psi) \\
    U_n &= X_n\beta + Cz_n^* + \nu_n \\
    I_n^* &= \alpha + \Lambda z_n^* + \varepsilon_n, \quad \varepsilon_n \sim N(0, \Theta)
\end{align*}
\]  

(1)

Latent variables $z_n^*$, $I_n^*$, and $U_n$ are unobservable: indicators

**Measurement equations**

\[
I_{rn} = \left\{ \begin{array}{ll}
1 & \text{if } \gamma_0 < I_{rn}^* \leq \gamma_1 \\
2 & \text{if } \gamma_1 < I_{rn}^* \leq \gamma_2 \\
\vdots & \\
Q & \text{if } \gamma_{Q-1} < I_{rn}^* \leq \gamma_Q
\end{array} \right.
\]

(4)

\[
y_{in} = \left\{ \begin{array}{ll}
1 & \text{if } U_{in} \geq U_{jn}, \forall j \in C_n, j \neq i \\
0 & \text{otherwise}
\end{array} \right.
\]

(5)
The choice kernel $P_n(i \mid z_n^*, X_n, \theta)$

- HCM classical full information estimation requires the evaluation of the joint probability $P(y_{in} = 1, I_n) \equiv P_n(i, l)$

$$P_n(i, l \mid X_n, w_n, \delta) = \int_{z_n^*} P_n(i \mid z_n^*, X_n, \theta) f(l_n \mid z_n^*, \Lambda, \Theta) g(z_n^* \mid w_n, B, \Pi, \Psi) dz_n^* \quad (6)$$

- This joint probability depends, first, on the discrete choice kernel $P_n(i \mid z_n^*, X_n, \theta)$
- Analytical form of the discrete choice kernel: standard discrete choice modeling (MNL, Probit, Logit Mixture)
- Importance of the assumptions regarding the distribution of the random term $\nu_n$ of the choice model
The Logit Mixture kernel

- We will decompose $\nu_n$ assuming a Normal distributed factor analytic structure $\nu_n = PT\xi_n + \nu_n \Rightarrow$

\[
P_n(i, l|X_n, w_n, \delta) = \int \int P_n(i|z_n^*, X_n, \theta, \xi_n) f(l_n|z_n^*, \Lambda, \Theta) g(z_n^*|w_n, B, \Pi, \Psi) N_\xi dz_n^* d\xi_n
\]

(7)

- $\nu$ i.i.d. extreme value of type 1 $\Rightarrow$

\[
P_n(i|z_n^*, X_n, \theta, \xi_n) = \frac{\exp(X_{in}\beta + C_i z_n^* + P_i T\xi_n)}{\sum_{j \in C_n} \exp(X_{jn}\beta + C_j z_n^* + P_j T\xi_n)}
\]

(8)

Classical estimation

The Logit Mixture kernel is the most convenient assumption to model flexible error structures
The latent variable distributions

- Measurement model: each equation that links the indicators and the latent variables corresponds to a continuous, a binary, or a multinomial ordered response

\[ f(I_n|z^*_n, \Lambda, \Theta) = \prod_{r=1}^{R} f(I_{rn}) \] (9)

- For example, if measurement equation \( r \) is continuous, then

\[ f(I_{rn}) = \frac{1}{\theta_r} \phi \left( \frac{I_{rn} - \alpha_r - \Lambda_r z^*_n}{\theta_r} \right) \] (10)

- Structural equation:

\[ g(z^*_n | w_n, B, \Pi, \Psi) \sim \text{MVN}((I_L - \Pi)^{-1}Bw_n, [(I_L - \Pi)^{-1}]\Psi[(I_L - \Pi)^{-1}]') \] (11)
The likelihood equation

- Maximum log-likelihood problem:

\[
\max_{\delta} L(\delta) = \sum_{n=1}^{N} \sum_{i \in C_n} y_{in} \ln P_n(i, l|X_n, w_n, \delta) 
\] (12)

Oops! How do we solve this?

Calculating the joint probability \( P_n(i, l|X_n, w_n, \delta) \) requires the evaluation of complex multidimensional integrals – the number of latent variables has an impact on the computation of this probability

- In practice, with a large number of latent variables (more than 3), we need to replace the multidimensional integral with a smooth simulator
The simulated log-likelihood solution

- Taking advantage of the expectation form, we can replace the probability with the following empirical mean:

\[
\tilde{P}_n(i, l | X_n, w_n, \delta) = \frac{1}{S} \sum_{s=1}^{S} \frac{\exp(X_{in}\beta + C_i z_{n}^{*s} + P_i T \xi_{n}^{s})}{\sum_{j \in C_n} \exp(X_{jn}\beta + C_j z_{n}^{*s} + P_j T \xi_{n}^{s})} f(l_n | z_n^{*s}, \Lambda, \Theta)
\]

(13)

- This simulator is known to be unbiased, consistent and smooth with respect to the unknown parameters
- Replacing \( P_n(i, l | X_n, w_n, \delta) \) with \( \tilde{P}_n(i, l | X_n, w_n, \delta) \) in the log likelihood leads to a maximum simulated likelihood (MSL) solution
The Survey

Personal vehicle choice data - Simon Fraser University 2003

Virtual personal vehicle choices made by Canadian consumers when faced with technological innovations – 866 completed surveys

- **Part 1:** Transportation options, requirements and habits
- **Part 2:** Personal vehicle choice (stated preferences experiment)
- **Part 3:** Transportation mode preferences
- **Part 4:** Views on transportation issues
- **Part 5:** Additional information (gender, education, income)
Incorporating pro-environmental preferences

Environmentally-conscious consumers are aware of the dangers of climate change and oil dependency:

- Their concerns about the role of transportation in global warming has a consequent change in their consuming behavior
- They are willing to pay more for sustainable solutions (low-emission vehicles) despite potential drawbacks (such as a reduced refueling availability).
- Current demand models have a hard time representing eco-friendly behavior
An HCM for vehicle purchase decisions

- **Key question**: how to incorporate the consumers’ environmental concerns into an economic model for private vehicle purchase decisions
- **DCM research**: current trend to enhance the behavioral representation of the underlying dynamics of the choice process

**Hybrid Choice Models (HCMs)**

HCMs integrate standard discrete choice and latent variables models, taking into account the impact of attitudes and perceptions on the decision process (psychological factors) thus enhancing the behavioral representation of the choice process. It also tackles the problem of measurement error.
The SP Experiment: Choice among four car types

- A conventional vehicle (operating on gasoline or diesel) (SGV)
- A natural-gas vehicle for alternative fuel vehicle (AFV)
- A hybrid vehicle (gasoline-electric) (HEV)
- A hydrogen fuel cell vehicle (HFC)
Characteristics of the vehicles

- **Capital Cost:** purchase price
- **Operating Cost:** fuel cost
- **Fuel Available:** percentage of stations selling the proper fuel type
- **Express Lane Access:** whether or not the vehicle would be granted express lane access
- **Emissions Data:** emissions compared to a standard gasoline vehicle
- **Power:** power compared to their current vehicle
Setting the Latent Variables

- Focus on three different attitudinal-perceptual questions of the survey:
  - Transport Policies Support (TPS)
  - Transport Problems Evaluation (TPE)
  - Car Attributes Importance (CAI)

- Considering the answers to these questions as indicators we identify as latent variables:
  - Environmental Concern (EC): related to transportation and its environmental impact
  - Appreciation of new car features (ACF): related to car purchase decisions and how important are the characteristics of this new alternative
Tackling the problem of measurement errors

Our model also includes a third latent variable, the **latent income variable** (REV), to account for the measurement error problem in quantifying the income variable.
The private vehicle HCM
# Car Choice Model: Estimations

<table>
<thead>
<tr>
<th>Hybrid Choice Model</th>
<th>MNL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>estimates</strong></td>
<td><strong>t-stat</strong></td>
</tr>
<tr>
<td>ASC_AFV</td>
<td>-6.626</td>
</tr>
<tr>
<td>ASC_HEV</td>
<td>-4.383</td>
</tr>
<tr>
<td>ASC_HFC</td>
<td>-6.403</td>
</tr>
<tr>
<td>Capital Cost</td>
<td>-0.943</td>
</tr>
<tr>
<td>Operating Cost</td>
<td>-0.849</td>
</tr>
<tr>
<td>Fuel available</td>
<td>1.384</td>
</tr>
<tr>
<td>Expess lane access</td>
<td>0.162</td>
</tr>
<tr>
<td>Power</td>
<td>2.710</td>
</tr>
</tbody>
</table>

## Latent Variables

- ACF on SGV: 3.160, 25.984
- EC on AFV: 0.798, 2.695
- ACF on AFV: 2.810, 30.708
- EC on HEV: 0.770, 3.965
- ACF on HEV: 2.810, 30.708
- EC on HFC: 1.085, 5.620
- ACF on HFC: 3.054, 31.048
- REV on HFC: 0.456, 3.373
### Structural Model: Equations

<table>
<thead>
<tr>
<th></th>
<th>EC</th>
<th></th>
<th>ACF</th>
<th></th>
<th>REV</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>est</td>
<td>t-stat</td>
<td>est</td>
<td>t-stat</td>
<td>est</td>
<td>t-stat</td>
</tr>
<tr>
<td>Intercept</td>
<td>2.434</td>
<td>9.660</td>
<td>-3.094</td>
<td>-18.495</td>
<td>1.293</td>
<td>7.374</td>
</tr>
<tr>
<td>Driving Alone User</td>
<td>-0.020</td>
<td>-0.428</td>
<td>-0.118</td>
<td>-2.244</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Car Pool User</td>
<td>0.097</td>
<td>1.135</td>
<td>0.100</td>
<td>0.938</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Transit User</td>
<td>0.204</td>
<td>2.572</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Female Dummy</td>
<td>0.258</td>
<td>7.392</td>
<td>0.283</td>
<td>6.474</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>High Income Dummy (&gt;80K$)</td>
<td>-0.011</td>
<td>-0.294</td>
<td>-0.001</td>
<td>-0.017</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Education: University</td>
<td>0.064</td>
<td>1.712</td>
<td>0.008</td>
<td>0.170</td>
<td>0.598</td>
<td>9.522</td>
</tr>
<tr>
<td>Age level: 26-40 years</td>
<td>0.187</td>
<td>2.279</td>
<td>0.328</td>
<td>3.961</td>
<td>0.592</td>
<td>5.363</td>
</tr>
<tr>
<td>Age level: 41-55 years</td>
<td>0.262</td>
<td>3.105</td>
<td>0.621</td>
<td>6.657</td>
<td>0.889</td>
<td>7.952</td>
</tr>
<tr>
<td>Age level: 56 years &amp; more</td>
<td>0.332</td>
<td>3.702</td>
<td>0.525</td>
<td>5.278</td>
<td>0.228</td>
<td>1.840</td>
</tr>
</tbody>
</table>
### Measurement Model: Estimations

<table>
<thead>
<tr>
<th>Transport Policies Support</th>
<th>estimates</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC on Expanding &amp; Upgrading Roads</td>
<td>-0.392</td>
<td>-5.405</td>
</tr>
<tr>
<td>EC on Road Tolls &amp; Gas Taxes</td>
<td>0.581</td>
<td>6.600</td>
</tr>
<tr>
<td>ACF on Road Tolls &amp; Gas Taxes</td>
<td>-0.091</td>
<td>-1.389</td>
</tr>
<tr>
<td>EC on Bike Lanes &amp; Speed Controls</td>
<td>0.532</td>
<td>8.507</td>
</tr>
<tr>
<td>EC on Reducing Car Emissions</td>
<td>0.478</td>
<td>7.944</td>
</tr>
<tr>
<td>ACF on Reducing Car Emissions</td>
<td>0.295</td>
<td>7.856</td>
</tr>
<tr>
<td>EC on High Occupancy Vehicles &amp; Transit Priorities</td>
<td>0.606</td>
<td>8.143</td>
</tr>
<tr>
<td>EC on Improving Transit Service</td>
<td>0.491</td>
<td>8.352</td>
</tr>
<tr>
<td>EC on Promoting Compact Communities</td>
<td>0.206</td>
<td>2.994</td>
</tr>
<tr>
<td>EC on Encouraging Short Work Weeks</td>
<td>0.396</td>
<td>7.159</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transport Problems Evaluation</th>
<th>estimates</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC on Traffic Congestion</td>
<td>0.735</td>
<td>9.154</td>
</tr>
<tr>
<td>EC on Traffic Noise</td>
<td>0.901</td>
<td>9.495</td>
</tr>
<tr>
<td>EC on Poor Local Air Quality</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>ACF on Poor Local Air Quality</td>
<td>-0.061</td>
<td>-1.416</td>
</tr>
<tr>
<td>EC on Accidents Caused by Bad Drivers</td>
<td>0.837</td>
<td>14.472</td>
</tr>
<tr>
<td>EC on Emissions &amp; Global Warming</td>
<td>1.113</td>
<td>16.200</td>
</tr>
<tr>
<td>EC on Speeding Drivers in Neighborhoods</td>
<td>1.107</td>
<td>15.790</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Car Attributes Importance</th>
<th>estimates</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACF on Purchase Price</td>
<td>-0.004</td>
<td>-0.087</td>
</tr>
<tr>
<td>ACF on Fuel Economy Importance</td>
<td>0.259</td>
<td>5.008</td>
</tr>
<tr>
<td>ACF on Horsepower Importance</td>
<td>0.433</td>
<td>7.220</td>
</tr>
<tr>
<td>ACF on Safety Importance</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>ACF on Seating Capacity Importance</td>
<td>0.684</td>
<td>11.563</td>
</tr>
<tr>
<td>ACF on Reliability Importance</td>
<td>0.537</td>
<td>16.241</td>
</tr>
<tr>
<td>ACF on Styling</td>
<td>0.371</td>
<td>6.660</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Income Class</th>
<th>estimates</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>REV on rev</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

![Diagram showing priorities and impacts](image)
Outcome from this work?

- Discrete choice modeling has evolved towards an explicit incorporation of psychological factors affecting decision making.

- Hybrid Choice Models ⇒ more realistic: perceptions and attitudes are now incorporated.

- Using real data, we showed that HCM is genuinely capable of adapting to practical situations:
  - We identified two travel related dimensions with a significant impact: environmental concern and appreciation of new car features.
  - We included in the choice model a latent income variable to account for the measurement errors associated with the reported income classes.
What’s next?

- MSL real data application: Analyzing the results of a prediction and simulation exercise
- Compare the predictions of a hybrid choice model versus those obtained from a conventional choice model
- Bayesian estimation: Implementation of the HCM Gibbs sampler
- Testing the general HCM Gibbs sampler to determine when Bayesian MCMC outperforms MSL according to empirical results based on correct identification restrictions and accurate predictions