Estimation of multinomial logit model using the Begg & Gray approximation

Hana Ševčíková and Adrian Raftery

February 21, 2012

1 MNL Model Specification

Our observations correspond to \( N \) individuals each of whom makes one choice out of \( J \) alternatives. The dependent variable, \( Y_n \), is the choice made by the \( n \)-th individual. The set of independent variables is divided into a set of variables that are individual-specific, say \( X_n = (x_{n1}, \ldots, x_{nK_0})^T \), and a set of variables that are alternative-specific, say \( W_{ni} = (w_{ni1}, \ldots, w_{niK_a})^T \), \( i = 1, \ldots, J \). The probability of individual \( n \) choosing alternative \( i \) is given by the standard multinomial logit formula

\[
P_{ni} = \frac{e^{V_{ni}}}{\sum_{j=1}^{J} e^{V_{nj}}} \quad \text{where} \quad V_{ni} = \alpha_{i0} + \alpha_i X_n + \beta_i W_{ni}
\]

Here, \( \alpha_i = (\alpha_{i1}, \ldots, \alpha_{iK_0}) \) and \( \beta_i = (\beta_{i1}, \ldots, \beta_{iK_a}) \). We call \( \alpha_{i0}, \ldots, \alpha_{J0} \) the alternative-specific constants.

It is often of interest to constrain the coefficients to be the same over the given set of alternatives, i.e. \( \alpha_{1k} = \alpha_{2k} = \cdots = \alpha_{Jk} \) for a given \( k \). The same applies to the \( \beta \) coefficients.

**Base Alternative**

In order to be able to use the Begg & Gray approximation [1], we need to set a base alternative and treat the remaining alternatives as differences to the base. Thus, if the base alternative is 1, \( V_{n1} = 0 \) for all \( n \). Furthermore,

\[
V_{ni} = \alpha_{i0} + \alpha_i X_n + \beta_i W'_{ni} \quad \text{where} \quad W'_{ni} = W_{ni} - W_{n1} \quad \text{for} \quad i = 2, \ldots, J
\]

2 Conversion

The conversion is done analogously to [2]. We decompose the original dataset into \( D_0 = \{D_b, D_r\} \), where \( D_b \) denotes the set of individuals that chose the base alternative, and \( D_r \) denotes the set of the remaining individuals. For each
\(i = 1, \ldots, J\), set \(N_i\) to be the number of individuals that chose alternative \(i\). Then the converted dataset is constructed as follows:

1. Form \(J\) matrices \(M_1, \ldots, M_J\) where each \(M_i\) has \(N_i\) rows and the columns consist of \(Y, X\) and \(U = W_i'\).

2. Form \(J-1\) blocks, \(D_2, \ldots, D_J\), where \(D_i\) has \((N_1 + N_i)\) rows and is formed as follows:
   
   (a) Take the rows of \(M_1\) and \(M_i\).
   
   (b) Add columns:
      
      \[
      \begin{align*}
      Y^* &= \begin{cases}
          0 : \ Y = 1 \\
          1 : \ \text{otherwise}
        \end{cases} \\
      \{Z_2, \ldots, Z_J\}, \text{ where } Z_j &= \begin{cases}
          1 : \ j = i \\
          0 : \ \text{otherwise}
        \end{cases} \\
      \{Z_2X, \ldots, Z_JX\} \\
      \{Z_2U, \ldots, Z_JU\}
      \end{align*}
      \]

3. Combine the rows of \(D_2, \ldots, D_J\).

The approximated binary logistic model is given by

\[
\logit(P[Y^* = 1]) = \gamma_2 + \sum_{l=3}^{J} \gamma_l Z_l + \sum_{k=1}^{K_o} Q(\delta(\cdot)_k, X_k) + \sum_{k=1}^{K_a} Q(\theta(\cdot)_k, U_k) \tag{3}
\]

where

\[
Q(\tau(\cdot)_k, S) = \begin{cases}
    \tau_k S & : \text{if the coefficients of } S \text{ are constrained to be the same for all alternatives, i.e. } \tau(\cdot)_k = \tau_k \\
    \sum_{l=2}^{J} \tau_l Z_l S & : \text{otherwise}
\end{cases}
\]

Given estimated coefficients \(\hat{\gamma}, \hat{\delta} \text{ and } \hat{\theta}\), estimators of the coefficients of the original model in Equation (2) are given by:

\[
\hat{\alpha}_{10} = 0, \quad \hat{\alpha}_{20} = \hat{\gamma}_2, \quad \hat{\alpha}_{i0} = \hat{\gamma}_i + \hat{\gamma}_2, \quad \text{for } i = 3, \ldots, J \\
\hat{\alpha}_{ik} = \hat{\delta}_{ik}, \quad \text{for } i = 2, \ldots, J, \ k = 1, \ldots, K_o \\
\hat{\beta}_{ik} = \hat{\theta}_{ik}, \quad \text{for } i = 2, \ldots, J, \ k = 1, \ldots, K_a
\]

3 Example

Consider the following toy dataset with eight individuals, four alternatives and two independent variables, \(X\) and \(W\):

\[
\begin{array}{cccccccc}
\text{id} & Y & X & W_1 & W_2 & W_3 & W_4 \\
\hline
1 & 1 & x_1 & w_{11} & w_{12} & w_{13} & w_{14} \\
2 & 1 & x_2 & w_{21} & w_{22} & w_{23} & w_{24} \\
3 & 2 & x_3 & w_{31} & w_{32} & w_{33} & w_{34} \\
4 & 2 & x_4 & w_{41} & w_{42} & w_{43} & w_{44} \\
5 & 3 & x_5 & w_{51} & w_{52} & w_{53} & w_{54} \\
6 & 3 & x_6 & w_{61} & w_{62} & w_{63} & w_{64} \\
7 & 4 & x_7 & w_{71} & w_{72} & w_{73} & w_{74} \\
8 & 4 & x_8 & w_{81} & w_{82} & w_{83} & w_{84} \\
\end{array}
\]
Setting the base alternative to 1, the converted dataset is of the form:

<table>
<thead>
<tr>
<th>Y</th>
<th>X</th>
<th>U</th>
<th>Y*</th>
<th>Z_2</th>
<th>Z_3</th>
<th>Z_4</th>
<th>Z_2X</th>
<th>Z_3X</th>
<th>Z_4X</th>
<th>Z_2U</th>
<th>Z_3U</th>
<th>Z_4U</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>w_{12} - w_{11}</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>x_1</td>
<td>0</td>
<td>0</td>
<td>w_{12} - w_{11}</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D_2</td>
<td>1</td>
<td>w_{22} - w_{21}</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>x_2</td>
<td>0</td>
<td>0</td>
<td>w_{22} - w_{21}</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>w_{32} - w_{31}</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>x_3</td>
<td>0</td>
<td>0</td>
<td>w_{32} - w_{31}</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>w_{42} - w_{41}</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>x_4</td>
<td>0</td>
<td>0</td>
<td>w_{42} - w_{41}</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D_3</td>
<td>1</td>
<td>w_{13} - w_{11}</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>x_1</td>
<td>0</td>
<td>0</td>
<td>w_{13} - w_{11}</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>w_{23} - w_{21}</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>x_2</td>
<td>0</td>
<td>0</td>
<td>w_{23} - w_{21}</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>w_{53} - w_{51}</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>x_5</td>
<td>0</td>
<td>0</td>
<td>w_{53} - w_{51}</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>w_{63} - w_{61}</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>x_6</td>
<td>0</td>
<td>0</td>
<td>w_{63} - w_{61}</td>
<td>0</td>
</tr>
<tr>
<td>D_4</td>
<td>1</td>
<td>w_{14} - w_{11}</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>x_1</td>
<td>0</td>
<td>0</td>
<td>w_{14} - w_{11}</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>w_{24} - w_{21}</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>x_2</td>
<td>0</td>
<td>0</td>
<td>w_{24} - w_{21}</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>w_{74} - w_{71}</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>x_7</td>
<td>0</td>
<td>0</td>
<td>w_{74} - w_{71}</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>w_{84} - w_{81}</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>x_8</td>
<td>0</td>
<td>0</td>
<td>w_{84} - w_{81}</td>
</tr>
</tbody>
</table>

An MNL model, specified in \texttt{mlogitBMA} by \( Y \sim 1 \mid X + U \), is approximated using the logit model

\[ Y^* \sim Z_3 + Z_4 + Z_2X + Z_3X + Z_4X + Z_2U + Z_3U + Z_4U \]

The MNL coefficients from Equation (2) correspond to:

\[
\begin{align*}
(\alpha_{20}, \alpha_2, \beta_2) &= (\text{Intercept}, \ Z_2X \text{coef.}, \ Z_2U \text{coef}) \\
(\alpha_{30}, \alpha_3, \beta_3) &= (\text{Intercept} + Z_3 \text{coef.}, \ Z_3X \text{coef.}, \ Z_3U \text{coef}) \\
(\alpha_{40}, \alpha_4, \beta_4) &= (\text{Intercept} + Z_4 \text{coef.}, \ Z_4X \text{coef.}, \ Z_4U \text{coef})
\end{align*}
\]

If we constrain the coefficients to be the same for all alternatives, i.e. \( \alpha = \alpha_2 = \alpha_3 = \alpha_4 \) and \( \beta = \beta_2 = \beta_3 = \beta_4 \), which is specified in \texttt{mlogitBMA} by \( Y \sim X + U \), the logit model

\[ Y^* \sim Z_3 + Z_4 + X + U \]

is used as an approximation. In this case, \( \alpha \) corresponds to the coefficient of \( X \) and \( \beta \) corresponds to the coefficient of \( U \).

References
