Available online at www.sciencedirect.com
ScienceDirect
Journal of Transport Geography 16 (2008) 126–133
JOURNAL OF TRANSPORT GEOGRAPHY
www.elsevier.com/locate/jtrangeo

Combining RP and SP data: biases in using the nested logit 'trick'– contrasts with flexible mixed logit incorporating panel and scale effects

David A. Hensher [a,*], John M. Rose [a], William H. Greene [b]

[a] Faculty of Economics and Business, Institute of Transport and Logistics Studies, The University of Sydney, NSW, Sydney 2006, Australia
[b] Stern School of Business, New York University, United States

Abstract

It has become popular practice that joint estimation of choice models that use stated preference (SP) and revealed preference (RP) data requires a way of adjusting for scale to ensure that parameter estimates across data sets are not confounded by differences in scale. The nested logit 'trick' presented by Hensher and Bradley in 1993 [Hensher, D.A., Bradley, M., 1993. Using stated response data to enrich revealed preference discrete choice models. Marketing Letters 4 (2), 139–152] continues to be widely used, especially by practitioners, to accommodate scale differences. This modelling strategy has always assumed that the observations are independent, a condition of all GEV models, which is not strictly valid within a stated preference experiment with repeated choice sets and between each SP observation and the single RP data point. This paper promotes the replacement of the NL 'trick' method with an error components model that can accommodate correlated observations as well as reveal the relevant scale parameter for subsets of alternatives. Such a model can also incorporate ''state'' or reference dependence between data types and preference heterogeneity on observed attributes. An example illustrates the difference in empirical evidence.

Keywords: Combined SP and RP data; Nested logit trick; Scale parameters; Error component mixed logit; Reference dependence

1. Introduction

It is common practice in estimation of many choice models which combine multiple data sources (e.g., RP and SP data sets) to use a nested logit (NL) structure as a 'trick or mechanism to reveal differences in scale between data sources' (Bradley and Daly, 1992, 1997; Hensher and Bradley, 1993). It is a trick in the sense that the underlying conditions to comply with utility maximisation such as the 0–1 bound on the inclusive value variable linking two levels in a nest (McFadden, 1981), while applicable between alternatives within SP and within RP choice sets, are not relevant *between* data sets – the scale differences between data sets (typically normalising to unity on one data source) is the only agenda.

In the majority of NL applications, the predictability of the set of NL structures studied is driven by the revelation of differences in SP–RP scale parameters and/or the partitioning of alternatives within a given data set in what is best described as 'commonsense' or intuitive partitions; for example, the marginal choice between car and public transport and then the choice between bus and train, conditional on choosing public transport. Hensher (1999) generalised the role of scale parameters through the use of the HEV search engine to allow for differences in scale, not only between data sets but between alternatives within and between data sets.

The NL model is a member of GEV models (McFadden, 1981) which cannot accommodate a number of specification

* Corresponding author. Tel.: +61 2 93510071; fax: +61 2 93510088.
E-mail addresses: Davidh@itls.usyd.edu.au (D.A. Hensher), Johnr@itls.usyd.edu.au (J.M. Rose), wgreene@stern.nyu.edu (W.H. Greene).

0966-6923/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtrangeo.2007.07.001

requirements of data that has repeated observations from the same respondent. This occurs with SP choice sets which exhibit potential correlation due to repeated observations or panel data. We need mixing of some kind for this, using the GEV model as a kernel (see Hess et al., 2005; Hess and Rose, 2006; Garrow and Bodea, 2005).

In addition to potential observation correlation, joint RP–SP estimation induces a potential 'state' or reference dependence effect, defined as the influence of the actual (revealed) choice on the stated choices of the individual. Reference dependence can manifest itself as a positive or negative effect of the choice of an alternative on the utility associated with that alternative in the stated responses (Bhat and Castelar, 2002). In a real sense it is a reflection of accumulated experience and the role that reference dependency plays in choosing, in the spirit of prospect theory (Kahneman and Tversky, 1979; Hensher, 2006; Hess et al., in press).

It is possible that the effect of reference dependence is positive for some individuals and negative for others (see Ailawadi et al., 1999), suggesting that an unconstrained analytical distribution for the random parameterisation of state dependence is appropriate. A positive effect may be the result of habit persistence, inertia to explore another alternative, or learning combined with risk aversion. A negative effect could be the result of variety seeking or the result of latent frustration with the inconvenience associated with the currently used alternative (Bhat and Castelar, 2002).

Thus, joint RP–SP estimation should not only recognise state dependence, but also accommodate heterogeneity in the reference dependence effect if it exists. Most RP–SP studies disregard reference dependence and adopt fixed parameters (i.e., homogeneity of attribute preference). Bhat and Castelar (2002) accommodate such unobserved heterogeneity in the reference dependence effect of the RP choice on SP choices. Brownstone et al. (1996), on the other hand, accommodate observed heterogeneity in the reference dependence effect by interacting the RP choice dummy variable with socio-demographic characteristics of the individual and SP choice attributes.

The paper outlines a very general mixed logit model which brings together the many recent contributions in the literature that deliver a flexible structure that can account for between-alternative error structure including correlated choice sets, RP–SP scale difference, unobserved preference heterogeneity, and reference dependency. An empirical example is used to illustrate the behavioural differences between the traditional NL-trick model and the flexible mixed logit model. We recognise parallel developments in GEV mixture models as summarised and extended by Hess et al. (2005), that can capture scale differences through an underlying NL model while also capturing other phenomena addressed herein through random terms.

2. The mixed logit framework

We begin with the basic form of the multinomial logit model, with alternative-specific constants α_{ji} and attributes x_{ji}, for individuals $i = 1, \ldots, N$ in choice setting t

$$\text{Prob}(y_{it} = j_t) = \frac{\exp(\alpha_{ji} + \boldsymbol{\beta}_i' \mathbf{x}_{jit})}{\sum_{q=1}^{J_i} \exp(\alpha_{qi} + \boldsymbol{\beta}_i' \mathbf{x}_{qit})}. \quad (1)$$

The random parameter model emerges as the form of the individual specific parameter vector, $\boldsymbol{\beta}_i$ is developed. The most familiar, simplest version of the model specifies

$$\begin{aligned} &\beta_{ki} = \beta_k + \sigma_k v_{ik} \text{ and} \\ &\alpha_{ji} = \alpha_j + \sigma_j v_{ji}, \end{aligned} \quad (2)$$

where β_k is the population mean for the kth attribute ($k = 1, \ldots, K$), v_{ik} is the individual specific heterogeneity, with mean zero and standard deviation one, and σ_k is the standard deviation of the distribution of β_{ik}'s around β_k. The term 'mixed logit' is increasingly used in the literature (e.g., Revelt and Train, 1998; Train, 2003; Hensher et al., 2005) for this model. The choice specific constants, α_{ji} and the elements of $\boldsymbol{\beta}_i$ are distributed randomly across individuals with fixed means.

The $v_{ij'}$s are individual and choice specific, unobserved random disturbances – the source of the heterogeneity. For the full vector of K random coefficients in the model, we may write the full set of random parameters as

$$\boldsymbol{\rho}_i = \boldsymbol{\rho} + \boldsymbol{\Gamma} \mathbf{v}_i, \quad (3)$$

where $\boldsymbol{\Gamma}$ is a diagonal matrix which contains σ_k on its diagonal. We can allow the random parameters to be correlated. An additional layer of individual heterogeneity may be added to the model in the form of the error components that capture influences that are related to alternatives in contrast to attributes. We do this by constructing a set of independent individual terms, E_{im}, $m = 1, \ldots, M \sim N[0,1]$ that can be added to the utility functions. This device allows us to create what amounts to a random effects model and, in addition, a very general type of nesting of alternatives. Let θ_m be the scale parameter (standard deviation) associated with these effects. Then, each utility function can be constructed as

$$U_{ijt} = \alpha_{ji} + \beta_i' x_{jit} + (\text{any of } \theta_1 E_{i1}, \theta_2 E_{i2}, \ldots, \theta_M E_{iM}). \quad (4)$$

Consider, for example, a four outcome structure

$$\begin{aligned} U_{i1t} &= V_{i1t} + \theta_1 E_{i1} + \theta_2 E_{i2}, \\ U_{i2t} &= V_{i2t} + \theta_2 E_{i2}, \\ U_{i3t} &= V_{i3t} + \theta_1 E_{i1} + \theta_3 E_{i3}, \\ U_{i4t} &= V_{i4t} + \theta_4 E_{i4}. \end{aligned}$$

Thus, U_{i4t} has its own uncorrelated effect, but there is a correlation between U_{i1t} and U_{i2t} and between U_{i1t} and U_{i3t}. This example is fully populated, so the covariance matrix is block diagonal with the first three freely correlated. The model might usefully be restricted in a specific applica-