The Journal of Choice Modelling 9 (2013) 39–56

Contents lists available at ScienceDirect

The Journal of Choice Modelling

journal homepage: www.elsevier.com/locate/jocm

The benefits of allowing heteroscedastic stochastic distributions in multiple discrete-continuous choice models

Sujan Sikder [a,1], Abdul Rawoof Pinjari [b,*]

[a] *Transportation Systems Modeler, Parsons Brinckerhoff, 400 SW Sixth Avenue, Suite 802, Portland, OR 97204, USA*
[b] *Department of Civil & Environmental Engineering, University of South Florida, 4202 E. Fowler Ave., ENB 118, Tampa, FL 33620, USA*

ARTICLE INFO

Available online 22 January 2014

Keywords:
Discrete-continuous choice models
Multiple discreteness
Heteroscedasticity
Distributional assumptions
Time use behavior
Spatial transferability

ABSTRACT

This paper investigates the benefits of incorporating heteroscedastic stochastic distributions in random utility maximization-based multiple discrete-continuous (MDC) choice models. To this end, first, a Multiple Discrete-Continuous Heteroscedastic Extreme Value (MDCHEV) model is formulated to allow heteroscedastic extreme value stochastic distributions in MDC models. Next, an empirical analysis of individuals′ daily time use choices is carried out using data from the National Household Travel Survey (NHTS) for three geographical regions in Florida. A variety of different MDC model structures are estimated: (a) the Multiple Discrete-Continuous Extreme Value (MDCEV) model with independent and identically distributed (IID) extreme value error structure, (b) the MDCHEV model, (c) the mixed-MDCEV model that allows heteroscedasticity by mixing a heteroscedastic distribution over an IID extreme value kernel, (d) the MDC generalized extreme value (MDCGEV) model that allows inter-alternative correlations using the multivariate extreme value error structure, (e) the mixed-MDCEV model that allows inter-alternative correlations using common mixing distributions across choice alternatives, and (f) the mixed-MDCEV model that allows both heteroscedasticity and inter-alternative correlations. Among all these model structures, the MDCHEV model provided the best fit to the current empirical data. Further, heteroscedasticity was prominent while no significant inter-alternative correlations were found. Specifically, the MDCHEV parameter estimates revealed the significant presence of heteroscedasticity in the random utility components of different activity type choice alternatives. On the other hand, the MDCEV model resulted in inferior model fit and systematic discrepancies between the observed and predicted distributions of time allocations, which can be traced to the thick right tail of the type-1 extreme value distribution. The MDCHEV model addressed these issues to a considerable extent by allowing tighter stochastic distributions for certain choice alternatives, thanks to its accommodation of heteroscedasticity among random utility components. Furthermore, spatial transferability assessments using different transferability metrics also suggest that the MDCHEV model clearly outperformed the MDCEV model.

* Corresponding author. Tel.: +1 813 974 9671; fax: +1 813 974 2957.
E-mail addresses: sikders@pbworld.com (S. Sikder), apinjari@usf.edu (A.R. Pinjari).
[1] Tel.: +1 813 507 3953; fax: +1 813 974 2957.

1755-5345/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jocm.2013.12.003

1. Introduction

1.1. Background

Numerous consumer choices are characterized by “multiple discreteness” where consumers can potentially choose multiple alternatives from a set of discrete alternatives available to them. Along with such discrete-choice decisions of which alternative(s) to choose, consumers typically make continuous-quantity decisions on how much of each chosen alternative to consume. Such multiple discrete-continuous (MDC) choices are being increasingly recognized and analyzed in a variety of scientific fields, including transportation, environmental economics, and marketing. Examples include (1) individuals′ daily time-use choices, which involve decisions to engage in different types of activities in a day along with the allocation of available time to each activity, (2) households′ recreational destination choices and time allocation to the chosen destinations over a season, and (3) grocery shoppers′ brand choice and purchase quantity decisions.

A variety of approaches have been used in the literature to model MDC choices. Among these, a particularly attractive approach is based on the classical microeconomic consumer theory of constrained utility maximization. Specifically, consumers are assumed to optimize a direct utility function $U(\mathbf{x})$ over a set of non-negative consumption quantities $\mathbf{x} = (x_1, ..., x_k, ..., x_K)$ subject to a linear budget constraint, as

$$\text{Max } U(\mathbf{x}) \text{ such that } \mathbf{x} \cdot \mathbf{p} = y \text{ and } x_k \geq 0 \,\forall k = 1, 2, ..., K \tag{1}$$

In the above equation, $U(\mathbf{x})$ is a quasi-concave, increasing and continuously differentiable utility function with respect to the consumption quantity vector $\mathbf{x}$, $\mathbf{p}$ is a vector of unit prices for all goods, and y is a budget for total expenditure. An increasingly popular approach for deriving the demand functions from the utility maximization problem in Eq. (1), due to Hanemann (1978) and Wales and Woodland (1983), is based on the application of Karush–Kuhn–Tucker (KKT) conditions of optimality with respect to the consumption quantities. Since the utility function is assumed to be randomly distributed over the population, the KKT conditions are also randomly distributed and form the basis for deriving the probability expressions for consumption patterns.

Over the past decade, the above-discussed KKT approach has received significant attention for the analysis of MDC choices. A stream of research in environmental economics (Phaneuf et al., 2000; von Haefen et al., 2004; von Haefen and Phaneuf, 2005; Phaneuf and Smith, 2005; Vasquez-Lavin and Hanemann, 2009) advanced the approach to model individuals′ recreational demand choices for non-market valuation of environmental goods. Several studies in marketing research (Kim et al., 2002; Satomura et al., 2011) employed the approach to model situations when consumers purchase a variety of brands of a product (e.g., yogurt). In the transportation field, the multiple discrete-continuous extreme value (MDCEV) model formulated by Bhat (2005) and enlightened further by Bhat (2008) lead to an increased use of the KKT approach for analyzing a variety of choices, including individuals′ activity participation and time-use (Bhat 2005; Habib and Miller, 2008; Pinjari et al., 2009; Chikaraishi et al., 2010; You et al., 2013), household vehicle ownership and usage (Ahn et al., 2008; Bhat et al., 2009; Jaggi et al., 2013), long-distance leisure destination choices (Van Nostrand et al., 2013), energy consumption choices (Pinjari and Bhat, 2011; Frontuto, 2011; Yu et al., 2012) and builders′ land-development choices (Kaza et al., 2012; Farooq et al., 2013). Clever use of stochastic specifications has led to model formulations with closed-form likelihood expressions. Specifically, consider an additive utility form, as follows:

$$U(x_1, ..., x_K) = \sum_{k=1}^{K} U(x_k) = \sum_{k=1}^{K} f(u(x_k), \varepsilon_k) \tag{2}$$

In the above equation, $U(x_k)$ is a random sub-utility function for good k, representing the utility derived from consuming x_k amount of good k, and expressed as a combination of a deterministic component $u(x_k)$ and a random component ε_k as $U(x_k) = f(u(x_k), \varepsilon_k)$. Assuming that the random components (ε_k) enter the sub-utility functions $U(x_k)$ in a multiplicative fashion, as $U(x_k) = u(x_k) \times e^{\varepsilon_k}$, and are type-1 extreme value and independent and identically distributed (IID) across the choice alternatives leads to a very simple and elegant consumption probability expression (Bhat, 2005), making it easy for parameter estimation. In addition, computationally efficient procedures are now available for using these model systems for forecasting and policy evaluation (see von Haefen et al. (2004) and Pinjari and Bhat (2011)). Thanks to these advances, KKT-based MDC models are being increasingly used in empirical research and have begun to be used in operational travel forecasting models.

1.2. Gaps in literature relevant to this paper

Recent research in this area has started to enhance the basic formulation in Eq. (1) along three specific directions: (a) toward more flexible, non-additively separable functional forms for the utility specification so as to accommodate rich complementarity and substitution patterns in consumption (Vasquez-Lavin and Hanemann, 2009; Bhat et al., 2013a), (b) toward greater flexibility in the specification of the constraints faced by the consumer, such as multiple linear budget constraints as opposed to a single constraint (Satomura et al., 2011; Castro et al., 2012; Pinjari and Sivaraman, 2013), and (c) toward more flexible stochastic specifications for the random utility functions. The reader is referred to Pinjari et al. (2013) for a more detailed discussion of recent advances along the first two directions.