ARTICLE IN PRESS

Transport Policy xxx (2017) 1–8

Contents lists available at ScienceDirect

Transport Policy

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

Trip distribution model for regional railway services considering spatial effects between stations

Rubén Cordera *, Roberto Sañudo, Luigi dell'Olio, Ángel Ibeas

Department of Transportation, Universidad de Cantabria, Av. de Los Castros s/n, 39005 Santander, Cantabria, Spain

ARTICLE INFO

Keywords:
Railway
Trip distribution models
Poisson regression
Spatial filtering

ABSTRACT

The railways are a priority transport mode for the European Union given their safety record and environmental sustainability. Therefore it is important to have quantitative models available which allow passenger demand for rail travel to be simulated for planning purposes and to evaluate different policies. The aim of this article is to specify and estimate trip distribution models between railway stations by considering the most influential demand variables. Two types of models were estimated: Poisson regression and gravity. The input data were the ticket sales and the prices between stations on a regional line in Cantabria (Spain) which were provided by the Spanish railway infrastructure administrator (ADIF – RAM). The models have also considered the possible existence of spatial effects between train stations. The results show that the models have a good fit to the available data, especially the gravity models constrained by origins and destinations. Furthermore, the gravity models which considered the existence of spatial effects between stations had a significantly better fit and provided a more realistic journey pattern in a future scenario than the Poisson models and the gravity models that did not consider these effects. The proposed models have therefore been shown to be good support tools for decision making in the field of railway planning.

1. Introduction

The European Commission transport roadmap (European Commission, 2011) gives priority to the railways because of their proven safety and environmental sustainability compared to road transport. One of the Commission's stated future goals is the creation of a unique European railway space, the introduction of new technological solutions and the construction of new infrastructure financed and priced intelligently.

In order to reach these goals, the European Commission has highlighted the need to evaluate transport projects to guarantee their social profitability and the added value they give to the EU. This evaluation needs to be supported by the available evidence and transport demand models which allow user behaviour to be accurately simulated.

Among the group of transport demand models are trip distribution models which allow the interaction between origin and destination points to be simulated. The most well-known and widely used distribution model has traditionally been the gravity model which, based on the analogy with Newtonian physics, has later been theorized from a probabilistic perspective as a maximum entropy model (Wilson and Bennett, 1985). The state of the art provides many calibration techniques for the parameters of both origin and destination as well as for travel cost (Ortúzar and Willumsen, 2011). Other researchers have insisted on the need to use Poisson type regression models given the discrete and non-negative nature of the journeys (Flowerdew and Aitkin, 1982).

This article proposes the estimation of trip distribution models based on the boarding and alighting data of passengers on a regional railway line. The data used has been obtained from ticket sales on the line provided by the Spanish railway infrastructure administrator (ADIF – RAM). The models were estimated based on two methods: a Poisson type nonlinear regression without any kind of constraint and a Wilson type gravity model doubly constrained to origins and destinations. Both types of models are compared by considering their goodness of fit with the data, in order to determine if the greater number of parameters estimated in the gravity models really does provide greater significance. The models have also been estimated with additional variables to consider the existence of spatial effects between stations to determine if these effects are significant and increase the explanatory capability of the models. Finally, the models have been applied to a rising demand future scenario to check their performance. The results show that gravity models restricted to origins and destinations with additional variables which consider spatial effects like contiguity between stations have a significantly better goodness of fit to the data and provide a more realistic

---

* Corresponding author.
E-mail address: corderar@unican.es (R. Cordera).

https://doi.org/10.1016/j.tranpol.2018.01.016
Received 22 July 2016; Received in revised form 21 June 2017; Accepted 18 January 2018

Please cite this article in press as: Cordera, R., et al., Trip distribution model for regional railway services considering spatial effects between stations, Transport Policy (2017), https://doi.org/10.1016/j.tranpol.2018.01.016

journey pattern in the future scenario.

A brief review of the state of the art in the field of trip distribution models and distribution models applied to the railways is presented in the following section. The methodology followed is summarised in Section 3 concentrating on Poisson type regression models and doubly constrained gravity models. Section 4 provides a description of the study area and presents and discusses the results obtained by the models. Finally, the conclusions drawn are summarised in Section 5.

2. State of the art about trip distribution models

Spatial interaction models were applied very early on in multiple fields of study for simulating the effects of spatial interaction such as the movement of people between urban areas (Ravenstein, 1885) or commercial flows (Huff, 1959). The first models proposed were based on an analogy with Newtonian gravity theory with the sizes of origins and destinations and the distances between them as explanatory variables. This type of model has a reasonably good fit to the data although they lacked theoretical justification. The theoretical base was provided by Wilson (1970) who showed the possibility of deriving a great number of models from the principle of maximum entropy by which the most probable distribution matrix is the one which maximises the microstates of a given macrostate (Fotheringham et al., 2000). Cochrane (1975) later proposed a derivation of the gravity model from the principle of utility maximization. Exponential gravity models, as derived by Wilson, have proven to be particularly useful for precisely representing the macro level behaviour of a wide variety of micro level interactions (Sen and Smith, 2012). Other authors have later insisted on the convenience of using Poisson type non-linear regression models given their greater adaptability to the trip generation and distribution phenomena (Flowerdew and Aitkin, 1982; Winkelmann and Zimmermann, 1995).

The currently available trip distribution models can be classified into two large groups depending on the data used: models based on aggregate data which uses, for example, ticket sales information and models based on surveys which use disaggregate data on an individual level. Cascetta et al. (2007) also differentiated mixed distribution models which incorporated characteristics of both aggregate and disaggregate models.

The specification of the travel cost function plays a key role in the distribution models so that the predictions fit as closely as possible with the distribution of the observed journeys (Tiefelsdorf, 2003). The most commonly used functional forms in practise are the potential, the exponential and the combined (also known as the Tanner deterrence function) (Cascetta, 2009). While the combined travel cost function is the most appropriate for urban environments where an increase in journeys may occur for small travel costs, in more extensive environments the potential and exponential functions should give a better fit. Travel cost is usually represented through a generalised cost which may include variables like journey time and fare. In the case where the generalised cost is expressed in terms of money, the journey time parameter can be interpreted as the value of time for users (Ortúzar and Willumsen, 2011).

Distribution models have been widely applied in the field of transport planning. Wang et al. (2016) applied a distribution model based on linear regression to the journeys obtained from the entrance and exit validations of contactless tickets at stations on the Beijing metro (China). The model allowed the authors to provide estimations about how journey distance and land use distribution affected journey patterns without having to estimate a complete four stage transport model. However, the authors did not consider the possible existence of spatial effects in the data, which is something that could affect the estimated parameters. In contrast, de Grange et al. (2011) estimated a gravity type distribution model considering spatial correlation. The model was estimated with data from the bus service of the city of Santiago (Chile). The authors concluded that explicitly considering spatial effects in a gravity model could significantly increase its explanatory and predictive capabilities.

In the field of trip distribution models relating specifically to railways, these models allow different planning alternatives to be evaluated. Among the aggregate models based on ticket sales, Wardman (2006) proposed an unrestricted generation-distribution model using time series data for the United Kingdom in the 1990s. The estimated model presented variables corresponding to the characteristics of the origin such as the population, GDP and the rate of motorisation, as well as to the journey such as the overall cost. The author found that GDP was the most important factor in explaining the growth of journeys, even though in a complete four stage model these types of variables are usually introduced into trip generation models. In a similar work applied to railway journeys to and from airports, Lythgoe and Wardman (2002) estimated a demand model based on linear regression which calculated elasticities for different variables like GDP, the fare or journey time.

Where disaggregate data is available, models based on user surveys allow researchers to simulate individual choices considering personal characteristics (age, gender, income, etc.) and transport service characteristics as well as origins and destinations (Ben-Akiva and Lerman, 1985). However, this type of disaggregate model based on random utility theory require greater effort during the data collection phase because they are generally estimated using fewer data than models based on ticket sales.

3. Methodology

Different authors have highlighted the specification problems involved in using a multiple linear regression model (MLR) to estimate the generation and distribution of journeys in a study area (Flowerdew and Lovett, 1988; Thill and Kim, 2005). The dependent variable in distribution models is of a discrete nature, whereas the MLR model assumes a continuous distribution. Therefore, it is desirable to use a model specified with a qualitative dependent variable such as the Poisson regression model (Gujarati and Porter, 2009). This model takes the form:

$$P(Y_i) = \frac{\mu^Y e^{-\mu}}{Y!} \tag{1}$$

The Poisson regression assumes that each dependent variable Y_i is extracted from a Poisson type discrete distribution with the distribution parameter μ_i of (1), logarithmically linked to a linear combination of explanatory variables:

$$\ln(u_i) = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_k X_{ki} \tag{2}$$

Where:

β_k are parameters to be estimated
X_{ki} are the independent variables

The Poisson model cannot be made linear, meaning that the parameters cannot be estimated using Ordinary Least Squares (OLS). Alternative estimation methods such as maximum likelihood implemented through algorithms as reweighted least squares have been proposed, producing good results (Green, 1984).

A particular case of the Poisson model appears when all the independent variables are specified as dummy variables. In this case the Poisson model is equivalent to a log-linear model as both the dependent variable and the independent are qualitative. Log-linear models are more frequently used for modelling contingency tables (Agresti and Kateri, 2011). This type of model can be specified as totally saturated, in other words, with a perfect fit to the data as a parameter is specified for each observation. Willekens (1983) has shown how log-linear models are equivalent to the gravity models if they are conveniently scaled, usually by equalling the equilibrium factors of the first origin and destination to 1.

The fit of a Poisson model can be evaluated through different indicators as the Akaike information criteria (AIC), the log–likelihood or through the difference in the log-likelihood of the model estimated with respect to the totally saturated model, in other words, using a likelihood