Journal of Transport Geography 18 (2010) 183–190

Contents lists available at ScienceDirect

Journal of Transport Geography

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

Sample size needed for calibrating trip distribution and behavior of the gravity model

Celik H. Murat *

Department of City and Regional Planning, Izmir Institute of Technology, Urla Izmir 35437, Turkey

ARTICLE INFO

Keywords:
Travel survey
Sample size
Gravity model
Trip distribution
Istanbul

ABSTRACT

Conventional calibration algorithms of trip distribution models assume that the analyst has a whole base year trip matrix. To attain a whole trip matrix, the sample size for travel surveys needed to be as large as possible. However, this could be very expensive especially in large cities. Some studies in the past showed a small sized sample would be enough to estimate functional parameters of observed trip length frequency distribution. But the performance of a gravity model with small sized samples has never been addressed. This empirical study has shown that sample sizes as small as 1000 (even smaller for quick response studies) could be as dependable as large sample surveys using a line search calibration algorithm.

1. Introduction

It is possible to classify trip distribution models into two broad categories as aggregate and disaggregate. The disaggregate models try to explain individuals' behaviors in selecting the origins and destinations of their spatial movements while the aggregate models analyze total number of flows between analysis zones. Since disaggregate models work at individual level, proponents of such models claim that the data requirement for calibration of these types of models may be significantly lower (Ruiter and Ben-Akiva, 1978). The calibrated disaggregate model is later used to estimate the total inter-zonal movements by aggregation. Disaggregate models are behavioral, and individual choices are explained by individual's characteristics and choice set attributes. While these models may require fewer travel samples to calibrate, their eventual aggregation may require very extensive data at zonal level such as proportions of the representative individuals in each zones. Every distinct movement between origin–destination pairs establishes the choice set of the disaggregate models. As the number of travel zones increases, the number of the alternatives in a choice set increases which may lead to decreased estimation sensitivity. Due to stated bottlenecks, the aggregate models are still frequently preferred in professional practices and computer packages.

Contrary to the disaggregate models, the aggregate models require total numbers of trip interchanges between zone pairs and inevitably need larger sample sizes for model calibration. For a satisfactory aggregate modeling effort, the textbooks' recommended sample size for the travel survey is around 10% for small to medium sized cities, and it is around 4% when the city population exceeds 1,000,000 (Dickey et al., 1983; Cambridge Systematics, 1996; Ortuzar and Willumsen, 2001). However, increasing budgetary constraints for urban areas caused decision makers and transportation professionals to reconsider the expenditures on these expensive surveys since especially the marginal accuracy of the urban travel modeling with respect to increased sample size has not been very well documented. Since then, transportation professionals around the world have been trying to develop alternative techniques (such as synthesizing or updating trip matrices using link counts) with considerably lower costs.

The Travel Survey Manual (TSM) by the US Department of Transportation (Cambridge Systematics, 1996) states that it is possible to calibrate aggregate trip distribution models with a sample size as small as 1000 for each trip purpose based on a study conducted by Pearson et al. (1974). Using 20 different travel surveys conducted by the Texas Highway Department, Pearson et al. demonstrated that the trip length distribution (TLD) of urban travel statistically showed best fit to the Gamma distribution among other similarly shaped distributions: (i) Poisson; (ii) Chi–Square; (iii) Pearson Type III; and (iv) Wiebull. They also concluded that approximately 1000 trip observations for each trip purpose would be enough to estimate the best fitting parameters of the underlying Gamma distribution.

However, we cannot easily use the probability distribution function directly in our trip distribution models. Instead, during calibration, we generally search for the parameter(s) of an aggregate trip distribution model (i.e. a singly or doubly constrained gravity model) that replicates the observed trip length frequency distribution (OTLD). Thus, enough sample size for estimating a statistical distribution's parameters does not necessarily mean that

* Tel.: +90 2327507030; fax: +90 2327507012.
E-mail address: muratcelik@iyte.edu.tr

0966-6923/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtrangeo.2009.05.013

this sample size would automatically be sufficient for the gravity model performance to replicate the OTLD. However trivial it is, this issue stands as a valid research question and further investigation on this subject might be interesting. It is the intention of this paper to search for a minimum sample size empirically for an aggregate trip distribution model using the Household Travel Survey Data of Istanbul Metropolitan Area conducted in 2006.

The issues concerning calibration algorithms are discussed in the next section. Data used in this study and description of study area are explained in Section 3. Section 4 is devoted to the methodology and the empirical findings, and the last section concludes the study.

2. Trip distribution and calibration

Even though there have been many alternative formulations for the aggregate trip distribution models (namely, Growth Factor, Fratar, Intervening Opportunities, Gravity or Regression Models), the gravity model is the most preferred one over the years despite all of its drawbacks. A typical doubly constrained gravity model, which is also used in this research, is expressed as follows:

$$T_{ij} = A_i * B_j * O_i * D_j * f(c_{ij}) \tag{1}$$

where

$$A_i = \frac{1}{\sum_j B_j D_j f(c_{ij})} \tag{2}$$

$$B_j = \frac{1}{\sum_i A_i O_i f(c_{ij})} \tag{3}$$

O_i = total trip production by zone i,
D_j = total trip attraction to the zone j,
A_i = balancing factor assuring $\sum_j T_{ij} = O_i$,
B_j = balancing factor assuring $\sum_i T_{ij} = D_j$,
$f(c_{ij})$ = friction function between zone i and zone j.

There are well known functional forms of the friction in the literature. These are exponential function, $e^{-\alpha(c_{ij})}$; power function, $c_{ij}^{-\beta}$; and Tanner (or Gamma) function, $a * e^{-\alpha(c_{ij})} * c_{ij}^{-\beta}$ (Rose, 1975). If TLD shows a Gamma distribution (i.e. TLD increases for the first intervals, and decreases for later), then usage of a Tanner function is recommended in the model. If TLD has a negative exponential distribution (i.e. TLD is highest in the first interval(s) and continuously decreases later), then usage of an exponential or a power function is preferred. Certain issues may have important effects on the performance of the gravity model: (i) choice of spatial separation measure, (ii) choice of travel mode, (iii) choice of matrix type (i.e. production–attraction (PA) or origin–destination (OD) matrix, (iv) choice of functional form of the spatial separation, (v) choice of time of day, and (vi) choice of model type (i.e. person or vehicle). Even if these issues were decided conveniently, there are still important discussions about calibration algorithms, convergence criteria and acceptable sample size for trip distribution modeling.

To calibrate a gravity model, a modeler needs a good representation of the base year trip matrix implying a very large sample size. This need is not a theoretical requirement but rather a mathematical property of the calibration algorithms which were mostly been developed during 1970s. One of the early algorithms is the maximum likelihood estimator minimizing the difference between the observed and estimated trips (Wilson, 1970). However, the computational burden for this analytic procedure is extensive. Several numerical computational procedures were also suggested by different scholars (Hyman, 1969; Evans, 1971; Williams, 1976; Openshaw, 1976; Easa, 1993). Among them, Hyman's calibration algorithm was found to be reasonably efficient (Williams, 1976). Hyman's algorithm uses a Furness' Bi-Proportional Balancing Procedure and the mean travel time as convergence criteria to obtain the calibration (see Williams, 1976 for details). These algorithms were the pioneering studies on the subject, established professional conventions and they are still used in calibration procedures of many computer packages, either in the form of continuous deterrence function or BPR discrete friction factor (Easa, 1993).

A common assumption of cited algorithms was that a complete base year matrix is present (Dickey et al., 1983; Ortuzar and Willumsen, 2001) otherwise a partially observed matrix may produce unstable balancing factors leading to inconsistent rows and columns totals (Ortuzar and Willumsen, 2001, pp. 187–188). One alternative suggestion to work under incomplete information is that "the analyst need not worry too much if he wants to do a calibration when there is information missing about some inter-zonal transfers. He may omit completely from his calibration all cells for which information is missing, and the rest assured that had the missing data conformed to his (calibrated) model, the trips he synthesizes for the partial matrix would be the same as those he would have obtained by synthesizing the whole matrix". However, this premise comes with certain assumptions (Kirby, 1979, p. 423) Satisfying these assumptions, on the other hand, may also be problematic (see Kirby, 1979 for details).

Regardless of the sample size, a planner always has to work with partial or incomplete trip matrices as an inevitable practical situation. Then the task of a planner should be estimating or synthesizing the base year trip matrix with the smallest sample size possible. To avoid the above mentioned algorithms' bottlenecks, a line search algorithm with a "Furness' Bi-Proportional Balancing Procedure" is used in this research. This algorithm, rather than searching the parameter iteratively that may end up a local optimum, gives the opportunity to see the model performance for each specific parameter in a given interval according to various converging criteria.

A computer code using SAS–IML was developed to conduct the analysis. Exponential and power functional forms are tried in the analysis for two different convergence criteria: "mean travel time" and "root mean squared error (RMSE) between the observed and estimated TLDs". The algorithm used in the study can be summarized as follows:

(1) Estimate normalized OTLD and observed mean travel time,
(2) Determine the search interval and divide it by 0.01,
(3) Take the next parameter value in the line,
(4) Distribute zonal total productions and attractions using the parameter,
(5) Normalize estimated TLD and estimate RMSE with normalized OTLD of step 1,
(6) Print estimated RMSE and mean travel time,
(7) Terminate the iteration if all values of interval are exhausted, go to step 3 otherwise
(8) Choose the best fitting parameters in the interval.

The literature on convergence criteria (Pearson et al., 1974; Rose, 1975) and our empirical research, as will be explained shortly, demonstrated that the parameters replicating the mean travel time and the OTLD could be different due to smoothness between observed and estimated TLD. As the mean and variance of travel time increase, the probability that those two parameters differ would increase, which was one of the findings of present research as well.

3. Description of study area and data

Geographically, Istanbul is located on both sides of the Bosphorus, the natural strait connecting the Marmara and Black Seas, and