Experimental analysis of the implicit choice set generation using the Constrained Multinomial Logit model

Michel Bierlaire, Ricardo Hurtubia and Gunnar Flötteröd

Transport and Mobility Laboratory
Introduction

Choice model: \( P_n(i|C_n) \)

- Common practice: \( C_n \) characterized by deterministic rules
- Modeling the choice set generation (Manski, 1977):
  \[
P_n(i) = \sum_{C_m \subseteq C} P_n(i|C_m) P_n(C_m)
\]
- Combinatorial complexity
- Operational instances:
  - Random constraints (Swait and Ben-Akiva, 1987, Ben-Akiva and Boccara, 1995)
  - MEV framework (Swait, 2001)
Introduction

- Heuristics:
  - Implicit Availability/Perception model (Cascetta and Papola, 2001)
  - Constrained Multinomial Logit model (Martinez et al., 2009)

Objective: analyze the quality of the CMNL as a choice set generation process.
Deterministic Choice Set Generation

- Availability conditions
- Exogenous variables

\[ A_{in} = \begin{cases} 
1 & \text{if alternative } i \text{ is considered by individual } n, \\
0 & \text{otherwise.} 
\end{cases} \]

- Choice model

\[ P_n(i|C_n) = \Pr (U_{in} \geq U_{jn}, \forall j \in C_n) \]
\[ = \Pr (U_{in} + \ln A_{in} \geq U_{jn} + \ln A_{jn}, \forall j \in C). \]

- Note: a choice model with deterministic choice set generation can always be written in terms of the universal choice set
Deterministic Choice Set Generation

Logit model:

\[ P_n(i) = \frac{e^{V_{in} + \ln A_{in}}}{\sum_{j \in C} e^{V_{jn} + \ln A_{jn}}} = \frac{A_{in} e^{V_{in}}}{\sum_{j \in C} A_{jn} e^{V_{jn}}} \]

What if variables \( A_{in} \) are not exogenously given?
Probabilistic Choice Set Generation

Approaches:

- Correct model: Manski (1977) most of the time impractical
- Sampling of alternatives:
  - Assume $C_n = C, \forall n$
  - Sample a subset for estimation
  - see Frejinger, Bierlaire and Ben-Akiva (forthcoming) for route choice
- Replace $A_{in}$ by a probability distribution
  - Availability/Perception (Cascetta and Papola, 2001)
  - Cutoffs (Martinez et al., 2009)
Cutoffs

Optimization problem of rational consumer $n$:

$$\max_{\delta_{ni}} \sum_{i \in C} \delta_{ni} U_{in}(X_i)$$

subject to

$$\sum_{i \in C} \delta_{ni} = 1, \quad \delta_{ni} \in \{0, 1\}, \forall i \in C$$

But attributes are meaningful only within some bounds

$$\ell_{nk} \leq X_{ik} \leq u_{nk} \quad \forall i \in C, \forall k$$

An alternative $i$ with one of its attributes is out of bounds is not considered
Cutoffs

Examples:

- Item too expensive
- Traveling by train involves a too long walking distance to the station
- etc.

If these rules are deterministic, the variables $A_{in}$ can be derived. If not, what can be done?
Cutoffs

Idea: relax the constraint in a probabilistic way
Example: constraint $\ell \leq X$

\[
\begin{align*}
V_{\text{not considered}} &= \ell + \varepsilon_1 \\
V_{\text{considered}} &= X + \varepsilon_2
\end{align*}
\]

\[
P(\text{considered}) = \frac{e^{\rho X}}{e^{\rho X} + e^{\rho \ell}} = \frac{1}{1 + e^{\rho (\ell - X)}}
\]

Example: constraint $X \leq u$

\[
P(\text{considered}) = \frac{e^{-\rho X}}{e^{-\rho X} + e^{-\rho u}} = \frac{1}{1 + e^{\rho (X - u)}}
\]
Example: $2 \leq X$
Cutoffs

Example: $X \leq 4$
Cutoffs

Constraint $\ell \leq X \leq u$

\[ P(\text{considered}) = \frac{1}{1 + e^{\rho(\ell - X)}} \cdot \frac{1}{1 + e^{\rho(X - u)}} \]

We denote this quantity by $\phi_n(X)$
Cutoffs

Example: $2 \leq X \leq 4$
Cutoffs

The utility function now becomes

\[ V_i = \sum_k \beta_k X_{ik} + \sum_{k^*} \frac{1}{\rho} \ln \phi_n(X_{ik^*}) \]

where \( k^* \) ranges only on constrained attributes. Note that

\[
\ln \phi(X) = -\ln(1 + e^{\rho(\ell - X)}) - \ln(1 + e^{\rho(X - u)})
\]

\[ = -\ln(1 + e^{\rho \ell} e^{-\rho X}) - \ln(1 + e^{\rho X} e^{-\rho u}) \]

Can be estimated, although it is difficult.
Comparison of CMNL and Manski

Simple example:

- Binary logit: $C = \{1, 2\}$
- Alternative 1 is always available
- Alternative 2 is considered with probability $\phi_2$

We have

- $P(C_n = \{1\}) = 1 - \phi_2$
- $P(C_n = \{2\}) = 0$
- $P(C_n = \{1, 2\}) = \phi_2$
Comparison of CMNL and Manski

Manski’s model

![Diagram showing choice sets and alternatives connected to a root node.]
Comparison of CMNL and Manski

Manski’s model

\[
P(1) = P(C_n = \{1\}) \frac{e^{V_1}}{e^{V_1} + e^{V_2}} + P(C_n = \{2\})0 + P(C_n = \{1, 2\}) \frac{e^{V_1}}{e^{V_1} + e^{V_2}} \\
= (1 - \phi_2) + \phi_2 \frac{e^{V_1}}{e^{V_1} + e^{V_2}}
\]

CMNL model

\[
P(1) = \frac{e^{V_1}}{e^{V_1} + e^{V_2} + \ln \phi_2}.
\]

Note: for given \(V\)’s, Manski is linear in \(\phi_2\), not CMNL
Comparison of CMNL and Manski

Equal utility

\[ V_1 = V_2 \]

\[ P_1 \]

\[ \phi_2 \]
Comparison of CMNL and Manski

Alt. 1 is dominant
Comparison of CMNL and Manski

Alt. 2 is dominant

![Graph showing comparison between CMNL and Manski models. The graph illustrates the difference in choice probabilities between the two models for different values of $V_2 - V_1 = 2$. The CMNL model is represented by a red line, and the Manski model by a green dotted line. The y-axis represents $P_1$, and the x-axis represents $\phi_2$. The graph demonstrates that Alt. 2 is dominant in terms of choice probabilities.]
Comparison of CMNL and Manski

Alt. 2 is even more dominant

![Graph showing comparison between CMNL and Manski models with the difference V_2 - V_1 = 4]
Comparison of CMNL and Manski

- CMNL underestimates the choice probability for alternative 1
- When alt. 1 is dominant, it makes no difference if it is preferred because of a high utility, or if because 2 is not even considered.
- When alt. 2 is dominant, the CMNL may be completely off
- Clearly, the model parameters could be adjusted to attenuate that error
Synthetic data

- Swissmetro data set, 5607 observations
  1. Driving a car (CAR)
  2. Regular train (TRAIN)
  3. Swissmetro, the future high speed train (SM)
- Exogenous variables come from the data set
- Synthetic choice set
  - TRAIN and SM always available
  - CAR available depending on travel time

\[ \phi_{\text{CAR}} = \frac{1}{1 + \exp(\omega (TT_{\text{CAR}}/60 - a))} \]

- Synthetic choice
## Synthetic data

### Postulated model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Car</th>
<th>Train</th>
<th>Swissmetro</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{ASC}_{\text{CAR}}$</td>
<td>0.3</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\text{ASC}_{\text{SM}}$</td>
<td>0.4</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$\beta_{\text{cost}}$</td>
<td>-0.001</td>
<td>Cost (CHF)</td>
<td>Cost (CHF)</td>
<td>Cost (CHF)</td>
</tr>
<tr>
<td>$\beta_{tt}$</td>
<td>-0.001</td>
<td>In veh. travel time (minutes)</td>
<td>In veh. travel time (minutes)</td>
<td>In veh. travel time (minutes)</td>
</tr>
<tr>
<td>$\beta_{he}$</td>
<td>-0.005</td>
<td>0</td>
<td>Headway (minutes)</td>
<td>Headway (minutes)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>3</td>
<td></td>
<td>Consideration threshold of car (hours)</td>
<td></td>
</tr>
<tr>
<td>$\omega$</td>
<td>1,2,3,5,10</td>
<td></td>
<td>Consideration dispersion of car</td>
<td></td>
</tr>
</tbody>
</table>
Synthetic data

- 100 choice data sets are simulated for each value of $\omega$
- Results:
  - mean of each parameter over 100 estimations
  - $t$-test against the true value, based on the empirical std. deviation.
### Estimation results for Manski's model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>real $\omega$ value</th>
<th>1 real value</th>
<th>1 estimate</th>
<th>1 t-test</th>
<th>2 real value</th>
<th>2 estimate</th>
<th>2 t-test</th>
<th>3 real value</th>
<th>3 estimate</th>
<th>3 t-test</th>
<th>5 real value</th>
<th>5 estimate</th>
<th>5 t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ASC_{\text{CAR}}$</td>
<td>0.3</td>
<td>0.304</td>
<td>0.027</td>
<td>0.288</td>
<td>0.113</td>
<td>0.300</td>
<td>0.010</td>
<td>0.301</td>
<td>0.012</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$ASC_{\text{SM}}$</td>
<td>0.4</td>
<td>0.396</td>
<td>0.044</td>
<td>0.399</td>
<td>0.010</td>
<td>0.405</td>
<td>0.053</td>
<td>0.401</td>
<td>0.017</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{\text{cost}}$</td>
<td>-0.01</td>
<td>-0.010</td>
<td>0.283</td>
<td>-0.010</td>
<td>0.001</td>
<td>-0.010</td>
<td>0.179</td>
<td>-0.010</td>
<td>0.052</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{\text{he}}$</td>
<td>-0.005</td>
<td>-0.005</td>
<td>0.241</td>
<td>-0.005</td>
<td>0.010</td>
<td>-0.005</td>
<td>0.048</td>
<td>-0.005</td>
<td>0.082</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{\text{time}}$</td>
<td>-0.01</td>
<td>-0.01</td>
<td>0.074</td>
<td>-0.010</td>
<td>0.050</td>
<td>-0.010</td>
<td>0.049</td>
<td>-0.010</td>
<td>0.003</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a$</td>
<td>3</td>
<td>2.963</td>
<td>0.019</td>
<td>3.008</td>
<td>0.118</td>
<td>3.000</td>
<td>0.100</td>
<td>2.998</td>
<td>0.081</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\omega$</td>
<td>see top</td>
<td>1.003</td>
<td>0.028</td>
<td>2.014</td>
<td>0.079</td>
<td>3.066</td>
<td>0.210</td>
<td>5.095</td>
<td>0.170</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Estimation results for CMNL model

<table>
<thead>
<tr>
<th>parameter</th>
<th>real $\omega$ value</th>
<th>1</th>
<th></th>
<th>2</th>
<th></th>
<th>3</th>
<th></th>
<th>5</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ASC_{CAR}</td>
<td>0.3</td>
<td>0.503</td>
<td>0.950</td>
<td>0.421</td>
<td>1.153</td>
<td>0.406</td>
<td>1.365</td>
<td>0.380</td>
<td>0.988</td>
</tr>
<tr>
<td>ASC_{SM}</td>
<td>0.4</td>
<td>0.565</td>
<td>2.013 *</td>
<td>0.550</td>
<td>2.375 *</td>
<td>0.536</td>
<td>1.804</td>
<td>0.506</td>
<td>1.485</td>
</tr>
<tr>
<td>$\beta_{cost}$</td>
<td>-0.01</td>
<td>-0.008</td>
<td>4.825 *</td>
<td>-0.008</td>
<td>3.580 *</td>
<td>-0.009</td>
<td>2.309 *</td>
<td>-0.009</td>
<td>1.182</td>
</tr>
<tr>
<td>$\beta_{he}$</td>
<td>-0.005</td>
<td>-0.005</td>
<td>0.202</td>
<td>-0.005</td>
<td>0.151</td>
<td>-0.005</td>
<td>0.071</td>
<td>-0.005</td>
<td>0.120</td>
</tr>
<tr>
<td>$\beta_{time}$</td>
<td>-0.01</td>
<td>-0.007</td>
<td>3.929 *</td>
<td>-0.008</td>
<td>3.645 *</td>
<td>-0.008</td>
<td>2.813 *</td>
<td>-0.009</td>
<td>2.316 *</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>3</td>
<td>2.186</td>
<td>1.753</td>
<td>2.656</td>
<td>3.073 *</td>
<td>2.773</td>
<td>3.762 *</td>
<td>-2.869</td>
<td>3.305 *</td>
</tr>
<tr>
<td>$\omega$</td>
<td>see top</td>
<td>1.043</td>
<td>0.239</td>
<td>2.094</td>
<td>0.403</td>
<td>3.118</td>
<td>0.431</td>
<td>5.238</td>
<td>0.424</td>
</tr>
</tbody>
</table>

(* indicates an insignificant parameter)
Synthetic data

- Manski model performs well, as expected
- CMNL may significantly bias the estimates
- The more deterministic the constraint, the better the CMNL
Conclusion

- CMNL is not adequate to model the choice set generation
- It is a model on its own, derived from semi-compensatory arguments
- Its complexity is linear in the number of alternatives, while Manski’s model is exponential.
- Research question: how can we modify the CMNL to be a better approximation of Manski’s model.