Discrete choice models incorporating revealed preferences and psychometric data

In: Econometric Models in Marketing, 2002

Taka Morikawa, Nagoya Univ.
Moshe Ben-Akiva, MIT
Daniel McFadden, Berkeley
DCM Framework

Product characteristics
Consumer characteristics

Preferences

Market behavior (RP data)

Observed variables
Latent variable
Observed variable
DCM Framework

- **Product characteristics**
- **Consumer characteristics**

**Preferences**

- $X_j$ is a set of product $j$'s characteristics
- $D_i$ is a set of consumer $i$'s characteristics

$$U_{ij} = V_{ij}(X_j, D_i) + \epsilon_{ij}$$

**Market behavior (RP data)**

$$P(y_i = j) = \Gamma(V_i.)$$
DCM Framework

Product characteristics
Consumer characteristics

Preferences

Market behavior (RP data)

$X_j$ is a set of product $j$'s characteristics
$D_i$ is a set of consumer $i$'s characteristics

$U_{ij} = X_j \beta + D_i \gamma + \epsilon_{ij}$
$\epsilon_{ij} \sim i.i.d. \text{EVI}$

$P(d_i = j) = \frac{e^{X_j \beta + D_i \gamma}}{\sum_k e^{X_k \beta + D_i \gamma}}$
DCM Framework

Challenges with RP data:
- Lack of the data about consumers
- Possible multicollinearity in product characteristics
- No data on the actual choice of only hypothetical alternatives:
  - New characteristics
  - New values of present characteristics
- Unknown actual choice set or consideration set
DCM Framework

Product characteristics
Consumer characteristics

Preferences

Opportunities with SP data:
• Preferences for non-existing alternatives or attributes
• The choice set is prespecified
• Multicollinearity is avoided
• Range of attribute values can be extended

Stated Preferences
(SP data)
Challenges with SP data:

- The respondent considers only the most important attribute
- The response is influenced by an ‘inertia’ of the current actual choice
- Respondent uses the survey as an opinion statement for his benefit (overstating)
- Not consider situational constraints
- Ignores or misinterprets an attribute if an attribute value lacks reality
DCM Framework

Product characteristics
Consumer characteristics

Preferences

Structural model:

\[ U_{ij}^{RP} = V_{ij}(X_j, D_i) + \epsilon_{ij}^{RP} \]
\[ U_{ij}^{SP} = \Upsilon_{ij}(X_j, D_i) + \rho d_{ij}^{RP} + \epsilon_{ij}^{SP} \]

\( V \) and \( \Upsilon \) may contain the same parameters and different parts

\( \rho d_{ij}^{RP} \) in SP captures ‘inertia’
DCM Framework

Product characteristics
Consumer characteristics

Situational constraints
Preferences

Market behavior
(RP data)

Stated preferences
(SP data)

Measurement (binary) model:

\[ d_{ij}^{RP} = \begin{cases} 1, & U_{ij}^{RP} \geq 0 \\ 0, & U_{ij}^{RP} < 0 \end{cases} \]

\[ d_{ij}^{SP} = \begin{cases} 1, & U_{ij}^{SP} \geq 0 \\ 0, & U_{ij}^{SP} < 0 \end{cases} \]
DCM Framework

Product characteristics
Consumer characteristics

Situational constraints
Preferences

Market behavior (RP data)

Stated preferences (SP data)

Estimation technique:

\[ L = \prod_{i \in RP} \prod_{j \in C_i} \left[ \Pr(d_{ij}^{RP}) \right]^{d_{ij}^{RP}} \times \prod_{i \in SP} \prod_{j \in C_i} \left[ \Pr(d_{ij}^{SP}) \right]^{d_{ij}^{SP}} \]
DCM Framework

Challenges with RP data:
• Heterogeneity with respect to latent consumer attributes
  • Perceptions
  • Attitudes
• Ex. in transport choice:
  • Convenience
  • Comfort
• Ex. in culture:
  • Beauty
  • Point of interest
  • Breathtaking

Product characteristics
Consumer characteristics
Preferences
Market behavior
(RP data)
DCM Framework

Product characteristics, $X$
Consumer characteristics, $D$

Preferences $U$

Market behavior (RP data), $d$

Attitudes Perceptions $w$

Perception indicators, $y$

Structural model:

$$U_{ij}^{RP} = V_{ij}(X_j, D_i, w_i) + \epsilon_{ij}^{RP}$$
$$w_i = BD_i + \epsilon_i$$

$w_i$ are latent perceptions for alternative or its characteristics
DCM Framework

Product characteristics, $X$
Consumer characteristics, $D$

Preferences $U$

Attitudes
Perceptions $w$

Market behavior (RP data), $d$

Perception indicators, $y$

Measurement (binary) model:

$$d_{ij}^{RP} = \begin{cases} 1, & U_{ij}^{RP} \geq 0 \\ 0, & U_{ij}^{RP} < 0 \end{cases}$$

$$y_i = \Lambda w_i + \nu_i$$

$w_i$ are latent perceptions
$y_i$ are perception indicators
DCM Framework

Product characteristics, \( X \)
Consumer characteristics, \( D \)

Preferences \( U \)
Attitudes Perceptions \( w \)

Market behavior (RP data), \( d \)
Perception indicators, \( y \)

Estimation technique:

First stage (LISRES):

\[ w_i = BD_i + \varepsilon_i \]
\[ y_i = \Lambda w_i + \nu_i \]
\[ y_i = \Lambda(BD_i + \varepsilon_i) + \nu_i \]

Obtain
\[ \hat{w}_i = \hat{\Lambda}^{-1}y_i \]
DCM Framework

- Product characteristics, $X$
- Consumer characteristics, $D$

Preferences $U$

Attitudes Perceptions $\hat{w}$

Market behavior (RP data), $d$

Perception indicators, $y$

Estimation technique:

Second stage:

$$L = \prod_{i \in RP} \prod_{j \in C_i} \left[ \Pr(d_{ij}^{RP} | X_j, D_i, \hat{w}_i) \right]^{d_{ij}^{RP}}$$
DCM Framework

Product characteristics, $X$
Consumer characteristics, $D$

Situational constraints

Preferences $U$

Attitudes Perceptions $w$

Market behavior (RP data), $d^{RP}$

Stated preferences (SP data), $d^{SP}$

Perception indicators, $y$
Application: train vs car choice

- Nijmegen – the city of interview
- Travel to Randstad (Amsterdam, Hague, Rotterdam)
  - By rail or by car, both approximately 2 hours
- Home conducted interview (228 respondents)
- Actual choice of intercity trip to Randstad during previous 3 months (RP data)
  - Level of service attributes (travel time, cost etc.)
  - Socio-economic characteristics (age, sex) and trip goal
  - Subjective rating of latent travel characteristics
- SP experiment of a choice between two different rail services (SP1 data, 2875 comparisons, ordered choice data)
- SP experiment of a choice between rail and car (SP2 data, 1577 comparisons, ordered choice data)
SP data: choice between two options

• Pairwise comparison:
  • SP1: two rail services
  • SP2: rail vs car
• Attributes:
  • Travel cost
  • Travel time
  • Number of transfers (for trains)
  • Luxury level of train (for trains)
• Answers:
  • Definitely choose the alternative 1
  • Probably choose the alternative 1
  • Not sure
  • Probably choose the alternative 2
  • Definitely choose the alternative 1
### SP data: choice between two options

<table>
<thead>
<tr>
<th></th>
<th>RP</th>
<th>SP1</th>
<th>SP2</th>
<th>RP + SP1</th>
<th>RP + SP2</th>
<th>RP + SP1 + SP2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rail constant (RP)</td>
<td>0.501</td>
<td></td>
<td></td>
<td>0.455</td>
<td>0.702</td>
<td>0.718</td>
</tr>
<tr>
<td></td>
<td>(1.8)</td>
<td></td>
<td></td>
<td>(1.8)</td>
<td>(3.0)</td>
<td>(3.4)</td>
</tr>
<tr>
<td>Rail constant (SP)</td>
<td></td>
<td>-0.970</td>
<td></td>
<td></td>
<td>-3.82</td>
<td>-3.82</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-9.8)</td>
<td></td>
<td></td>
<td>(-4.0)</td>
<td>(-4.0)</td>
</tr>
<tr>
<td>Cost per person</td>
<td>-0.0270</td>
<td>-0.0828</td>
<td>-0.0111</td>
<td>-0.0279</td>
<td>-0.0338</td>
<td>-0.0337</td>
</tr>
<tr>
<td></td>
<td>(-4.4)</td>
<td>(-25.4)</td>
<td>(-5.6)</td>
<td>(-5.2)</td>
<td>(-6.5)</td>
<td>(-6.8)</td>
</tr>
<tr>
<td>Line-haul time</td>
<td>-0.342</td>
<td>-0.967</td>
<td>-0.156</td>
<td>-0.327</td>
<td>-0.401</td>
<td>-0.394</td>
</tr>
<tr>
<td></td>
<td>(-1.4)</td>
<td>(-11.6)</td>
<td>(-1.9)</td>
<td>(-4.9)</td>
<td>(-2.1)</td>
<td>(-6.1)</td>
</tr>
<tr>
<td>Terminal time</td>
<td>-1.61</td>
<td>-0.272</td>
<td>-1.60</td>
<td>-1.46</td>
<td>-1.47</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-4.83)</td>
<td>(-1.9)</td>
<td>(-4.9)</td>
<td>(-4.63)</td>
<td>(-4.77)</td>
<td></td>
</tr>
<tr>
<td>Number of transfers</td>
<td>-0.139</td>
<td>-0.140</td>
<td>0.0433</td>
<td>-0.0478</td>
<td>-0.0348</td>
<td>-0.0569</td>
</tr>
<tr>
<td></td>
<td>(-1.0)</td>
<td>(-4.3)</td>
<td>(0.8)</td>
<td>(-3.4)</td>
<td>(-0.3)</td>
<td>(-3.8)</td>
</tr>
<tr>
<td>Comfort</td>
<td>0.493</td>
<td></td>
<td>0.166</td>
<td></td>
<td></td>
<td>0.201</td>
</tr>
<tr>
<td></td>
<td>(14.4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(6.24)</td>
</tr>
<tr>
<td>Business trip dummy</td>
<td>0.902</td>
<td>-0.115</td>
<td>0.887</td>
<td>0.358</td>
<td>0.363</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.2)</td>
<td>(-1.2)</td>
<td>(3.2)</td>
<td>(1.74)</td>
<td>(1.78)</td>
<td></td>
</tr>
<tr>
<td>Female dummy</td>
<td>0.488</td>
<td>-0.102</td>
<td>0.488</td>
<td>0.230</td>
<td>0.232</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.4)</td>
<td>(-1.5)</td>
<td>(2.4)</td>
<td>(1.4)</td>
<td>(1.5)</td>
<td></td>
</tr>
<tr>
<td>Inertia dummy</td>
<td>1.60</td>
<td></td>
<td>5.68</td>
<td></td>
<td>5.70</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(18.7)</td>
<td></td>
<td></td>
<td></td>
<td>(4.7)</td>
<td>(4.8)</td>
</tr>
</tbody>
</table>
Predicting the latent attributes

• For both chosen and unchosen modes
• Perceptual indicators
  • Relaxation during the trip (relax)
  • Reliability of arrival time (relia)
  • Flexibility of choosing departure time (flex)
  • Ease of travelling with children or heavy luggage (ease)
  • Safety during the trip (safety)
  • Overall rating of the mode
• Each indicator is valued by 5-point scale
• Overall rating is valued by 10-point scale
• Two latent attributes:
  • Ride comfort ($w_1$)
  • Convenience ($w_2$)
Predicting the latent attributes

- Two latent attributes:
  - Ride comfort \( (w_1) \)
  - Convenience \( (w_2) \)
- \( w \) affected by consumer attributes \( D \) through \( B \)
- \( w \) affect perceptual indicators \( y \) through \( \Lambda \)

\[
\hat{B}' = \begin{bmatrix}
(w_1^*) & (w_2^*) \\
-0.427(-2.4) & 0.378(2.4) \\
-0.323(-1.7) & 0 \\
0 & -1.98(-9.0) \\
0.281(0.9) & 0 \\
0 & -0.396(-3.7) \\
0 & 0.482(3.5) \\
-0.339(-1.3) & 0 \\
\end{bmatrix}
\]

\[
\hat{\Lambda} = \begin{bmatrix}
(w_1^*) & (w_2^*) \\
0.433(7.6) & 0.280(3.2) \\
0.527(12.5) & 0.661(10.2) \\
0 & 0.815(14.7) \\
0 & 0.794(14.2) \\
0.462(11.6) & 0.311(5.2) \\
0.784(8.5) & 1.76(14.1) \\
\end{bmatrix}
\]
## RP model with latent attributes

<table>
<thead>
<tr>
<th></th>
<th>Model w/o Latent Attributes</th>
<th>Sequential Estimation Model</th>
<th>Simultaneous Estimation Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rail constant</td>
<td>0.583</td>
<td>0.322</td>
<td>−1.81</td>
</tr>
<tr>
<td></td>
<td>(2.0)</td>
<td>(1.0)</td>
<td>(−0.9)</td>
</tr>
<tr>
<td>Cost per person</td>
<td>−0.0268</td>
<td>−0.0338</td>
<td>−0.0379</td>
</tr>
<tr>
<td></td>
<td>(−4.2)</td>
<td>(−4.1)</td>
<td>(−4.3)</td>
</tr>
<tr>
<td>Line-haul time</td>
<td>−0.405</td>
<td>0.0751</td>
<td>0.379</td>
</tr>
<tr>
<td></td>
<td>(−1.6)</td>
<td>(0.2)</td>
<td>(0.9)</td>
</tr>
<tr>
<td>Terminal time</td>
<td>−1.57</td>
<td>−1.18</td>
<td>−0.818</td>
</tr>
<tr>
<td></td>
<td>(−4.2)</td>
<td>(−2.6)</td>
<td>(−2.3)</td>
</tr>
<tr>
<td>Number of transfers</td>
<td>−0.195</td>
<td>−0.316</td>
<td>−0.230</td>
</tr>
<tr>
<td></td>
<td>(−1.3)</td>
<td>(−1.7)</td>
<td>(−1.2)</td>
</tr>
<tr>
<td>Business trip dummy</td>
<td>0.942</td>
<td>1.33</td>
<td>1.28</td>
</tr>
<tr>
<td></td>
<td>(3.6)</td>
<td>(3.6)</td>
<td>(3.3)</td>
</tr>
<tr>
<td>Female dummy</td>
<td>0.466</td>
<td>0.652</td>
<td>0.700</td>
</tr>
<tr>
<td></td>
<td>(2.3)</td>
<td>(2.6)</td>
<td>(2.9)</td>
</tr>
<tr>
<td>$w_1^*$ (comfort)</td>
<td></td>
<td>0.882</td>
<td>1.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.7)</td>
<td>(1.8)</td>
</tr>
<tr>
<td>$w_2^*$ (convenience)</td>
<td></td>
<td>1.39</td>
<td>1.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(4.1)</td>
<td>(4.7)</td>
</tr>
</tbody>
</table>
Conclusion

• RP+SP+Latent variables give:
  • Identification of preferences for new alternatives/attributes (SP vs RP)
  • Bias correction for SP (SP+RP)
  • Efficiency (SP+RP+Latent variables)