The Simulated Mixed Logit Model in the Preference Space and Willingness-to-Pay Space

John Paul Helveston

November 2, 2014

Abstract

The mixed logit model is a commonly used discrete choice model to estimate heterogeneous preference models. Because the model coefficients are sometimes difficult to interpret, the coefficients are sometimes converted from the preference space into the "willingness-to-pay" (WTP) space, which is measured in dollars rather than units of utility. Estimation using simulated maximum likelihood is sometimes slow, but providing an analytic gradient function can speed up the optimization procedure by reducing the number of log-likelihood function evaluations. This document describes in detail the maximum simulated likelihood procedures for estimating a mixed logit model in both the preference and WTP space, including derivations of the analytic gradients in each space. The summary is limited to mixed logit models with random coefficients that are assumed normally or log-normally distributed with uncorrelated heterogeneity covariances (i.e. a diagonal heterogeneity covariance matrix). This document was primarily prepared to accompany the *logitr* program written in the *R* statistical programming language.

Contents

1 The Mixed Logit Model in the Preference Space

1.1 The Model

Train (2009) provides the following description of a mixed logit model in the preference space using simulation to estimate the preference coefficients. In a preference space model, utility is given by

$$u_{njt} = \boldsymbol{\beta}_n' \mathbf{x}_{jt} + \varepsilon_{njt} = v_{njt} + \varepsilon_{njt} \tag{1}$$

where $\boldsymbol{\beta}_n$ is a vector of preference taste coefficients for individual n and $\mathbf{x}_{jt}$ is a vector of attribute values for each attribute of alternative j in each choice situation t. The quantity $v_{nj} = \boldsymbol{\beta}_n' \mathbf{x}_j$ is referred to as the "observed utility." The coefficients $\boldsymbol{\beta}_n$ are assumed to be distributed with density $f(\boldsymbol{\beta}_n \mid \boldsymbol{\theta}_n)$, where $\boldsymbol{\theta}_n$ refers collectively to the parameters of this distribution, such as the mean and covariance if assumed normally distributed. The researcher specifies the functional form $f(\cdot)$ and wants to estimate the parameters $\boldsymbol{\theta}_n$. The choice probabilities are given by

$$P_{ni} = \int L_{ni}(\boldsymbol{\beta}_n) f(\boldsymbol{\beta}_n \mid \boldsymbol{\theta}_n) d\,\boldsymbol{\beta}_n, \tag{2}$$

where

$$L_{ni}(\boldsymbol{\beta}_n) = \frac{e^{v_{ni}}}{\sum_{j=1}^{J} e^{v_{nj}}} \tag{3}$$

For any given value of $\boldsymbol{\theta}_n$, the probabilities in equation 2 are approximated through simulation by the following procedure:

1. Draw a value of $\boldsymbol{\beta}_n$ from $f(\boldsymbol{\beta}_n \mid \boldsymbol{\theta}_n)$, and label it $\boldsymbol{\beta}_n^r$ with the superscript $r = 1$ referring to the first draw.
2. Calculate the logit formula (equation 3) with this draw.
3. Repeat steps 1 and 2 many times, and average the results.

The average over all draws is the simulated probability, given by

$$\hat{P}_{ni} = \frac{1}{R} \sum_{r=1}^{R} L_{ni}(\boldsymbol{\beta}_n^r) \tag{4}$$

where R is the number of draws. $\hat{P}_{ni}$ is an unbiased estimator of P_{ni} by construction. The simulated probabilities are inserted into the log-likelihood function to give a simulated log likelihood (SLL):

$$\mathrm{SLL} = \sum_{n=1}^{N} \sum_{j=1}^{J} d_{nj} ln \hat{P}_{nj} \tag{5}$$

where $d_{nj} = 1$ if n chose j and zero otherwise.The maximum simulated likelihood estimator (MSLE) is the value of $\boldsymbol{\theta}$ that maximizes SLL.

1.2 The Estimation Procedure Using Standard Normal Draws

When actually estimating the mixed logit model, we typically do not make draws of $\boldsymbol{\beta}_n$ directly from $f(\boldsymbol{\beta}_n \mid \boldsymbol{\theta}_n)$. Instead, we make R draws from a standard normal distribution, $\boldsymbol{\Delta} \sim \mathrm{N}(0, 1)$, and shift them by $\boldsymbol{\theta}_n$ in order to get the simulated draws of $\boldsymbol{\beta}_n$ in each iteration through the optimization procedure. Note that $\boldsymbol{\Delta}$ is a matrix of draws of the dimensions $R \times N$, where N is the number of parameters in $\boldsymbol{\beta}_n$.

If $\boldsymbol{\beta}_n$ is assumed normally distributed, $\boldsymbol{\beta}_n \sim \mathrm{N}(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n)$, then a single draw, $\boldsymbol{\beta}_n^r$, is

$$\boldsymbol{\beta}_n^r = \delta^r \, \boldsymbol{\sigma}_n + \boldsymbol{\mu}_n \tag{6}$$

where $\boldsymbol{\sigma}_n$ is the diagonal of $\boldsymbol{\Sigma}_n$ and δ^r is a vector of standard normal draws of dimension $R \times 1$ take from $\boldsymbol{\Delta}$. Likewise, if $\boldsymbol{\beta}_n$ is assumed log-normally distributed, $\boldsymbol{\beta}_n \sim \mathrm{logN}(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n)$, then a single draw, $\boldsymbol{\beta}_n^r$, is

$$\boldsymbol{\beta}_n^r = \exp\left(\delta^r \, \boldsymbol{\sigma}_n + \boldsymbol{\mu}_n\right) \tag{7}$$

1.3 The Gradient of the Mixed Logit Model in the Preference Space

Providing the analytic gradient to the SLL in Equation 5 can dramatically improve the speed of the optimization procedure to find the $\boldsymbol{\theta}$ that maximizes the SLL. To compute the gradient we take the partial derivative of Equation 5 with respect to $\boldsymbol{\theta}$:

$$\begin{aligned}
\frac{\partial \mathrm{SLL}}{\partial \boldsymbol{\theta}} &= \sum_{n=1}^{N} \sum_{j=1}^{J} d_{nj} ln(\hat{P}_{nj}) \\
&= \sum_{n=1}^{N} \sum_{j=1}^{J} d_{nj} \frac{\partial ln(\hat{P}_{nj})}{\partial \boldsymbol{\theta}} \\
&= \sum_{n=1}^{N} \sum_{j=1}^{J} d_{nj} \frac{1}{\hat{P}_{nj}} \frac{\partial (\hat{P}_{nj})}{\partial \boldsymbol{\theta}} \\
&= \sum_{n=1}^{N} \sum_{j=1}^{J} d_{nj} \frac{1}{\hat{P}_{nj}} \frac{1}{R} \sum_{r=1}^{R} \frac{\partial}{\partial \boldsymbol{\theta}} \left[L_{nj}(\boldsymbol{\beta}_n^r \mid \boldsymbol{\theta}, \delta^r) \right] \\
&= \sum_{n=1}^{N} \sum_{j=1}^{J} d_{nj} \frac{1}{\hat{P}_{nj}} \frac{1}{R} \sum_{r=1}^{R} \frac{\partial}{\partial \boldsymbol{\theta}} \left[\frac{e^{v_{nj}(\boldsymbol{\beta}_n^r \mid \boldsymbol{\theta}, \delta^r)}}{\sum_{j=1}^{J} e^{v_{nj}(\boldsymbol{\beta}_n^r \mid \boldsymbol{\theta}, \delta^r)}} \right] \\
&= \sum_{n=1}^{N} \sum_{j=1}^{J} d_{nj} \frac{1}{\hat{P}_{nj}} \frac{1}{R} \sum_{r=1}^{R} \frac{\left(\sum_{j=1}^{J} e^{v_{nj}^r}\right) e^{v_{nj}^r} \frac{\partial (v_{nj}^r)}{\partial \boldsymbol{\theta}} - e^{v_{nj}^r} \left(\sum_{j=1}^{J} e^{v_{nj}^r} \frac{\partial (v_{nj}^r)}{\partial \boldsymbol{\theta}}\right)}{\left(\sum_{j=1}^{J} e^{v_{nj}^r}\right)^2} \\
&= \sum_{n=1}^{N} \sum_{j=1}^{J} d_{nj} \frac{1}{\hat{P}_{nj}} \frac{1}{R} \sum_{r=1}^{R} \frac{e^{v_{nj}^r}}{\sum_{j=1}^{J} e^{v_{nj}^r}} \left[\frac{\left(\sum_{j=1}^{J} e^{v_{nj}^r}\right) \frac{\partial (v_{nj}^r)}{\partial \boldsymbol{\theta}} - \left(\sum_{j=1}^{J} e^{v_{nj}^r} \frac{\partial (v_{nj}^r)}{\partial \boldsymbol{\theta}}\right)}{\sum_{j=1}^{J} e^{v_{nj}^r}} \right] \\
&= \sum_{n=1}^{N} \sum_{j=1}^{J} d_{nj} \frac{1}{\hat{P}_{nj}} \frac{1}{R} \sum_{r=1}^{R} L_{nj}^r \left[\frac{\partial (v_{nj}^r)}{\partial \boldsymbol{\theta}} - \left(\sum_{j}^{J} L_{nj}^r \frac{\partial (v_{nj}^r)}{\partial \boldsymbol{\theta}} \right) \right]
\end{aligned} \tag{8}$$

The $\frac{\partial v_{nj}^r}{\partial \boldsymbol{\theta}}$ term in equation 8 depends on the distributional assumptions of $\boldsymbol{\beta}$.

If $\boldsymbol{\beta}_n$ is assumed normally distributed, $\boldsymbol{\beta}_n \sim \mathrm{N}(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n)$, then v_{nj}^r and and $\frac{\partial (v_{nj}^r)}{\partial \boldsymbol{\theta}}$ are

$$v_{nj}^r = \boldsymbol{\beta}_n^{r\prime} \, \mathbf{x}_j = (\delta^r \, \boldsymbol{\sigma}_n + \boldsymbol{\mu}_n)' \, \mathbf{x}_j \tag{9}$$

$$\frac{\partial (v_{nj}^r)}{\partial \boldsymbol{\mu}_n} = \mathbf{x}_j, \quad \frac{\partial (v_{nj}^r)}{\partial \boldsymbol{\sigma}_n} = \delta^{r\prime} \, \mathbf{x}_j \tag{10}$$

Likewise, if $\boldsymbol{\beta}_n$ is assumed log-normally distributed, $\boldsymbol{\beta}_n \sim \mathrm{logN}(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n)$, then v_{nj}^r and $\frac{\partial (v_{nj}^r)}{\partial \boldsymbol{\theta}_n}$ are

$$v_{nj}^r = \boldsymbol{\beta}_n^{r\prime} \, \mathbf{x}_j = \exp(\delta^r \, \boldsymbol{\sigma}_n + \boldsymbol{\mu}_n)' \, \mathbf{x}_j \tag{11}$$

$$\frac{\partial(v_{nj}^r)}{\partial\,\boldsymbol{\mu}_n} = \exp(\delta^r\,\boldsymbol{\sigma}_n + \boldsymbol{\mu}_n)'\,\mathbf{x}_j, \quad \frac{\partial(v_{nj}^r)}{\partial\,\boldsymbol{\sigma}_n} = \exp(\delta^r\,\boldsymbol{\sigma}_n + \boldsymbol{\mu}_n)'\,\delta^r\,\mathbf{x}_j \tag{12}$$

For other distributional assumptions on $\boldsymbol{\beta}_n$, the gradient of the SLL can simply be computed by substituting the appropriate value for $\frac{\partial(v_{nj}^r)}{\partial\,\boldsymbol{\theta}_n}$ into equation 8.

2 The Mixed Logit Model in the Willingness-to-Pay Space

2.1 The Model

In a WTP space model, the utility specification in equation 1 is modified such that the unit of the estimated attribute parameters becomes the dollar rather than units of utility. We begin this modification by separating ou the price parameter, α, and price attribute, p, from the non-price parameters and attributes. The utility specification now becomes

$$u_{njt} = \alpha_n p_{jt} + \boldsymbol{\beta}_n'\,\mathbf{x}_{jt} + \varepsilon_{njt} \tag{13}$$

Train & Weeks (2005) suggest substituting a WTP coefficient defined as $\boldsymbol{\omega}_n = \frac{\boldsymbol{\beta}_n}{\alpha_n}$ into equation 13 to derive the WTP space model, given by

$$u_{njt} = \alpha_n(p_{jt} + \boldsymbol{\omega}_n'\,\mathbf{x}_{jt}) + \varepsilon_{njt} \tag{14}$$

Note that the observed utility in the WTP space is $v_{njt} = \alpha_n(p_{jt} + \boldsymbol{\omega}_n'\,\mathbf{x}_{jt})$, which is nonlinear in parameters.

2.2 The Gradient of the Mixed Logit Model in the WTP Space

Since the only difference between preference and WTP space models is the specifications of v_{nj}, the SLL and gradient of the SLL for a WTP space model is the same except for the $\frac{\partial(v_{nj}^r)}{\partial\,\boldsymbol{\theta}_n}$ term in equation 8. This depends on the distributional assumptions on α_n and $\boldsymbol{\omega}_n$.

For this document, we only provide the equations for the assumptions that α_n is fixed and $\boldsymbol{\omega}_n$ is normally distributed, $\boldsymbol{\omega}_n \sim \mathrm{N}(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n)$. Using these assumptions, v_{nj}^r and and $\frac{\partial(v_{nj}^r)}{\partial\,\boldsymbol{\theta}_n}$ are given by

$$\begin{aligned} v_{nj}^r &= \alpha_n(p_j + \boldsymbol{\omega}_n^{r\prime}\,\mathbf{x}_j) \\ &= \alpha_n[p_j + (\delta^r\,\boldsymbol{\sigma}_n + \boldsymbol{\mu}_n)'\,\mathbf{x}_{nj}] \end{aligned} \tag{15}$$

$$\frac{\partial(v_{nj}^r)}{\partial\alpha_n} = p_j + \boldsymbol{\omega}_n^{r\prime}\,\mathbf{x}_j, \quad \frac{\partial(v_{nj}^r)}{\partial\,\boldsymbol{\mu}_n} = \alpha_n\,\mathbf{x}_j, \quad \frac{\partial(v_{nj}^r)}{\partial\,\boldsymbol{\sigma}_n} = \alpha_n\,\delta^{r\prime}\,\mathbf{x}_j \tag{16}$$

To compute the gradient of the SLL in the WTP space, simply substitute the values for $\frac{\partial(v_{nj}^r)}{\partial\,\boldsymbol{\theta}}$ into equation 8.

References

Train, Kenneth E. 2009. *Discrete Choice Methods with Simulation*. 2nd edn. Cambridge University Press.

Train, Kenneth E., & Weeks, Melyvn. 2005. *Discrete choice models in preference space and willingness-to-pay space*. Springer Netherlands.