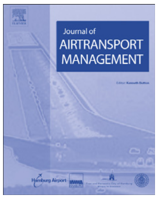
Journal of Air Transport Management 34 (2014) 123–130

Contents lists available at ScienceDirect

Journal of Air Transport Management

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



Analysis of bilateral air passenger flows: A non-parametric multivariate adaptive regression spline approach

Li-Yen Chang*

Department of Marketing and Logistics/Transportation, National Chia-Yi University, 580 Sin-Min Road, Chia-Yi 60054, Taiwan

Keywords:
Airlines
Passenger flows
Gravity model
Multivariate adaptive regression spline model (MARS)

ABSTRACT

Gravity models have long served as a framework for analyzing bilateral flows, trade and many other activities. To estimate a gravity model, (parametric) linear regression techniques have been commonly employed to develop the relationship between passenger flows and factors that can significantly influence these flows. This study explores the application of an alternative method, the non-parametric multivariate adaptive regression spline (MARS) technique, to identify the determinants for air passenger flows between pairs of countries. The data of 2006 and 2007 air passenger flows between pairs of countries in the APEC region were collected to develop the MARS models. Results indicate that distance between the countries, annual import value, national per capita income, unemployment rate and consumer price index of the origin country, as well as GDP, annual import value, and consumer price index of the destination country are significant determinants for bilateral air passenger flows.

1. Introduction

For many years airlines, economists, and air transport authorities, have had a serious interest in finding the determinants for the bilateral demand for air transport. For the airlines, the capability of accurately predicting the bilateral air transport demand, and of examining the potential factors that can influence this demand, is critical to their profits and operating strategies. For the air transport authorities, realistic forecasts of passenger demand are also essential for planning air transport infrastructure. Therefore, the problems of accurately predicting passenger and cargo demand, and of identifying the determinants for this demand, have attracted considerable research interest.

Gravity models, which can effectively describe the spatial interactions between two bodies, have been commonly adopted by past research as a framework to analyze bilateral activities, such as passenger flows, cargo flows, trade, investment and other activities. Econometric methodologies, particularly linear regression techniques, are then applied to estimate the gravity models and also to explore the determinants that can significantly affect the bilateral flows or activities.

Following past studies, this present study adopts the gravity model as the framework, but employs an alternative estimation technique, the non-parametric multivariate adaptive regression spline (MARS) model, to explore the determinants and the extent of their influence on the demand for cross-country air transport. MARS has been widely applied in many scientific fields and has demonstrated its effectiveness for dealing with prediction problems. In contrast to linear regression models, MARS is a non-parametric and nonlinear regression model, which not only requires no assumptions for the underlying relationships between the response variable and explanatory variables, but also provides greater flexibility to explore the nonlinear relationships between the response variable and explanatory variables by fitting the data into a series of spline functions. This technique is particularly helpful because some explanatory variables, such as distance and per capita income, might have nonlinear effects on the bilateral demand for air transport. By applying the MARS modeling technique, the potential nonlinear effects of explanatory variables can be effectively identified. Therefore, this study applies the MARS modeling technique to analyze the determinants and the extent of their influence on international air passenger flows between country pairs in the APEC (Asia–Pacific Economic Cooperation) region.

The rest of this paper is organized as follows. The next section provides a brief review of previous literature on modeling bilateral passenger flows, cargo flows and economic activities. Section 3 provides a description of the methodological approach and Section 4 discusses the data. Section 5 presents the modeling results of MARS analysis. Section 6 discusses the limitations of this study, and the final section concludes with a summary and directions for future research.

---

* Tel.: +886 5 2732933; fax: +886 5 2732932.
E-mail address: liyen@mail.ncyu.edu.tw.

0969-6997/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jairtraman.2013.09.003

2. Literature review

The gravity model originates from the concept in Newtonian physics. Newton's law of gravity in mechanics states that two bodies attract each other in proportion to the product of each body's mass divided by the square of the distance between their respective centers of gravity. This law has been adapted to analyze the spatial interactions and spatial flows of people, goods and information in the fields of business, economics, geography, regional science and transport. As described by Chang (2012), the linear (or natural logarithm) form of the gravity equation has been commonly applied to estimate the effects of economic characteristics of origin and destination countries on the bilateral activities. The linear form of the gravity model can be expressed as follows:

$$\log(T_{ij}) = \beta_0 + \beta_1 \log(P_i) + \beta_2 \log(A_j) - \beta_3 \log(d_{ij}) + \varepsilon \qquad (1)$$

where T_{ij} represents the number of trips produced by origin country i and attracted to destination country j, P_i is the production factors of country i (such as gross domestic product (GDP), population and other economic characteristics of origin country), A_j is the attracting factors of country j, d_{ij} is the distance between country i and country j, ε is an error term and β's are statistically estimated coefficients. In recent years, many studies have applied gravity models to analyze cross border activities such as passenger flows, cargo flows, tourist flows and trade flows. Some key studies are summarized in Table 1.

Most of these studies adopted the gravity model as their framework and applied linear regression techniques to identify the factors that could significantly influence cross border passenger flows (Chang, 2012; Endo, 2007; Grosche et al., 2007; Hazledine, 2009; Matsumoto, 2004, 2007; Wei and Hansen, 2006). The results consistently indicate that GDP and income of origin and destination countries have positive effects on bilateral passenger flows, whereas distance has negative effects. Other factors that can significantly influence bilateral passenger flows include fare, service frequency, open sky policy, language and location advantage.

In addition to the use of the gravity model in analyzing cross border passenger flows, recent applications of the model have also focused on cross border cargo flows (Hwang and Shiao, 2011; Matsumoto, 2004; Yamaguchi, 2008), tourist flows between pairs of cities or countries (Khadaroo and Seetanah, 2008; Marrocu, and Paci, 2012), and international trade (e.g., import and export) between pairs of countries (Anwar and Nguyen, 2011; Nitsch and Schumacher, 2004). For the cargo flows, per capita GDP, population, distance, freight rate, and open sky agreements have been significant determinants. Distance, income, accessibility, attractions and a destination's population density have been key determinants for the tourist flows.

Overall, economic characteristics (e.g., GDP and income), geographic characteristics (e.g., distance), and culture variables (e.g., population and language difference) have been identified as significant determinants for bilateral passenger or cargo flows. In order to effectively predict the passenger (or cargo) flows and identify the determinants for these flows, gravity models have commonly been adopted and a wide variety of modeling methods have been proposed to estimate these gravity models, including both parametric linear regression and non-parametric regression tree models. With recent advancements in computing power and technology, more and more non-parametric models have been developed and applied in various scientific fields. However, only limited studies have applied non-parametric models to predict the air passenger (or air cargo) flows. Therefore, this study applies the non-parametric MARS technique to analyze the relationships between air passenger flows and the variables of economic, geographic and cultural characteristics.

3. Methodology

MARS is a non-parametric and nonlinear regression methodology, first introduced by Friedman (1991). With its many advantages, such as no need to specify the function form as the parametric linear regression technique, and its greater flexibility to explore the nonlinear relationships between a response variable and explanatory variables, MARS has been commonly applied in many scientific fields for dealing with prediction problems. The basic principle of MARS is to allow different functions over different intervals. In other words, the nonlinear relationship between a response variable and an explanatory variable is approximated by the use of separate regression slopes in distinct intervals of the explanatory variable region. In addition, MARS searches for interactions between variables by checking all degrees of interactions. Since it allows for all functional forms and interactions, MARS is able to effectively track the complex data structures hidden in high-dimensional data. The general MARS function can be expressed using the following equations:

$$\hat{y} = b_0 + \sum_{m=1}^{M} b_m B_m(x) \qquad (2)$$

where $\hat{y}$ is the predicted response, x is the explanatory variable, b_0 and b_m are estimated coefficients to yield the best fit of data and M is the number of basis functions included into the model. $B_m(x)$ is the mth basis function, which can be either one single spline function or a product of two or more spline functions for different explanatory variables. The spline basis function, $B_m(x)$, can be specified as

$$B_m(x) = \prod_{k=1}^{k_m} \left[s_{km} \left(x_{v(k,m)} - t_{k,m} \right) \right]_+ \qquad (3)$$

where k_m is the number of knots, s_{km} can be either 1 or -1 to indicate the right/left regions of the associated step function, $v(k,m)$ is the label of the explanatory variable and $t_{k,m}$ is the knot location. Fig. 1 shows a simple example of how MARS uses a piecewise linear spline function to fit nonlinear data. Note that y and x in Fig. 1 are the response and explanatory variable, respectively. The MARS equation can be specified as

$$\hat{y} = 29.4 + 0.9 \times \max(0, 4 - x) - 2.2 \times \max(0, x - 4) + 1.9 \times \max(0, x - 8) \qquad (4)$$

This estimated MARS model contains three basis functions, which are a single spline function. The knots are located at $x = 4$ and 8, delimiting three intervals where different linear relationships between y and x are identified. The slope is -0.9 (i.e., $0.9 \times (-1)$) when x is less than 4, becomes -2.2 when x is between 4 and 8, and then -0.3 (i.e., $-2.2 + 1.9$) when x is greater than 8. As shown in Fig. 1, the non-linearity of a model can be effectively approximated through the use of separate linear regression slopes in distinct intervals of the explanatory variable space.

The variables to be used and the knot locations of the intervals for each explanatory variable are determined via an intensive search procedure. An optimal MARS model is developed through a two-stage forward/backward procedure. In the forward procedure, the MARS algorithm starts with the constant basis function in the model. Then a pair of basis functions, which produces the largest fit among all the possible pairs of basis functions, is successively added to the model each time. The pair of basis functions is the same except that a