Transportation Research Part A 116 (2018) 130–150

Contents lists available at ScienceDirect

Transportation Research Part A

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

Taste heterogeneity as an alternative form of endogeneity bias: Investigating the attitude-moderated effects of built environment and socio-demographics on vehicle ownership using latent class modeling

Sung Hoo Kim, Patricia L. Mokhtarian*

School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

ARTICLE INFO

Keywords:
Latent class model
Vehicle ownership
Taste heterogeneity
Self-selection
Success table
Endogeneity bias

ABSTRACT

Vehicle ownership (VO) is of vital interest to transportation planning and policy, both in its own right and as a correlate of other travel-related behaviors. This study explores disaggregate relationships among socio-economic/demographic (SED) traits, the built environment (BE), and VO. Many previous studies have assumed that the importance of SED and BE variables to VO is homogeneous across the population, and have focused on the direct and mediated effects of those variables on VO. Here, we aim to account for heterogeneity in the effects of BE and SED, allowing those effects on VO to be moderated as a function of attitudes. Specifically, we use Latent Class Modeling (LCM), which probabilistically segments the sample so as to be homogeneous within and heterogeneous across segments, with respect to the choice process. Applied to a sample of 2385 commuters in Northern California, LCM outperforms an ordinary multinomial logit model and a deterministic segmentation model. The LCM identifies two classes: an "auto-oriented" segment, for which household size and income have stronger influences on vehicle ownership (compared to the other segment); and an "urbanite" segment, for which the BE generally has a stronger influence. This study contributes to better understanding the heterogeneity of an important travel-related choice process, and offers an alternative explanation to residential self-selection for an attitude-related endogeneity bias.

1. Introduction

Vehicle ownership (VO) is undoubtedly a key consideration when it comes to regional planning and transportation policies. In the realm of behavioral research, owning a vehicle is not only an important choice in itself, but it is also a major factor in other travel behaviors. For example, VO is strongly related to travel/activity patterns in the short term, and to residential location in the long term (Eluru et al., 2010; van Wee, 2009). For prediction purposes, VO could be modeled at either the aggregate or disaggregate level. However, most recent studies (including this one) have focused on disaggregate models because of their superior ability to identify causal relationships (Bhat and Pulugurta, 1998; Potoglou and Susilo, 2008). Readers can refer to de Jong et al. (2004) or Ortúzar and Willumsen (2011) for aggregate models.

In disaggregate models, numerous variables have been employed to explain VO. Socio-economic/demographic (SED) traits are fundamental. Household SED characteristics include size (e.g. numbers of members, children, adults, workers, or license holders) and

* Corresponding author.
E-mail addresses: skim3020@gatech.edu (S.H. Kim), patmokh@gatech.edu (P.L. Mokhtarian).

https://doi.org/10.1016/j.tra.2018.05.020
Received 10 November 2017; Received in revised form 10 April 2018; Accepted 31 May 2018
0965-8564/

income; the consensus is that number of vehicles tends to increase as either of these increases. Another set of key explanatory variables includes built environment (BE) attributes (or residential location). Some studies included BE as subjective/perceived attributes (Cao et al., 2007; Gao et al., 2008), while others measured it using objective attributes. The BE attributes studied range from simple binary indicators (e.g. urban vs. suburban area) to detailed area characteristics (e.g. population density, transit frequency). Recent papers dealing with relationships between BE and travel-related behavior (including VO) mainly rely on five types of "D variables" – density, diversity (non-automobile-mode-oriented) design, destination accessibility, and distance to transit (Cervero and Kockelman, 1997; Ewing and Cervero, 2010; Kockelman, 1997). The hypothesis is that VO will tend to decrease as the first four Ds increase and/or the fifth decreases, reflecting the reduced reliance on the car made possible by increasing the proximity to desired activities and the attractiveness of alternative means of travel. Empirical results generally support this hypothesis.

With respect to the relationship between travel behavior and BE, there has been an extensive discussion of residential self-selection (e.g. Mokhtarian and Cao, 2008; Cao et al., 2009). In this context, self-selection can be broadly defined as "the tendency of people to make choices that are relevant for travel behavior, based on their abilities, needs and preferences" (van Wee, 2009, p. 280). If someone self-selects into a certain residential location, the true effect of BE could differ from the effect predicted by a model that does not account for that bias. Since most models control for SED sources of self-selection (such as income), the key concern is for attitudinal sources, which are often unobserved.

The present paper positions residential self-selection (RSS) as one "branch" on the "endogeneity bias tree", and taste heterogeneity (TH) as a different branch of the same tree. We focus on capturing TH as a way of resolving an endogeneity bias in estimating the influence of the BE on VO. Specifically, we hypothesize that attitudes influence the *impact* of the BE on VO (i.e., the *coefficients* of BE variables in a model of VO). In particular, we expect that a certain increase in residential density will offer a stronger disincentive to own (more) cars for someone who likes transit, active transportation, and a denser/more diverse neighborhood than for someone who is the opposite.

A few other studies have dealt with TH in the context of self-selection. Kamruzzaman et al. (2013) deterministically segmented their sample into transit-oriented development (TOD) dwellers and non-TOD dwellers, while Salon (2015) deterministically segmented based on geographically-defined residential location clusters. In the latter case, the segmentation was an integral component of the sample selection approach to treating RSS, involving the insertion of additional terms into the outcome models for each residential location segment, which corrected the estimated model coefficients for self-selection into the segment. To our knowledge, this self-selection correction approach is only associated with deterministic segmentation, where the segments represent various treatment (and/or control) states. Choi et al. (in preparation) also deals with TH as an alternative paradigm to self-selection and is something of a companion to the present paper. However, it investigates another travel outcome (vehicle-miles driven), accordingly (since the dependent variable is continuous, rather than discrete as in the present case) employs a different model (latent class linear regression), and uses an entirely different dataset. As far as we are aware, the present paper offers a more systematic treatment of the conceptual relationship between RSS and TH than can be found elsewhere to date.

The rest of the paper is organized as follows. Section 2 points to some key literature pertaining to VO models and TH, particularly latent class modeling (LCM), and presents a typology of approaches to analyzing TH. Section 3 describes the two datasets used in the study, including their key variables. Section 4 delineates the modeling framework – explaining how self-selection and TH are distinct examples of endogeneity bias – and presents the LCM approach to accounting for TH. In Section 5, we examine the model results and describe the segments identified by the model. Section 6 compares the LCM results to those obtained from deterministically-segmented and pooled models, accompanied by a discussion of which of four possible success tables to use in judging a model's predictive ability. Section 7 summarizes the results and discusses some limitations, implications, and directions for future research.

2. Literature review

A number of functional forms have been used to model VO, including choice models (e.g. probit, or binary/multinomial, ordered, and nested logit), count models (e.g. Poisson and negative binomial regression), and structural equations models. We do not describe the literature in detail here, because readers can refer to the comprehensive summary in Anowar et al. (2014a). In particular, multinomial logit (MNL) and ordered logit are the most frequently used, and several papers have compared the two models to determine which is better for this purpose (Bhat and Pulugurta, 1998; Potoglou and Susilo, 2008). Although ordered logit produces more parsimonious solutions, its associated "parallel lines" assumption is often invalid (Brant, 1990; Fullerton and Xu, 2012), and MNL generally outperforms it on goodness of fit, even after penalizing MNL for involving more parameters. Accordingly, although we also explored other functional forms, we ultimately chose MNL as the base model for this study.

TH (or taste variation) refers to the population variability of coefficients in models of behavior or decision processes, in contrast to the default "one size fits all" assumption of constant coefficients for everyone. The importance of accounting for TH has been underscored by a large body of studies (Bergantino et al., 2013). In some literatures, TH is more often referred to as a *moderation effect* on the explanatory variables of interest (e.g. Wu and Zumbo, 2007). Fig. 1 illustrates a typology of approaches to treating TH. The approaches can be classified based on whether the allowed taste variations are continuous or discrete, and whether the identification of segments is deterministic or random. Deterministic segmentation approaches have been in use for a very long time (see, e.g., Kitamura, 1981; Ben-Akiva and Lerman, 1985), so the present discussion focuses on the relatively newer stochastic segmentation approaches: random coefficients and latent classes.

Many studies have employed mixed logit to capture random TH (e.g. Bhat and Guo, 2007; Bergantino et al., 2013). All such random coefficients models, however, require prior assumptions about the mixture distribution for each coefficient, not to mention assumptions about the correlations among coefficients (Vij et al., 2013). In addition, simply modeling coefficients as randomly