The Journal of Choice Modelling 6 (2013) 44–59

Contents lists available at SciVerse ScienceDirect

The Journal of Choice Modelling

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

Stated response and multiple discrete-continuous choice models: Analyses of residuals

Boris Jäggi*, Claude Weis[1], Kay W. Axhausen[2]

IVT, ETH Zürich, CH-8093 Zürich, Switzerland

ARTICLE INFO

Available online 28 May 2013

Keywords:
Stated preference
Household fleet choice
MDCEV
Forecast
Residual
Validation

ABSTRACT

In sophisticated transport models, choice modelling is used to capture a wide range of behaviour, such as mode choice, vehicle choice and route choice. A newly developed approach to improving realism is the multiple discrete-continuous extreme value (MDCEV) model, which allows researchers to model the allocation of continuous amounts of a consumer good. Before implementing this model in overall frameworks, it is important to determine the accuracy of the forecasting. In this paper, an MDCEV model of household fleet choice based on data collected in a stated adaptation survey is presented. The model was used to predict the annual mileage of households with regard to 17 different types of cars, and the results of that forecast were compared to the actual data by calculating the residuals. The residual analysis showed that the model performed significantly better than a completely random model, but the share of wrongly allocated mileage, 70% of the total, remained high. However, the results of only one model were not sufficient to assess the procedure. The differences between two submodels, one with and one without public transport, regarding the distribution of the residuals indicated that model specification has a significant influence on performance. Therefore, more work on forecasting additional MDCEV models was necessary to have a basis for comparison. We compared two further MDCEV models to obtain a fuller understanding of their performance.

1. Introduction

In sophisticated transport models like the SACSIM model of the Sacramento metropolitan area in California, USA (Bradley et al., 2010), the ILUTE model of Toronto, Canada (Salvini and Miller, 2005) or the ALBATROSS model of the Netherlands (Beckx et al., 2009), choice modelling is used to capture a wide range of behaviour, including mode choice, household fleet choice and route choice. Discrete choice models in their standard formulations can integrate neither multivariate choices nor the associated continuous attributes of those choices. Such a capability would benefit research on a number of questions, however, as it would allow modellers to improve the realism of the description. One prime example is the composition of a set of mobility tools (Simma et al., 2002; Simma and Axhausen, 2003, 2001b,a) and the vehicles' associated mileage. A recent development of the MCDEV framework by Bhat (2005) offers a new approach to address this gap.

* Corresponding author. Tel.: +41 44 633 67 37; fax: +41 44 633 10 57.
E-mail addresses: boris.jaeggi@ivt.baug.ethz.ch (B. Jäggi), weis@ivt.baug.ethz.ch (C. Weis), axhausen@ivt.baug.ethz.ch (K.W. Axhausen).
[1] Tel.: +41 44 633 39 52; fax: +41 44 633 10 57.
[2] Tel.: +41 44 633 39 43; fax: +41 44 633 10 57.

1755-5345/$ - see front matter
http://dx.doi.org/10.1016/j.jocm.2013.04.005

A general transport model currently being developed at the Institute for Transport Planning and Systems (IVT) at the Swiss Federal Institute of Technology, Zürich in collaboration with the Technische Universität (TU) Berlin is MATSim (Balmer, 2007; Meister et al., 2009; Balmer et al., 2008), an agent-based micro-simulation tool for modelling travel demand and traffic flow. This paper is a part of ongoing work on implementing multiple discrete-continuous extreme-value (MDCEV) models into the model frameworks of various fields. Given how recently this approach was developed, our experience in its use is limited, and there is still little to be learned from the literature. Therefore, an evaluation of recent forecasting results and their residuals seems prudent.

In MATSim, travel demand is a generated activity based on using activity chains from the Mikrozensus (Swiss Federal Statistical Office, 2006), a Swiss national travel diary survey, which is conducted every five years. In the current version of MATSim, agents can conduct activities (e.g. staying at home, shopping, work, leisure, etc.) inside facilities (buildings). In an iterative solution process, the agents optimize their given activity chains. The agents are neither designated as members of households, nor do they have specific car types allocated to them yet. They only have an attribute that describes their car availability for mode choice processes.

In a future development of MATSim, the agents will be pooled into households and specific car types will be allocated to the households. This will not only enable energy consumption to be analysed on a microscopic level, but will also allow the implementation of a behavioural model to forecast the development of car usage, and as a result, energy consumption. The first step toward these enhancements is a car ownership model. We estimated such a model by applying the MDCEV approach developed by Bhat (2005) to the stated adaptation data collected in a survey conducted by Erath and Axhausen (2010). To test the performance of the model for an application like MATSim, we repeatedly applied a forecast of the MDCEV approach and analysed the residuals by comparing the results to the actual choices made in the survey.

Since 2004, when the MDCEV approach was originally developed to analyse time use (Bhat, 2005), various researchers have used it to estimate preferences. In his dissertation, Sen (2006) presented an MDCEV application for examining decisions of households regarding vehicle types and usage. See also Bhat and Sen (2006). The impact of demographics, built environment attributes, vehicle characteristics and petrol prices on the same issue was analysed by Bhat et al. (2009). Pinjari et al. (2009) analysed residential self-selection effects in time use models, and Spissu et al. (2009) presented an analysis of weekly out-of-home activity participation. Copperman and Bhat (2007) analysed the determinants of children's participation in weekend activities. Pinjari and Bhat (2010b) introduced the nested version of the MDCEV approach, the multiple discrete-continuous nested extreme value (MDCNEV) model, and presented an application for capturing non-workers' time use behaviour. A detailed description of the MDCEV approach and the role of its parameters can be found in Bhat (2008).

Pinjari and Bhat (2010a) have presented an efficient forecasting procedure for such models. We tested the performance of MDCEV forecasting results with our stated adaption data set on household fleet choice, energy savings in households and private transport, and induced demand. Unfortunately we were not able to find any useful literature on the disaggregate validation of (multiple) discrete-continuous models. In most MNL models, validation occurs at an aggregate level when actual and predicted market shares are compared, as researchers have done since MNL models were first introduced; see for example Train (1978). In the case of multiple discrete-continuous models, however, a disaggregate validation would be more valuable because it would give more insight into their characteristics.

2. Data

2.1. The survey

The primary data set used here was collected in the context of a project on long-term fuel price elasticity and the effects on mobility tool ownership and residential location choice (Erath and Axhausen, 2010), which was funded by the Swiss Federal Office of Energy and the Federal Office for the Environment. In the survey, 409 households were questioned about their long-term reactions to rising fuel costs. The survey was divided into a questionnaire on the respondents' socioeconomic situation and mobility tool ownership and a three-stage stated response survey. In the first stage, the respondents were presented with six fuel-price scenarios, ranging from CHF 1.5/l to CHF 5.5/l for petrol. The survey was conducted via face-to-face interviews. The interviewer was equipped with computer software that simultaneously calculated personalized mobility costs (with fixed costs kept separate from variable costs) based on previously collected personal information. The respondents could choose cars and the annual mileage for each chosen car to a high degree of detail, including the type of car, its engine size and drivetrain system, and whether they would buy a new or a used car, all the while being supported by the real-time calculations of the computer. Public transport season tickets, a popular alternative, were always an available option. Fig. 1 shows a screen shot of the survey.

Respondents could also choose and/or change the mileage travelled by public transport. In the second stage of the survey, respondents were given six different residential locations as well as varying fuel prices and were again asked to choose the preferred mobility tool (and mileage) for each situation. For the third stage of the stated preference survey, another six choice situations were presented. The choice sets consisted of two alternatives, one from each of the previous stages. The data used in this paper is from the first stage only.