Transportation Research Part A 70 (2014) 281–293

Contents lists available at ScienceDirect

Transportation Research Part A

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

Heterogeneity among motorists in traffic-congested areas in southern California

J.F. Jennifer Lee [a,*], Peter K. Kwok [b], Jeffrey Williams [a]

[a] Department of Agricultural and Resource Economics, University of California Davis, One Shields Avenue, Davis, CA 95616, USA
[b] Peter K. Kwok, 1640 Maple Avenue, Unit 1001, Evanston, IL 60201, USA

ARTICLE INFO

Article history:
Received 26 August 2013
Received in revised form 15 August 2014
Accepted 16 October 2014
Available online 19 November 2014

Keywords:
Heterogeneity
Survival analysis
Remote sensing data
Census proxy
Traffic congestion
Southern California

ABSTRACT

Estimation of congestion costs, presumed to be one of the largest external costs of automobile travel, is typically based on a single value of time delay for motorists in metropolitan areas. However, the estimation may be wrong if the profiles of motorists are different at different times of day. This study uses a survival model to examine the demographic and socioeconomic profiles of motorists at different times of day at congested locations in southern California, by using on-road remote-sensing measurements and license plates images obtained in 2007 and 2008 by the California South Coast Air Quality Management District (SCAQMD). More than 80,000 vehicles were observed from fifteen selected study sites over fifteen days. Their plates, through anonymized registration records, revealed addresses at the census block group level, which have homogenous profiles by construction. Motorists' profiles at different times of day display large variation, however, according to extended Cox model with a non-parametric baseline hazard, which is used to accommodate both the time-invariant and time-varying effects of the covariates. This study thus proposes a new approach to examine heterogeneity among motorists.

1. Introduction

Traffic congestion is one of the largest externalities related to automobile travel (other externalities include accidents and pollution) (Delucchi and McCubbin, 2011; Lemp and Kockelman, 2008; Parry et al., 2007). The major contribution to congestion costs is the cost of time delay (Grant-Muller and Laird, 2007; Hanks and Lomax, 1991; Schrank et al., 2011), which is typically estimated based on a typical regional wage rate. Estimation based on a single value of time delay for a homogenous population may over- or under-estimate the congestion costs. When the profiles of motorists are different at different times of day, the value of travel time will also vary during the course of a day. Better understanding of the heterogeneity among motorists would also allow better modeling of the temporal distribution of travel demand, better assessment of transportation management, and better pricing policies, whether congestion pricing, parking prices, or distance-based fees, and thus contribute to better decisions (Small et al., 2005, 2006; Verhoef and Small, 2004).

Past research on travel behavior has examined the travel behavior of specific demographics, such as groups defined by age (the elderly), gender, income, race, and social disadvantage (for example, Dodson et al., 2010; Evans, 2001; Gardenhire and Sermons, 2001; Georggi and Pendyala, 2001; Mallett, 2001; Niemeier and Morita, 1996; Polzin et al., 2001; Rosenbloom and

* Corresponding author. Tel.: +1 (530) 220 3882.
E-mail addresses: jfjlee@ucdavis.edu (J.F.J. Lee), pkyk2001@yahoo.com (P.K. Kwok), williams@primal.ucdavis.edu (J. Williams).

http://dx.doi.org/10.1016/j.tra.2014.10.011
0965-8564/

Waldorf, 2001). These studies are all based on surveys, which were collected from each household over a long period of time. Thus, a comprehensive picture of the motorists on roads at any one moment is usually unknown.

Only a few empirical studies have incorporated motorists' heterogeneity in transportation policy design, in particular transportation pricing strategies. Verhoef and Small (2004) model the properties of various types of public and private pricing on a congested road network with heterogeneous users. They find that revenue-maximizing pricing is much less efficient than welfare-maximizing pricing, but the difference is mitigated by the product differentiation made possible with heterogeneous users. Small et al. (2006) use surveys to collect data of people traveling on the congestion pricing segment of the SR 91 (SR 91 express lanes) in Orange County, and propose a differentiated road-pricing scheme based on variation among highway users. They model three choices by motorists, including whether to acquire a transponder required for using the SR 91 express lanes, whether to travel on the express lanes or the free lanes, and whether to have more than one occupant. Their results show that implementing various highway lane pricing schemes can have social welfare gains, and that the road prices can be adjusted to reflect motorists' varying behavior. However, they acknowledge that because the lack of information on how congestion varies on roads besides the SR 91 study corridor, they are not able to model the choice of what time of day people travel (Small et al., 2006). Small et al. (2005) find that motorists who use the SR 91 express lanes exhibit substantial heterogeneity in value of time and reliability by using the same survey data as Small et al. (2006). Winston and Yan (2011) investigate the potential for public highway privatization by considering motorists' widely varying values of travel time and reliability. They find that highway privatization can benefit road users and increase social welfare. A private operator(s) may respond to the heterogeneity among road users, by allocating capacity and charging tolls that result in differentiated services, in ways that public highway authorities have not (Winston and Yan, 2011).

Several studies have analyzed the user demographics of roads with congestion pricing. A study of SR 91 express lane users by Sullivan (2000) shows that the proportion of commuters who chose the SR 91 express lanes over the free lanes is higher for females than for males, and intermediate age categories are likely to be frequent toll lane users (Sullivan, 2000, pp. 80–81). A survey by Supernak et al. (2002) concludes that I-15 congestion pricing road users were "from higher-income households, more highly educated, predominantly 35–54 years old, more likely to be homeowners, more likely to be middle-aged women, and from two-vehicle households" compared with other I-15 users. Insights Worldwide Research (2009, p. 16) characterizes the survey respondents of the SR 91 express lanes "are likely to be in their early 50 s, male, and with some college education." Among the survey respondents, over 70 percent were Caucasian (Insights Worldwide Research, 2009, p. 16). Although this degree of heterogeneity does not seem surprising, very little is known about the demographics of travelers affected by congestion on roads without pricing.

This study analyzes the demographic and socioeconomic characteristics of highway road users in congested areas of southern California at different times of specific days without relying on survey data. The study exploits an existing data source in conjunction with linking to other records, to yield the data needed for our analysis. In particular, it uses on-road remote-sensing measurements and images of license plates, which are connected to anonymized registration records. The unavailable demographic and socioeconomic profiles of the vehicles' owners are supplemented with Census data. More than 80,000 vehicles were observed at fifteen selected study sites over fifteen days.

Different kinds of estimation models, depending on whether to treat the time variable as discrete intervals or as continuous, can be used to examine the varying profiles of motorists by time of day. The profiles can be measured by count or by risk, namely, the proportion in the study population. Because traffic volumes vary across sites due to different travel speeds and different neighborhoods, a risk model should give a more informative description of usage than a count model. A fast increase (or decrease) of the proportion of site users associated with a particular attribute would indicate a surge (or drop) of their demand relative to other users at that site with different attributes. Accordingly, a survival model is proposed to analyze the varying demographic and socioeconomic profiles of motorists during the course of a day for the case of weekdays in Los Angeles, Orange, and Riverside counties in California.

This study presents three major contributions. (1) This is the first study to use on-road data to analyze the profiles of motorists. Previous empirical studies of travel behavior rely heavily on data collected from household travel surveys based on questionnaires. However, surveys suffer from inherent biases and the sample sizes of surveys are usually small due to associated costs. (2) This study demonstrates the use of a continuous time model to examine the varying socioeconomic profiles of motorist at different times of day, instead of using discrete time choice models, such as logit models. (3) The study analyzes the profiles of motorists in traffic congested areas in the Los Angeles/Long Beach/Santa Ana metropolitan region, which could help policymakers and transportation planners to devise more effective strategies to mitigate traffic congestion.

1.1. Background of survival analysis

Survival analysis, also called duration analysis, transition analysis, or failure-time analysis, deals with the time until the occurrence of an event, often called a failure. Survival analysis has been used traditionally in biometrics to study the length of time patients survive, and in engineering to determine the factors influencing the length of time until machine parts fail. However, survival analysis has found many applications that are not failure-related, for example, the duration of purpose-specific activities, such as shopping trips. The appropriateness of survival analysis, broadly defined, depends on the statistical properties of the situation.

The use of duration models in the transportation field started in the late 1980s (Hensher and Mannering, 1994). Bhat (2000), who reviews the use of duration models in the transport field since the mid-1990s, notes that applications to activity