Advanced discrete choice models with applications to transport demand

Stephane Hess
Imperial College London

The Expanding Sphere of Travel Behaviour Research
11\textsuperscript{th} International Conference on Travel Behaviour Research
Kyoto, August 16\textsuperscript{th} - 20\textsuperscript{th} 2006
Topics

- Model specification
- Model estimation
- Interpretation of results
- Use of advanced models in practice
Topics

- Model specification
- Model estimation ⇒ strongly related
- Interpretation of results
- Use of advanced models in practice
Model specification I: correlation structure

- Standard models assume covariance homogeneity

- Mixed covariance structures

\[ P_n(i) = \int_{\lambda} P_n(i | \lambda) f(\lambda | \Omega) d\lambda \]

- Allowing for covariance heterogeneity
  - Improvements in model fit
  - Differences in forecasts
  - Differences in trade-offs
## Mixed Covariance example

<table>
<thead>
<tr>
<th></th>
<th>MNL</th>
<th>NL</th>
<th>MCOV(_U)</th>
<th>MCOV(_N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null LL</td>
<td>-3517.56</td>
<td>-3517.56</td>
<td>-3517.56</td>
<td>-3517.56</td>
</tr>
<tr>
<td>Final LL</td>
<td>-2329.07</td>
<td>-2320.23</td>
<td>-2191.42</td>
<td>-2189.8</td>
</tr>
<tr>
<td>Par</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>Adj.(\rho^2)</td>
<td>0.3339</td>
<td>0.3361</td>
<td>0.3725</td>
<td>0.3729</td>
</tr>
<tr>
<td>VTTS(_{car})</td>
<td>54.64</td>
<td>46.03</td>
<td>42.43</td>
<td>41.96</td>
</tr>
<tr>
<td>VTTS(_{PT})</td>
<td>23.64</td>
<td>27.25</td>
<td>30.13</td>
<td>30.94</td>
</tr>
</tbody>
</table>

## Changes in market shares after 20% inc. in base TT

<table>
<thead>
<tr>
<th></th>
<th>MNL</th>
<th>NL</th>
<th>MCOV(_U)</th>
<th>MCOV(_N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>base alt.</td>
<td>-13.33%</td>
<td>-18.54%</td>
<td>-17.96%</td>
<td>-18.51%</td>
</tr>
<tr>
<td>early dep.</td>
<td>30.72%</td>
<td>46.17%</td>
<td>51.01%</td>
<td>53.36%</td>
</tr>
<tr>
<td>late dep.</td>
<td>31.27%</td>
<td>47.51%</td>
<td>39.22%</td>
<td>40.39%</td>
</tr>
<tr>
<td>PT</td>
<td>26.27%</td>
<td>17.81%</td>
<td>15.62%</td>
<td>14.93%</td>
</tr>
</tbody>
</table>
Model specification II: taste heterogeneity

- Continuous vs discrete mixtures

\[ P(i) = \int_{\beta} [P(i | \beta) f(\beta, \Omega)] \, d\beta \]

- Specification issue: choice of distributions

\[ P(i) = \sum_{j_1=1}^{m_1} \cdots \sum_{j_K=1}^{m_K} P\left(i \mid \beta = \langle \hat{\beta}^1_{j_1}, \ldots, \hat{\beta}^K_{j_K} \rangle \right) \pi^1_{j_1} \cdots \pi^K_{j_K} \]

- Specification issue: selection of number of support points
### Taste heterogeneity example I

<table>
<thead>
<tr>
<th>Model</th>
<th>MNL</th>
<th>MMNL(N)</th>
<th>DM(2)</th>
<th>DM(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respondents:</td>
<td>185</td>
<td>185</td>
<td>185</td>
<td>185</td>
</tr>
<tr>
<td>Observations:</td>
<td>1,421</td>
<td>1,421</td>
<td>1,421</td>
<td>1,421</td>
</tr>
<tr>
<td>Final LL:</td>
<td>-880.96</td>
<td>-849.65</td>
<td>-845.40</td>
<td>-844.60</td>
</tr>
<tr>
<td>adj. $\rho^2$:</td>
<td>0.1036</td>
<td>0.1343</td>
<td>0.1366</td>
<td>0.1354</td>
</tr>
<tr>
<td>Estimation time (s):</td>
<td>1</td>
<td>75</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Mean VTTS (DKK/hour):</td>
<td>19.77</td>
<td>30.41</td>
<td>32.81</td>
<td>34.29</td>
</tr>
<tr>
<td>VTTS standard deviation</td>
<td>-</td>
<td>33.70</td>
<td>36.55</td>
<td>41.86</td>
</tr>
</tbody>
</table>
Taste heterogeneity example II

CDF for $\beta_T$

- True distribution
- MMNL
- DM(6)
Model specification III: joint structures

- Correlation & random taste heterogeneity
  - phenomena taking place in unobserved part of utility

- Both processes can act at the same time
  - Solution:
    - use GEV mixture models, or joint RCL-ECL formulation

- Majority of applications account only for one of two phenomena
  - Big risk of confounding
    * wrongly retrieved taste heterogeneity: issues in CBA, etc
    * wrongly retrieved correlation: issues in forecasting
<table>
<thead>
<tr>
<th>True model</th>
<th>MNL</th>
<th>NL (rail-SM)</th>
<th>NL (rail-car)</th>
<th>RCL</th>
<th>NL mixture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final LL</td>
<td>-1562.29</td>
<td>-1560.55</td>
<td>-1525.55</td>
<td>-1163.73</td>
<td>-1147.78</td>
</tr>
<tr>
<td>adj. $\rho^2(0)$</td>
<td>0.5239</td>
<td>0.5241</td>
<td>0.5347</td>
<td>0.6433</td>
<td>0.6478</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>est. t-stat.</th>
<th>est. t-stat.</th>
<th>est. t-stat.</th>
<th>est. t-stat.</th>
<th>est. t-stat.</th>
<th>est. t-stat.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{TC} (\mu)$</td>
<td>-0.100</td>
<td>-0.035</td>
<td>-0.035</td>
<td>-0.028</td>
<td>-0.152</td>
<td>-0.093</td>
</tr>
<tr>
<td>$\beta_{TC} (\sigma)$</td>
<td>0.035</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.061</td>
</tr>
<tr>
<td>$\beta_{HW} (\mu)$</td>
<td>-0.020</td>
<td>-0.023</td>
<td>-0.022</td>
<td>-0.014</td>
<td>-0.034</td>
<td>-0.018</td>
</tr>
<tr>
<td>$\beta_{HW} (\sigma)$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.009</td>
</tr>
<tr>
<td>$\beta_{TT,car} (\mu)$</td>
<td>-0.030</td>
<td>-0.021</td>
<td>-0.020</td>
<td>-0.015</td>
<td>-0.054</td>
<td>-0.028</td>
</tr>
<tr>
<td>$\beta_{TT,car} (\sigma)$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.016</td>
</tr>
<tr>
<td>$\beta_{TT,rail} (\mu)$</td>
<td>-0.040</td>
<td>-0.032</td>
<td>-0.032</td>
<td>-0.022</td>
<td>-0.065</td>
<td>-0.037</td>
</tr>
<tr>
<td>$\beta_{TT,rail} (\sigma)$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.002</td>
</tr>
<tr>
<td>$\beta_{TT,SM} (\mu)$</td>
<td>-0.035</td>
<td>-0.022</td>
<td>-0.022</td>
<td>-0.013</td>
<td>-0.059</td>
<td>-0.034</td>
</tr>
<tr>
<td>$\beta_{TT,SM} (\sigma)$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.003</td>
</tr>
<tr>
<td>$\lambda_{rail,SM}$</td>
<td>0.50</td>
<td>1.00</td>
<td>0.89</td>
<td>1.00</td>
<td>1.00</td>
<td>0.48</td>
</tr>
<tr>
<td>$\lambda_{rail,car}$</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.49</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Model estimation I: type of draws

- Simulation-based estimation of continuous mixture models
- Various alternatives to standard PMC draws
- Halton draws:
  - Problems even with only 4 – 5 dimensions
  - Also problems with various adapted versions
- MLHS
  - outperforms Halton sequences, but high variance
- Advanced approaches
  - difficult to implement, lack of available code
  - sometimes no guarantee of better performance
Model estimation II: number of draws

- Number of draws has very significant impacts on results

- Guidelines in literature
  - Worrying number of applications using 125 Halton draws, in up to 8 dimensions!

- Other approaches for reducing estimation cost
  - MNL starting values
  - Pre-estimation with lower number of draws

- Advisable to use multiple runs with different draws and starting values
Model interpretation

- Model specification has effects on model results
  - important to recognise in model interpretation

- Example: choice of distributions in MMNL
  - shape assumptions impact model results

- Application: Danish VOT data
  - compare models with different distributional assumptions
Model interpretation example

$\beta_{TT} \sim \text{Normal}$

$\beta_{TT} \sim \text{Lognormal, with sign-change}$

$\beta_{TT} \sim S_B \text{ (symmetrical)}$

$\beta_{TT} \sim \text{Triangular}$
Model interpretation example (continued)

- Significant differences in implied distributions

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>Lognormal</th>
<th>$S_B$</th>
<th>Triangular</th>
</tr>
</thead>
<tbody>
<tr>
<td>adj. $\rho^2$</td>
<td>0.1484</td>
<td>0.1518</td>
<td>0.1512</td>
<td>0.1486</td>
</tr>
<tr>
<td>% positive</td>
<td>14.53%</td>
<td>0%</td>
<td>0%</td>
<td>14.6%</td>
</tr>
</tbody>
</table>

- Differences in model fit very small

- Guidance: use flexible distributions
Theoretical rules vs practical considerations

- Applied part...

- Modelling air travel choice behaviour

- Joint choice of airport, airline and access mode
  - departure from state-of-the-art
  - major gains in model performance

- Three datasets
  - RP data (SF-bay area & Greater London)
  - SP data (US, internet-based)
Random taste heterogeneity

- Some gains in model fit, and insights into behaviour
- Huge increases in estimation cost
- Use of Normal almost unavoidable
- Use of high number of draws almost impossible
- London data (> 10,000 observations, 324 alternatives)
  - basic MMNL models take in excess of 1 month to estimate
- Flexible interactions and non-linearities massively reduce scope for retrieving random taste heterogeneity
Correlation structure

- Shortcomings of NL model for multi-dimensional choice processes
2-level NL

Combined alternative: airport k, airline l, access-mode m
3-level NL

Root

Airport 1

Airport K

Airline 1

Airline L

Airline 1

Airline L

Combined alternative: airport k, airline l, access-mode m

XXX

XXX

Composed nest

λ

Airport nesting parameter

π

Airline nesting parameter

k l m

Imperial College London

20/26
CNL

Root

Airport 1 --→ Airport K --→ Airline 1 --→ Airline L --→ Access-mode 1 --→ Access-mode M

λ_1, λ_k, π_1, π_L, ψ_1, ψ_M

1 1 1 1 L M 1 1 K 1 L M

λ_x

Airport nesting parameter

π_x

Airline nesting parameter

ψ_x

Access-mode nesting parameter

XXXXX

Composite nest

k | l | m

Elementary alternative: airport k, airline l, access-mode m
Correlation structure

- Shortcomings of NL model for multi-dimensional choice processes
- Need cross-nesting structure

<table>
<thead>
<tr>
<th></th>
<th>Par</th>
<th>Adj. $\rho^2$</th>
<th>est. time</th>
</tr>
</thead>
<tbody>
<tr>
<td>MNL</td>
<td>55</td>
<td>0.3445</td>
<td>minutes</td>
</tr>
<tr>
<td>NL airport</td>
<td>59</td>
<td>0.3465</td>
<td>hours</td>
</tr>
<tr>
<td>NL airline</td>
<td>74</td>
<td>0.3469</td>
<td>hours</td>
</tr>
<tr>
<td>NL access</td>
<td>60</td>
<td>0.3499</td>
<td>hours</td>
</tr>
<tr>
<td>CNL</td>
<td>91</td>
<td>0.3578</td>
<td>weeks</td>
</tr>
</tbody>
</table>
Advanced structures

- GEV mixtures

- Covariance heterogeneity

- Essentially impossible to use
  - Potential confounding and other mis-specification effects
Conclusions

- Major issues in specification, estimation and interpretation
- Certain guidelines
- Often need to be violated in practice
Discussion

- State-of-the-art has moved forward at great speed
- State-of-practice is trailing behind
- Advanced models often almost inapplicable
  - Estimation and application cost
  - Data requirements
- Often rather small gains in performance
  - log-likelihood is not everything
- Mixture models over-hyped
- Need to educate, and sell our models better
Thank you ...