Transportation Research Part A 43 (2009) 1–12

Contents lists available at ScienceDirect

Transportation Research Part A

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

A framework for analyzing rank-ordered data with application to automobile demand

John K. Dagsvik *, Gang Liu

Research Department, Statistics Norway, P.O. Box 8131, Dep. N-0033 Oslo, Norway

ARTICLE INFO

Article history:
Received 17 August 2007
Received in revised form 24 May 2008
Accepted 7 June 2008

Keywords:
Random utility models
GEV rank-ordered models
Car demand
Alternative fuel vehicles

ABSTRACT

In this paper, we develop a general random utility framework for analyzing data on individuals' rank-orderings. Specifically, we show that in the case with three alternatives one can express the probability of a particular rank-ordering as a simple function of first choice probabilities. This framework is applied to specify and estimate models of household demand for conventional gasoline cars and alternative fuel vehicles in Shanghai based on rank-ordered data obtained from a stated preference survey. Subsequently, the framework is extended to allow for random effects in the utility specification to allow for intrapersonal correlation in tastes across stated preference questions. The preferred model is then used to calculate demand probabilities and elasticities and the distribution of willingness-to-pay for alternative fuel vehicles.

1. Introduction

During the last three decades, there has been a rapid development of theoretical and empirical approaches to analyzing individual choice behavior of the demand for differentiated products, such as the choice among brands of cars, types of houses, etc. Important contributions in this area are due to McFadden and collaborators; see for example, McFadden (2001). Specifically, random utility models have been applied extensively to analyze urban travel behavior. As regards empirical behavioral analysis, the application of data obtained by means of stated preference (SP) type of surveys has become increasingly popular, see for example, Louviere et al. (2000), Brownstone et al. (2000), Calfee et al. (2001) and Potoglou and Kanaroglou (2007). Recall that by an SP survey, it is understood that individuals in a sample are exposed to hypothetical choice situations. SP data are useful in situations where market transactions are not available to reveal information about individual preferences. Contrary to the conventional revealed preference method, one important advantage of the SP method is that one can obtain several (hypothetical) choice observations for each respondent.

This paper makes two contributions. The main contribution is to show that for a general additive random utility model the probabilities for individuals' rank-orderings of alternatives in the case with three alternatives can be expressed as a simple function of first choice probabilities. This implies that if the first choice probabilities can be expressed on closed from, such as in the case of the generalized extreme value (GEV) random utility model, the corresponding ranking probabilities can also be expressed in a simple closed form. Second, we apply this framework to analyze the demand for conventional and alternative fuel cars in the city of Shanghai.

In China, the rapid increase in the demand for private cars is an important and sensitive issue. On the one hand, there is the expressed intention of the Chinese government to use the car industry as an engine to promote industrial and economic

* Corresponding author. Tel.: +47 95256437; fax: +47 21090040.
E-mail address: john.dagsvik@ssb.no (J.K. Dagsvik).

0965-8564/$ - see front matter
doi:10.1016/j.tra.2008.06.005

growth. On the other hand, one realizes the pressing need to adequately address serious pollution problems owing to car traffic in urban areas. There also appears to be a growing awareness within China about the role transportation sources play in increasing greenhouse gas emissions. Finally, there is the concern that an uncontrolled increase in the number of private cars may lead to very serious congestion problems. Traffic problems in a number of large cities in developing countries may serve as a warning of what may happen if the increase in private cars in China is not kept under control.

To the authors' best knowledge, studies on car demand undertaken in China are based on historical aggregated data, and these studies are mostly only loosely founded on microeconomic theory, and do not base their analysis on explicitly formulated behavioral models, see for example, Guo (2001) and Zhai (2000). In contrast, the empirical analysis conducted in this paper represents a first attempt to undertake a behavioral empirical study of the demand for cars in the city of Shanghai, and it is based on micro data and the theory of discrete choice. The data are obtained from a SP survey collected in Shanghai during the summer of 2001. The survey approach we follow is similar to Dagsvik et al. (2002). Specifically, in our survey each household is presented with 15 choice experiments and is asked to rank-order several hypothetical car alternatives characterized by car-specific attributes (price, size, power, fuel consumption) that vary from one choice experiment to another. We apply the collected data to estimate several model versions within the framework developed here. In the first model version the preferences are assumed independent across experiments, but allowed to be correlated across alternatives. Subsequently, we introduce random effects in the utility specification to allow for time invariant unobserved population heterogeneity in preferences. The estimation results show that this type of heterogeneity is important. Unfortunately, the sample is rather small, and we have therefore only specified and estimated models with rather limited observed population heterogeneity. This is clearly unsatisfactory, and it is desirable to obtain a larger sample in future research.

The behavioral car demand model estimated in this paper enables the prediction of demand and the computation of demand elasticities with respect to price and other car attributes conditional on car attributes. It also allows us to calculate welfare measures such as willingness-to-pay for alternative fuel vehicles (AFV).

The remainder of the paper is structured as follows: in section 2 the theoretical results about rank-order probabilities are obtained. Section 3 describes the survey method and the data. In section 4, the empirical specification of the different models as well as the estimation results are presented. In section 5, we present the results on demand and their elasticities. In section 6, we use the model to calculate willingness-to-pay estimates for alternative fuel vehicles.

2. The relationship between first choice- and rank-ordered probabilities in the case with three alternatives

To analyze data on the rank-ordering of alternatives, a particular methodological framework is required. The development of choice models for rank-ordered data originated with work by Luce (1959), Block and Marschak (1960) and Luce and Suppes (1965), whereas Beggs et al. (1981) represents an early application of such models to SP data with observations on the potential demand for electric vehicles.

In this section we shall derive the probabilities of specified rank-orderings when the set of feasible alternatives contain three elements. Previous models for rank-ordering data are often based on the assumption that the random error terms of the utility function are independent across alternatives. A particularly simple expression for the probability of a specific rank-order follows readily when these error terms are i.i. extreme value distributed (see for example, Beggs et al., 1981). For general random utility models, however, there does not seem to be a simple closed form expression for the ranking probabilities. In this section we show that in the case with three alternatives one can express the ranking probabilities as a simple function of the first choice probabilities. This is of interest in cases where the first choice probabilities can be expressed on closed form, such as in the case when the random terms of the utilities are multivariate extreme value distributed, (the GEV model) because in this case simple closed form expressions for the first choice probabilities exist. Consequently, within the GEV class the corresponding probabilities for rank-orderings follow.

Let U_j denote the utility of alternative, $j = 1,2,3$. We assume that $U_j = v_j + \varepsilon_j$, where v_j is a deterministic component and ε_j, $j = 1,2,3$, are random terms (taste shifters) with joint cumulative distribution function (c.d.f.) that is independent of $\{v_j\}$ and is continuously differentiable.

Consider the probability that alternative 2 is ranked on top, alternative 1 is the second and alternative 3 is the third preference. Note that since the statement $\{U_1 > \max(U_2, U_3)\}$ can be expressed as $\{(U_2 < U_1) \cap (U_1 > U_3)\}$, it follows that:

$$\{U_1 > \max(U_2, U_3)\} \cup \{U_2 > U_1 > U_3\} \Longleftrightarrow \{(U_2 < U_1) \cap (U_1 > U_3)\} \cup \{(U_2 > U_1) \cap (U_1 > U_3)\} \Longleftrightarrow \{U_1 > U_3\}.$$

From this it follows immediately that:

$$P(U_1 > \max(U_2, U_3)) + P(U_2 > U_1 > U_3) = P(U_1 > U_3). \tag{2.1}$$

Consequently, (2.1) implies that

$$P(U_2 > U_1 > U_3) = P(U_1 > U_3) - P(U_1 = \max_{q \leqslant 3} U_q).$$

Similarly, we obtain that

$$P(U_j > U_k > U_r) = P(U_k > U_r) - P(U_k = \max_{q \leqslant 3} U_q) \tag{2.2}$$