Flexible Multinomial Logit Models with Preference Space and Willingness-to-Pay Space Utility Specifications in R: The **logitr** Package

John Paul Helveston
George Washington University

Abstract

In many applications of discrete choice models, modelers are interested in estimating consumer's marginal "willingness-to-pay" (WTP) for different attributes. WTP can computed by dividing the estimated parameters of a utility model in the preference space by the price parameter or by estimating a utility model in the WTP space. For homogeneous models, these two procedures generally produce the same estimates of WTP, but the same is not true for heterogeneous models where model parameters are assumed to follow a specific distribution. The **logitr** package was written to allow for flexible estimation of multinomial logit models with preference space and WTP space utility specifications. The package supports homogeneous multinomial logit (MNL) and heterogeneous mixed logit (MXL) models, including support for normal and log-normal parameter distributions. Since MXL models and models with WTP space utility specifications are non-convex, an option is included to run a multi-start optimization loop with random starting points in each iteration. The package also includes a market simulation function to estimate the expected market shares of a set of alternatives using an estimated model.

Keywords: multinomial logit, preference space, willingness-to-pay space, discrete choice, R.

1. Introduction

In many applications of discrete choice models, modelers are interested in estimating consumer's marginal "willingness-to-pay" (WTP) for different attributes. WTP can be estimated in two ways:

1. Estimate a discrete choice model in the "preference space" where parameters have units of utility and then compute the WTP by dividing the parameters by the price parameter.
2. Estimate a discrete choice model in the "WTP space" where parameters have units of WTP.

While the two procedures generally produce the same estimates of WTP for homogenous models, the same is not true for heterogeneous models where model parameters are assumed to follow a specific distribution, such as normal or log-normal (Train and Weeks 2005). For example, in a preference space specification, a normally distributed attribute parameter divided by a log-normally distributed price parameter produces a strange WTP distribution with large tails. In contrast, a WTP space specification allows the modeler to directly assume

WTP is normally distributed. The **logitr** package was developed to enable modelers to choose between these two utility spaces when estimating multinomial logit models.

2. The random utility model

The random utility model is a well-established framework in many fields for estimating consumer preferences from observed consumer choices (Louviere, Hensher, and Swait 2000, Train (2009)). Random utility models assume that consumers choose the alternative j a set of alternatives that has the greatest utility u_j. Utility is a random variable that is modeled as $u_j = v_j + \varepsilon_j$, where v_j is the "observed utility" (a function of the observed attributes such that $v_j = f(\mathbf{x}_j)$) and ε_j is a random variable representing the portion of utility unobservable to the modeler.

Adopting the same notation as in (Helveston, Feit, and Michalek 2018), consider the following utility model:

$$u_j^* = \boldsymbol{\beta}^{*\prime}\mathbf{x}_j - \alpha^* p_j + \varepsilon_j^*, \qquad \varepsilon_j^* \sim \text{Gumbel}\left(0, \sigma^2 \frac{\pi^2}{6}\right) \tag{1}$$

where $\boldsymbol{\beta}^*$ is the vector of coefficients for non-price attributes $\mathbf{x}_j$, α^* is the coefficient for price p_j, and the error term, ε_j^*, is an IID random variable with a Gumbel extreme value distribution of mean zero and variance $\sigma^2(\pi^2/6)$. This model is not identified since there exists an infinite set of combinations of values for $\boldsymbol{\beta}^*$, α^*, and σ that produce the same choice probabilities. In order to specify an identifiable model, the modeler must normalize equation (1). One approach is to normalize the scale of the error term by dividing equation (1) by σ, producing the "preference space" utility specification:

$$\left(\frac{u_j^*}{\sigma}\right) = \left(\frac{\boldsymbol{\beta}^*}{\sigma}\right)'\mathbf{x}_j - \left(\frac{\alpha^*}{\sigma}\right)p_j + \left(\frac{\varepsilon_j^*}{\sigma}\right), \qquad \left(\frac{\varepsilon_j^*}{\sigma}\right) \sim \text{Gumbel}\left(0, \frac{\pi^2}{6}\right) \tag{2}$$

The typical preference space parameterization of the multinomial logit (MNL) model can then be written by rewriting equation (2) with $u_j = (u_j^*/\sigma)$, $\boldsymbol{\beta} = (\boldsymbol{\beta}^*/\sigma)$, $\alpha = (\alpha^*/\sigma)$, and $\varepsilon_j = (\varepsilon_j^*/\sigma)$:

$$u_j = \boldsymbol{\beta}'\mathbf{x}_j - \alpha p_j + \varepsilon_j \qquad \varepsilon_j \sim \text{Gumbel}\left(0, \frac{\pi^2}{6}\right) \tag{3}$$

The vector $\boldsymbol{\beta}$ represents the marginal utility for changes in each non-price attribute, and α represents the marginal utility obtained from price reductions. In addition, the coefficients $\boldsymbol{\beta}$ and α are measured in units of *utility*, which only has relative rather than absolute meaning.

The alternative normalization approach is to normalize equation (1) by α^* instead of σ, producing the "willingness-to-pay (WTP) space" utility specification:

$$\left(\frac{u_j^*}{\alpha^*}\right) = \left(\frac{\boldsymbol{\beta}^*}{\alpha^*}\right)'\mathbf{x}_j - p_j + \left(\frac{\varepsilon_j^*}{\alpha^*}\right), \qquad \left(\frac{\varepsilon_j^*}{\alpha^*}\right) \sim \text{Gumbel}\left(0, \frac{\sigma^2}{(\alpha^*)^2}\frac{\pi^2}{6}\right) \tag{4}$$

Since the error term in equation is scaled by $\lambda^2 = \sigma^2/(\alpha^*)^2$, we can rewrite equation (4) by multiplying both sides by $\lambda = (\alpha^*/\sigma)$ and renaming $u_j = (\lambda u_j^*/\alpha^*)$, $\boldsymbol{\omega} = (\boldsymbol{\beta}^*/\alpha^*)$, and $\varepsilon_j = (\lambda\varepsilon_j^*/\alpha^*)$:

$$u_j = \lambda \left(\boldsymbol{\omega}' \mathbf{x}_j - p_j \right) + \varepsilon_j \qquad \varepsilon_j \sim \text{Gumbel}\left(0, \frac{\pi^2}{6}\right) \tag{5}$$

Here $\boldsymbol{\omega}$ represents the marginal WTP for changes in each non-price attribute, and λ represents the scale of the deterministic portion of utility relative to the standardized scale of the random error term.

The utility models in equations 3 and 5 represent the preference space and WTP space utility specifications, respectively. In equation 3, WTP is estimated as $\hat{\boldsymbol{\beta}}/\hat{\alpha}$; in equation 5, WTP is simply $\hat{\boldsymbol{\omega}}$.

3. The logitr package

3.1. Installation

First, make sure you have the **devtools** library installed:

```
R> install.packages("devtools")
```

Then load the **devtools** library and install the **logitr** package:

\begin{CodeChunk}

\begin{CodeInput} R> library("devtools") R> install_github("jhelvy/logitr") R> library("logitr") \end{CodeInput} \end{CodeChunk}

3.2. Data format

References

Helveston JP, Feit EM, Michalek JJ (2018). "Pooling stated and revealed preference data in the presence of RP endogeneity." *Transportation Research Part B: Methodological*, **109**, 70–89.

Louviere JJ, Hensher DA, Swait J (2000). *Stated Choice Methods: Analysis and Applications.* Cambridge University Press, Cambridge, Massachusetts.

Train KE (2009). *Discrete Choice Methods with Simulation.* 2nd edition. Cambridge University Press.

Train KE, Weeks M (2005). "Discrete Choice Models in Preference and Willingness-to-Pay Space." In *Appl. Simul. Methods Environ. Resour. Econ.*, chapter 1, pp. 1–16.

Affiliation:

John Paul Helveston
George Washington University

Science & Engineering Hall
800 22nd St NW
Washington, DC 20052
E-mail: jph@gwu.edu
URL: http://jhelvy.com