Transportation Research Part A 78 (2015) 161–177

Contents lists available at ScienceDirect

Transportation Research Part A

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

Factors affecting public transportation usage rate: Geographically weighted regression

Yu-Chiun Chiou [a,*], Rong-Chang Jou [b], Cheng-Han Yang [a]

[a] Department of Transportation and Logistics Management, National Chiao Tung University, 4F, 118 Sec. 1, Chung-Hsiao W. Rd., Taipei 10012, Taiwan, ROC
[b] Department of Civil Engineering, National Chi Nan University, No. 1, University Rd, Puli, Nantou County 54561, Taiwan, ROC

ARTICLE INFO

Article history:
Received 22 December 2014
Received in revised form 24 April 2015
Accepted 21 May 2015
Available online 7 June 2015

Keywords:
Public transportation usage rate
Tobit regression
Geographically weighted regression
Spatial autocorrelation

ABSTRACT

As the number of private vehicles grows worldwide, so does air pollution and traffic congestion, which typically constrain economic development. To achieve transportation sustainability and continued economic development, the dependency on private vehicles must be decreased by increasing public transportation usage. However, without knowing the key factors that affect public transportation usage, developing strategies that effectively improve public transportation usage is impossible. Therefore, this study respectively applies global and local regression models to identify the key factors of usage rates for 348 regions (township or districts) in Taiwan. The global regression model, the Tobit regression model (TRM), is used to estimate one set of parameters that are associated with explanatory variables and explain regional differences in usage rates, while the local regression model, geographically weighted regression (GWR), estimates parameters differently depending on spatial correlations among neighbouring regions. By referencing related studies, 32 potential explanatory variables in four categories, social-economic, land use, public transportation, and private transportation, are chosen. Model performance is compared in terms of mean absolute percentage error (MAPE) and spatial autocorrelation coefficient (Moran' *I*). Estimation results show that the GWR model has better prediction accuracy and better accommodation of spatial autocorrelation. Seven variables are significantly tested, and most have parameters that differ across regions in Taiwan. Based on these findings, strategies are proposed that improve public transportation usage.

1. Introduction

Although travellers typically differ in their transportation choices, private transportation has the advantages of better accessibility and convenience than public transportation, which have engendered rapid growth in the number of private vehicles, increasing both traffic congestion and environmental pollution. Many governments worldwide have focused on improving their public transportation services and restricting ownership and usage of private vehicles. However, public transportation patronage in many developed countries with adequate public transportation systems remains low. Thus, whether governments should develop their public transportation services has not been thoroughly explored. In literature, two main approaches are applied to conduct this exploration: individual and collective approaches. The individual approach

* Corresponding author. Tel.: +886 2 23494940.
E-mail address: ycchiou@mail.nctu.edu.tw (Y.-C. Chiou).

http://dx.doi.org/10.1016/j.tra.2015.05.016
0965-8564/

developed mode choice models based on data from a questionnaire survey of travellers. Covariates of travel time (e.g. walking time, waiting time, and on-board time) and travel cost (e.g. parking cost, fuel cost, and bus fare) of different transportation modes, along with demographics and trip characteristics of travellers, are commonly considered. Estimation results by the individual approach provide insights into the effects of covariates on travellers' mode choice behaviours, which is helpful when designing public transportation systems. However, aggregation of individual mode choice behaviours to represent regional usage rates usually leads to biased predictions due to small sample sizes or excessive sampling errors. In contrast, the collective approach directly correlates ridership or usage rate of public transportation to potential covariates, which is helpful for decisions about whether to invest in public transportation systems in specific regions in order to increase the market share of public transportation, but individual preferences for public transportation systems are hard to be predicted based on the collective models. Therefore, to support decisions about investing in and designing public transportation systems, these two approaches are necessary. The collective approach can investigate the necessity of public transportation investment in various regions, while the individual approach can support decisions regarding the design of public transportation systems. Although the individual approach has several advantages over the collective approach, especially at the level of details in associating the mode choice changes with the improved service quality of public transportation, the commonly adopted explanatory variables of the individual approach, such as travel time, waiting time, walking distance of various transportation systems and social-demographics and trip characteristics of individual travellers, are rather expensive to be collected for model estimation and prediction for a large-scale area. To assess governmental investment in public transportation in various regions nationwide, this study adopts the collective approach to identify the key factors contributing to usage rate of public transportation systems.

To objectively review the benefit of investment in public transportation, a demand-side response should be evaluated scientifically and systematically to assist decision-makers in fully utilizing limited resources towards policy objectives. This information can also be used as feedback to amend short- and long-term strategies as well as avoid ineffective investments. Previous studies on aggregate public transportation usage rate modelling were limited to large unit (e.g., countries or cities) and thereby did not examine differences in inter-regional influential factors. Moreover, since the service level of transportation systems in neighbouring regions may be closely related (i.e. spatial autocorrelation), this study applies the geographically weighted regression (GWR) model to investigate the effects of key factors on public transportation usage rates locally rather than globally. The GWR model allows estimated parameters to vary across regions to accommodate potential spatial dependencies. Comparisons of model performance and estimation results with traditional global regression models are also conducted. As the dependent variable in this study is a percentage, which is truncated and does not have a normal distribution, the Tobit regression model (TRM) is used.

The rest of this paper is organized as follows. Section 2 presents a brief review of prior research. Section 3 briefly introduces the TRM and GWR models used in this paper along with performance indices for model comparisons. Section 4 addresses empirical data and their descriptive statistics. Section 5 presents estimation results along with implications of the two models. Model comparisons are given in Section 6. Section 7 presents conclusions and suggestions for further research.

2. Prior research

The models used to identify key factors contributing transit ridership and usage rate along with contributing factors considered are in previous studies are briefly reviewed.

2.1. Global and local regression models

Numerous studies have been conducted to examine the effects of key factors to transit ridership or usage rate globally or locally. The global regression models generally treat each observation independently and estimate one set of parameters to represent all observations. The commonly adopted models depend upon the dependent variables: transit ridership (i.e. number of boarding) and public transportation usage rate. For transit ridership, multiple regression analysis (Cervero, 1996; Kuby et al., 2004; Taylor et al., 2009; Mulley and Tanner, 2009; Swimmer and Klein, 2010; Blainey, 2010; Cervero et al., 2010; Souche, 2010) and simultaneous regression (Taylor et al., 2009; Swimmer and Klein, 2010) were commonly adopted. For public transportation usage rate, various models were constructed, including multiple regression analysis (Chow et al., 2006; Messenger and Ewing, 2007), simultaneous regression (Messenger and Ewing, 2007), the aggregate logit model (Buehler, 2011; Coldren et al., 2003; Chen et al., 2009), and the Tobit regression model (Boame, 2004) and so on.

Fotheringham et al. (2000) noted that the global regression model only estimates a set of parameters for relationships between an independent and dependent variables, and the estimated parameters do not vary with space therefore reflect spatial characteristics. That is, the major drawback of the models is that any geographical variation in the relations between variables is masked (Lloyd and Shuttleworth, 2005; Cardozo et al., 2012) by ignoring the existence of local variations due to the spatial autocorrelation.

In contrast, the local models, such as spatial proximity regression, GWR, distance-decay weighted regression, hierarchical linear model, are recommended when spatial autocorrelation is present. Several studies have been conducted for