



# 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.

© 2009 Elsevier Ltd. All rights reserved.

## 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 med-

ium 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) Weibull. 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](mailto:muratcelik@iyte.edu.tr)

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:

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

where

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

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

$O_i$  = 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 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 convergence 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

separating the two continents: Asia and Europe. The city was the capital of Ottoman Empire for a very long time. Even though the capital moved to Ankara after the Turkish Independence War and the abolition of the Ottoman Monarchy in 1923, the city has remained as the focus of all leading economic activities as well as the primary destination of internal migration after the 1950s. The population of the city in 2005 was around 10.5 million while it was around 1 million in 1950. Istanbul has a divided mono-centric central business as shown in the circled area in Fig. 1. In this figure, the darker areas show the forests and environmental assets around the city, light colored areas are the valuable agricultural lands, and the shaded areas are occupied by urban uses. The main destination of morning commuting is toward the CBD, and in the evenings towards the fringes. This situation is the main cause of traffic congestion like in many other metropolitan areas of the world. There are two suspension bridges connecting the continents, and these bridges are very congested towards the west in the morning and the east in the evening.

According to the 2006 Household Travel Survey, there are approximately 21 million daily trips in the metropolitan area and 6% of them cross the Bosphorus between Asia and Europe. Approximately 300,000 of the continent crossing trips are carried by ferries while the remaining 1,000,000 trips use bridges. This continent crossing traffic is the major source of congestion not only for continent crossing but also for intra-continent traffic. The expressways serving the bridges are also the major arterials of the highway network of the city that carry intra-continent traffic. Thus, any congestion taking place over these expressways also hampers intra-continent circulations.

The leading transportation mode is walking (49%) and it can be said that Istanbul is a pedestrian oriented city. Private car usage is only around 14%, and public transit is around 36%. The backbone of the public transit system is rubber-tired public transit. Only 2.3% is rail transit due to a very short rail network, and 1% ferries due to a poorly integrated transit system. At the moment, there are 113 km of rail transit network in Istanbul. As well, there are some 68 km of rail transit lines under construction with different combinations of rail technologies, and some 116 km in the phase of sub-contracting. When all are completed the rail transit network is expected to total 297 km by the year 2012.

Data used in this study comes from the 2006 Household Travel Survey conducted by the Transportation Department of the Metropolitan Municipality of Istanbul. The sample size in this study included 90,000 households and sampling rate was approximately

3%. Surveyed households were determined as two-stage random cluster sampling. At the first stage, approximately 4000 primary sampling units (PSUs) were drawn from the latest household list of the municipality. At the second stage, 90 households around these PSUs were recorded as the secondary sampling units (SSUs) and at most 30 households out of 90 SSU were surveyed randomly in 451 travel analysis zones (TAZs) throughout the metro area. Fig. 2 presents the TAZ of Istanbul. The surveying technique was face to face. Sample replacement was strictly forbidden to avoid non-sampling errors. Every household member older than six was surveyed for their previous week-day travels as well as the household and personal information. Since sample replacement was not allowed, on average three visits were paid to each household to catch every eligible member at home or to convince the household to participate in the survey. At the end of the field work, a remarkable overall response rate of 80% was achieved.

With this response rate, 264,000 people in 72,000 households were surveyed. A total of 356,000 trips were recorded between 451 OD pairs making 203,401 distinct movements. 127,000 of these trips were home-based-work (HBW); 94,000 trips were home-based-school (HBS); 115,000 were home-based-other; and 20,000 were non-home-based (NHB) trips.

#### 4. Methodology and empirical findings

This empirical study was structured in three steps. In the first step, using the whole data set (i.e. 356,000 trips) a best fitting model was estimated for each trip purpose (HBW, HBS, HBO, and NHB). In the second step, 15 random subsets from the original data were drawn with the sample sizes ranging from 200 to 3000 increasing by 200, and the model parameters of each subsample were estimated and compared with the original models. At the third step, for the suggested sample size (i.e. 1000), 30 separate random samples were drawn from the original data, and the model parameters, parameters' means, standard deviations, and standard errors of the means were estimated to confirm the sufficiency of the suggested sample size.

Instead of a vehicle model, a person trip model was preferred since it was not possible to know future mode split. For the spatial friction, it was not possible to use survey stated times since it was not possible to catch trips between each OD pair and for each travel mode in the survey. For this reason, a system produced car travel time matrix was used in the models. To obtain the travel time matrix, an initial set of models with free flow speed travel time were

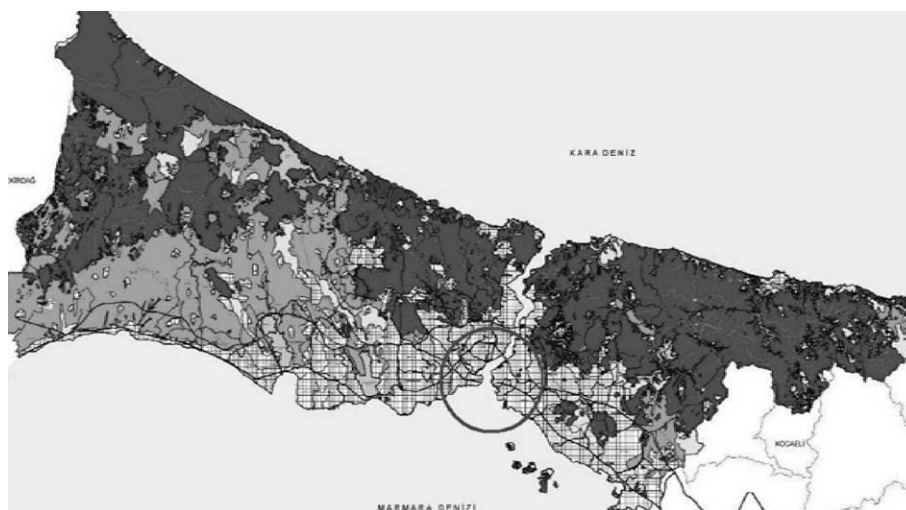


Fig. 1. Land use in the Istanbul Metropolitan Area.



Fig. 2. Traffic analysis zones and highway network of Istanbul Metropolitan Area.

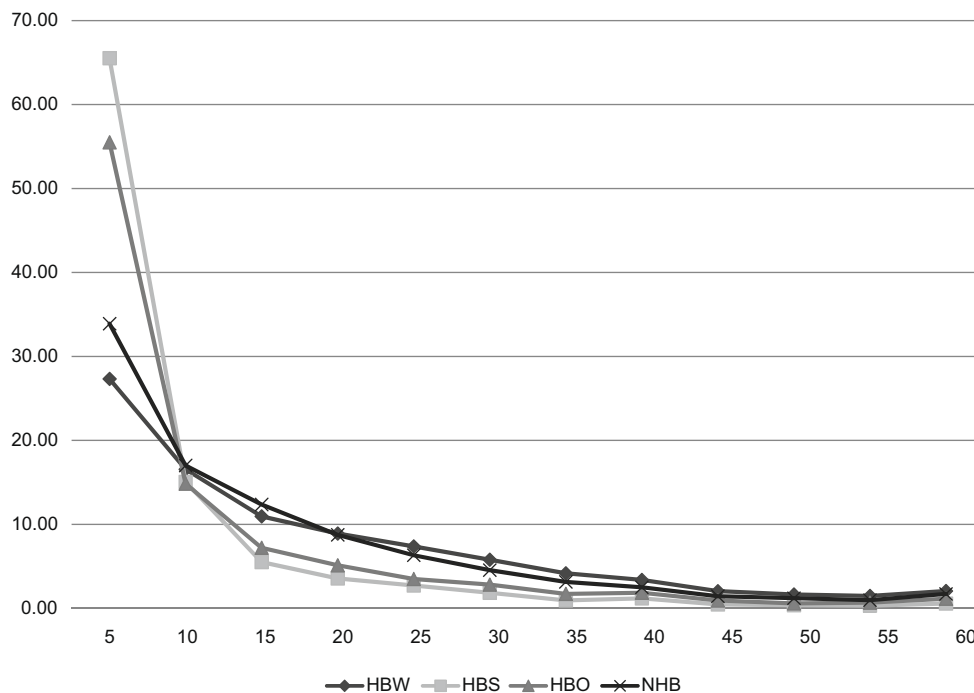


Fig. 3. Observed trip length frequency distributions by trip purposes.

estimated, and with observed percentages of mode choices, these initial travel matrices were assigned to the highway network. These steps were repeated three times until a reasonable convergence for the congested travel times were obtained, and this resultant travel time matrix is used in the study.

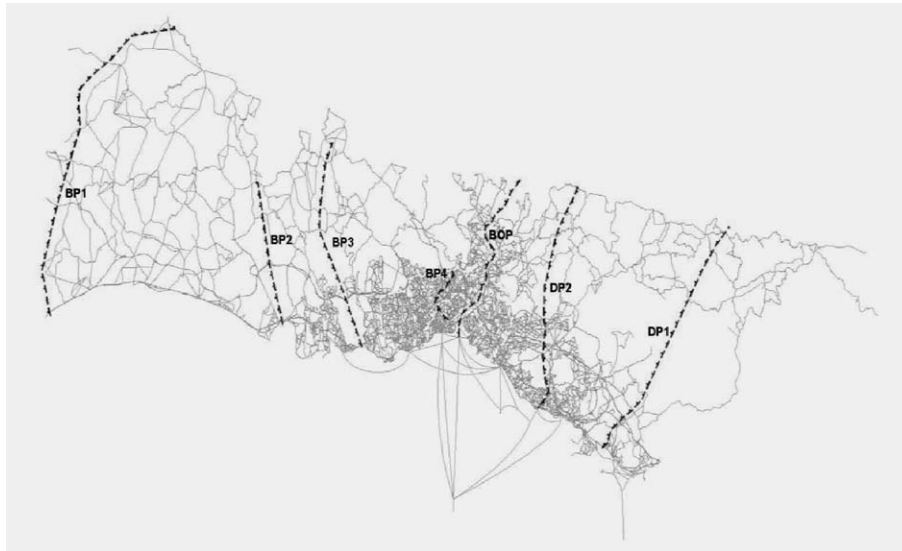
For original calibration, the previously described line search algorithm was employed. The search intervals were  $-4$  to  $0$  for the inverse power and  $-1$  to  $0$  for the exponential models. The OTLD classified into 31 and 5 minute bins. OTLDs for each trip purpose are presented in Fig. 3. Each trip purpose is modeled with both exponential and inverse power functional forms according to two convergence criteria: RMSE (for TLDs) and mean travel times. The best fitting parameters were presented in Table 1. In this way, for each trip purpose, four different models were estimated.

To determine the best fitting set of the models to the ground counts at the screen lines was the next task of the first step. Fig. 4 shows the screen lines of Istanbul Metro area. Using each model, a base year's trip matrices were synthesized and summed at the screen lines, and these sums were compared with the ground counts. The results are presented in Table 2. It was concluded that the best fitting set of models was an inverse power model using the TLD RMSE as the convergence criterion (Table 2).

During this study, even though it was not the initial intention, an interesting behavior of the gravity model with exponential or power function was witnessed. Previously, it had been stated that when the OTLD shows a negative exponential distribution, usage of exponential or power function is recommended expecting that these functions would produce a TLD showing a negative exponential distribution. This statement was confirmed for the most of the

**Table 1**  
Estimated models' parameters.

	Observed mean travel time	Observed SD	Power model parameters		Exponential model parameters	
			Mean travel time	TLD RMSE	Mean travel time	TLD RMSE
HBW	21.13	23.57	-2.05	-1.77	-0.105	-0.19
HBS	8.24	14.3	-2.98	-2.90	-0.28	-0.64
HBO	11.89	18.43	-2.39	-2.41	-0.18	-0.48
NHB	17.46	21.52	-1.74	-1.63	-0.09	-0.29



**Fig. 4.** Screen lines of Istanbul Metropolitan Area.

**Table 2**  
Screen line ground counts and model produced crossings at the screen lines in persons.

	BOP	BP4	BP3	BP2	DP1	DP2	Total RMSE
Ground counts	<b>1,244,419</b>	<b>2,030,433</b>	<b>808,797</b>	<b>255,451</b>	<b>394,284</b>	<b>1,156,505</b>	
Power mean travel time	1,122,783	1,773,959	830,851	309,716	361,865	1,014,254	629,099
Power RMSE	<b>1,324,858</b>	<b>2,029,227</b>	<b>943,140</b>	<b>352,191</b>	<b>392,584</b>	<b>1,135,339</b>	<b>335,594</b>
Expo mean travel time	668,351	1,703,056	652,828	177,534	363,593	1,137,980	1,186,547
Expo RMSE	674,800	787,870	351,962	88,874	189,596	564,326	3,232,461

**Table 3**  
Bifurcation points for trip purposes and models.

Trip purposes	Exponential function	Power function
HBW	-0.27	-1.05
HBS	-0.25	-1.05
HBO	-0.24	-1.00
NHB	-0.19	-0.85

parameter values. However, it was discovered that there is a bifurcation point for these functions, after which the model produced TLDs demonstrate a Gamma shaped distribution. Table 3 presents the values of these bifurcation points for trip purposes. For example, from negative infinity to -1.05, the power model produced TLD shows a negative exponential type of distribution. After that point towards 0, the TLD becomes a Gamma shaped distribution. Behavior of the HBW Power Model for certain parameters are presented as an example in Fig. 5.

In the second step of this study, in order to determine a minimum sample size that could produce approximately the same parameters estimated with full data, 15 subsamples (starting from 200 up to 3000 with an equal interval of 200) were randomly

drawn from the survey data, and the gravity model parameter was re-calibrated with each subsample. The estimated parameters and their respective percentage errors for each trip purpose are given in Tables 4–7.

Looking at the tables, it is possible to say that independent of sample size, each subsample was able to capture the respective parametric magnitude. Table 8 shows the average parameter estimate of 15 subsamples for HWB trips was -1.76 with a standard deviation (SD) of 0.040 and its original parameter estimate was -1.77 and -2.88 with a SD of 0.067 for HBS trips and its original parameter was -2.90 and -2.42 with a SD of 0.038 for HBO trips and the original estimate was 2.41 and -1.62 with a SD of 0.032 for NHB trips while the original parameter was 1.63. In each case, average parameter estimates were very close to the originally estimated parameters with very small variances. Percent errors of TLDs reported in the last columns of the tables confirmed this fact as well. The mean percent error of 15 subsamples for HBW was 0.77 with a SD of 1.40; 1.28 with a SD 0.91 for HBS; 1.08 with a SD of 0.80 for HBO; and 0.97 with a SD of 0.97. Tables 4–7 reveal that only at four cases out of 60 had a percent error of more than 2, and the rest were around or below 2%. Error fluctuations with respect to sample sizes and trip purposes are presented in Fig. 6.

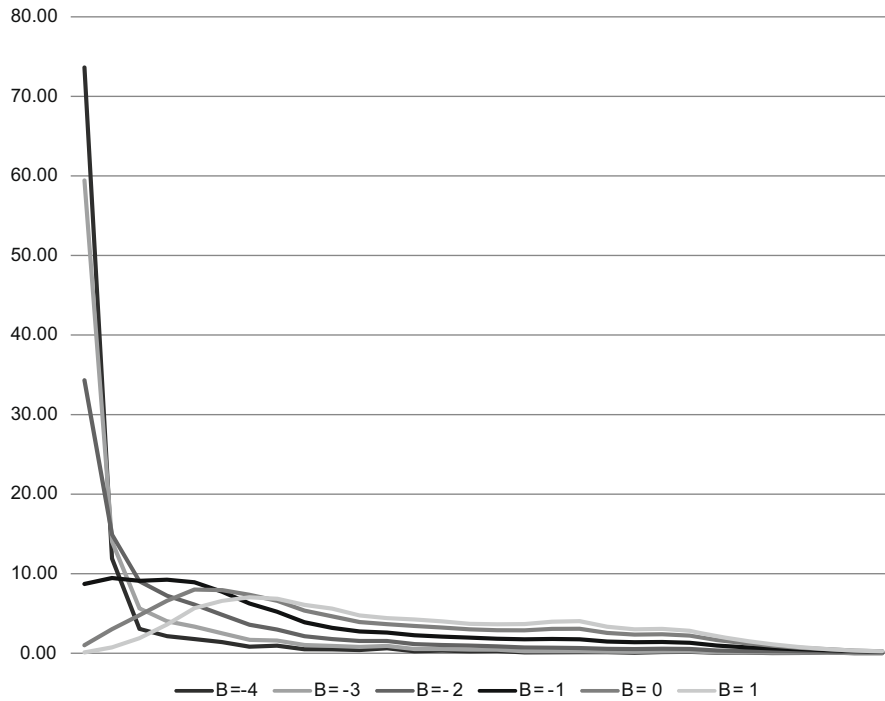


Fig. 5. HBW Power Model TLD for certain parameters.

**Table 4**  
HBW model parameters and errors for different sample sizes.

Sample sizes	Sample mean travel time	RMSE parameter	% Error
200	21.71	-1.75	<b>0.62</b>
400	23.11	-1.63	<b>4.30</b>
600	19.65	-1.73	<b>1.24</b>
800	21.43	-1.81	<b>1.25</b>
1000	20.97	-1.75	<b>0.62</b>
1200	20.70	-1.77	<b>0.00</b>
1400	20.41	-1.78	<b>0.31</b>
1600	21.57	-1.77	<b>0.00</b>
1800	20.89	-1.79	<b>0.62</b>
2000	21.00	-1.79	<b>0.62</b>
2200	21.00	-1.76	<b>0.31</b>
2400	21.06	-1.79	<b>0.62</b>
2600	20.95	-1.76	<b>0.31</b>
2800	21.73	-1.76	<b>0.31</b>
3000	21.00	-1.78	<b>0.31</b>
<b>127,598</b>	<b>21.13</b>	<b>-1.77</b>	<b>3.12</b>

**Table 6**  
HBO model parameters and errors for different sample sizes.

Sample sizes	Sample mean travel time	RMSE parameter	% Error
200	12.18	-2.47	<b>1.96</b>
400	10.86	-2.41	<b>0.00</b>
600	11.78	-2.47	<b>1.96</b>
800	10.71	-2.47	<b>1.96</b>
1000	11.93	-2.43	<b>0.66</b>
1200	11.26	-2.37	<b>1.34</b>
1400	12.02	-2.41	<b>0.00</b>
1600	10.79	-2.49	<b>2.60</b>
1800	12.48	-2.37	<b>1.34</b>
2000	12.95	-2.39	<b>0.67</b>
2200	11.84	-2.40	<b>0.33</b>
2400	11.78	-2.41	<b>0.00</b>
2600	11.89	-2.38	<b>1.00</b>
2800	11.39	-2.46	<b>1.64</b>
3000	11.98	-2.43	<b>0.66</b>
<b>114,921</b>	<b>11.89</b>	<b>-2.41</b>	<b>1.19</b>

**Table 5**  
HBS model parameters and errors for different sample sizes.

Sample sizes	Sample mean travel time	RMSE parameter	% Error
200	9.85	-2.82	<b>1.82</b>
400	6.67	-3.01	<b>2.32</b>
600	9.28	-2.74	<b>3.77</b>
800	7.69	-2.92	<b>0.44</b>
1000	7.89	-2.99	<b>1.91</b>
1200	7.75	-2.89	<b>0.22</b>
1400	8.64	-2.83	<b>1.59</b>
1600	8.74	-2.83	<b>1.59</b>
1800	8.49	-2.92	<b>0.44</b>
2000	8.47	-2.87	<b>0.67</b>
2200	7.90	-2.86	<b>0.90</b>
2400	8.27	-2.85	<b>1.13</b>
2600	8.41	-2.92	<b>0.44</b>
2800	8.23	-2.84	<b>1.36</b>
3000	7.86	-2.93	<b>0.65</b>
<b>94,052</b>	<b>8.24</b>	<b>-2.90</b>	<b>1.02</b>

**Table 7**  
NHB model parameters and errors for different sample sizes.

Sample sizes	Sample mean travel time	RMSE parameter	% Error
200	16.75	-1.64	<b>0.40</b>
400	15.61	-1.65	<b>0.80</b>
600	17.75	-1.61	<b>0.79</b>
800	17.79	-1.58	<b>1.97</b>
1000	16.94	-1.65	<b>0.80</b>
1200	17.05	-1.63	<b>0.00</b>
1400	17.64	-1.58	<b>1.97</b>
1600	18.66	-1.53	<b>3.92</b>
1800	17.90	-1.61	<b>0.79</b>
2000	17.86	-1.61	<b>0.79</b>
2200	17.31	-1.62	<b>0.40</b>
2400	17.94	-1.63	<b>0.00</b>
2600	17.58	-1.66	<b>1.19</b>
2800	17.46	-1.62	<b>0.40</b>
3000	17.33	-1.62	<b>0.40</b>
<b>20,109</b>	<b>17.46</b>	<b>-1.63</b>	<b>2.62</b>

**Table 8**

Test and descriptive statistics for subsample group with sample size less than 3000.

N = 15	B	Mean B	SD	t Statistics	Mean error	Error SD
HBW	1.77	1.76	0.040	-0.966	0.77	1.31
HBS	2.90	2.88	0.067	-1.154	1.28	0.91
HBO	2.41	2.42	0.038	1.009	1.08	0.80
NHB	1.63	1.62	0.032	-1.694	0.97	0.97

To gain a more analytical insight for our findings, we classified our subsamples into three arbitrary groups, estimated mean and standard deviation of parameter estimates, and conducted a student *t* test. The first group included all subsamples and their test statistics were presented in Table 8. In this table, using all the subsamples' means and standard deviations, *t* statistics of estimated parameters were estimated. All *t* tests were lower than the critical table value of 2.145 (at 0.025 error range with a d.f. 14) indicating that the average parameter estimate of these 15 subsamples is not significantly different than the originally estimated model parameters for all each purposes.

The same analysis was conducted for the groups with sample sizes less than 2000 and 1000 to find the minimum threshold of the sample size. As it can be seen in Table 9, *t* statistics were lower than the critical table value of 2.262 (at 0.025 error range with a d.f. 9) again indicating that none of the mean parameter values of 10 subsamples is significantly different from the originally estimated parameter values for all trip purposes.

For the subsamples including a sample size of less than 1000 (see Table 10), the analysis has revealed that sample sizes around 1000 would approximately produce a parameter that is not significantly different from the originally estimated parameter for the HBW, HBS, and NHB trips. Their *t* statistics are lower than the critical table value of 2.776 (at 0.025 error range with a d.f. 4). However, to be able to confident about the parameter estimate of HBO, the sample size should be increased a little bit. But this increase should not be too much since it is still not significantly different from the 0.010 critical table value (which is 3.747) at this

**Table 9**

Test and descriptive statistics for subsample group with sample size less than 2000.

N = 10	B	Mean B	SD	t Statistics	Mean error	Error SD
HBW	1.77	1.75	0.047	-1.346	0.96	1.40
HBS	2.90	2.88	0.078	-0.813	1.48	1.03
HBO	2.41	2.42	0.042	0.748	1.25	0.84
NHB	1.63	1.61	0.035	-1.807	1.22	1.07

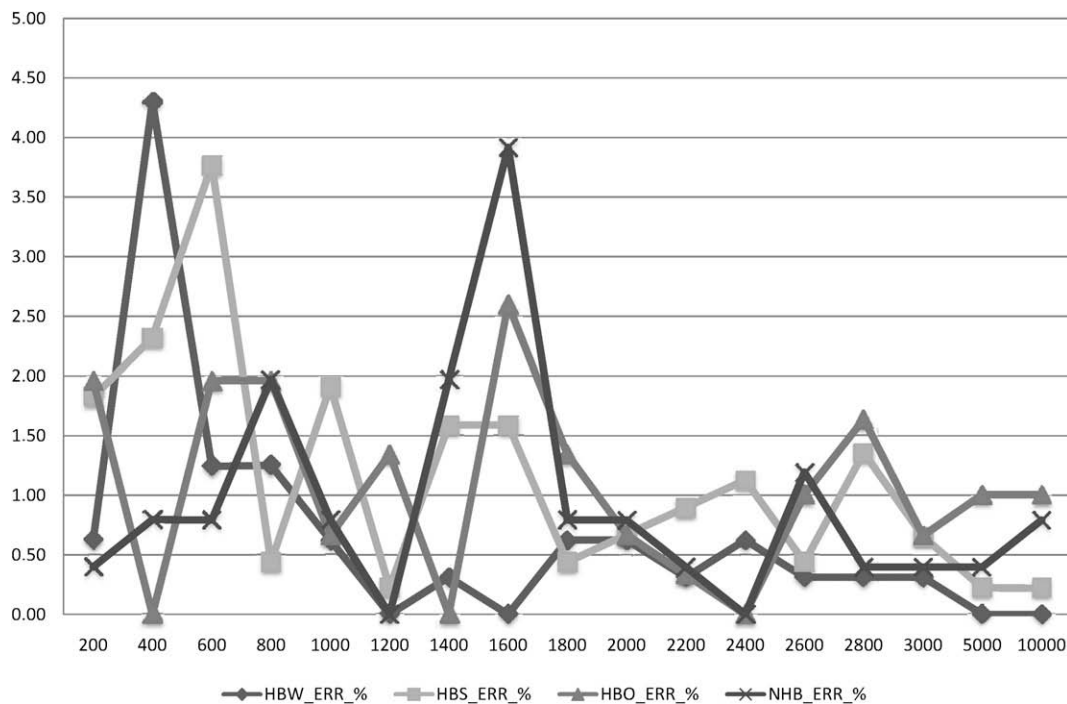
**Table 10**

Test and descriptive statistics for subsample group with sample size less than 1000.

N = 5	B	Mean B	SD	t Statistics	Mean error	Error SD
HBW	1.77	1.73	0.059	-1.529	1.60	1.38
HBS	2.90	2.89	0.105	-0.213	2.05	1.14
HBO	2.41	2.45	0.025	3.535	1.31	0.57
NHB	1.63	1.62	0.027	-0.822	0.79	0.56

threshold, and we have shown that we could be confident with a sample size of less than 2000 for HBO trips.

The experiment so far has shown that a sample size of about 1000 would be enough to estimate the parameters that are not significantly different from the original parameter. In the third step of the study, 30 separate subsamples with a sample size of 1000 were drawn from the whole data set, and parameters were estimated for each subsample and each trip purpose. The results are included in Table 11. The average parametric value for HBW is 1.78 with a standard error of 0.008. The confidence level for the mean is 98% meaning that with a sample size of 1000, 98 times out of every 100, the parameter estimate would be between 1772 and 1788. Considering our original parameter value is 1.77 for HBW, this result confirmed that the sample size of 1000 would produce a parametric value that is not significantly different than larger size samples. Actually this result is also confirmed for all the trip purposes. The mean is 2.89 with a standard error of 0.015 and 97% confidence level for HBS; 2.41 with a standard error of 0.007 and 99% confidence level for HBO and 1.62 with a standard error of



**Fig. 6.** Model's percent error change for different sample sizes and trip purposes.

**Table 11**  
Descriptive statistics of 30 subsamples of sample size 1000.

	HBW	HBS	HBO	NHB
Original	1.77	2.90	2.41	1.63
Mean	1.78	2.89	2.41	1.62
Standard deviation	0.043	0.082	0.040	0.035
Standard error of mean	0.008	0.015	0.007	0.006
Confidence of mean	0.02	0.03	0.01	0.01
N	30	30	30	30

0.006 and 99% confidence level for NHB. In all cases, the original parametric values remained in the interval bounded with the standard errors of the means.

## 5. Conclusion

Our analysis has confirmed the premise that a sample size around 1000 for each trip purpose would produce approximately the same parameter estimate as the gravity model with larger sample sizes, beyond estimating the TLD statistical distribution parameter as stated by *Travel Survey Manual* (Cambridge Systematics, 1996). The analysis also reveals that as the sample size increases, the parameter and error variances decrease but this decrease makes only a marginal contribution to the model accuracy implying an unnecessary, wasted investment in acquiring a larger sample.

The findings of this study can be generalized as long as three basic assumptions are satisfied. (i) The urban area has one distinct TLD: if the TLD within an urban area differs from region to region, then each region's parameters should be estimated separately as long as these regions can be identified and separated into sub-matrices. In our case, only the Bosphorus crossing showed a slightly different TLD. However, since the share of the Bosphorus crossings were relatively low (i.e. 6%), the TLD was not distorted overall. (ii) The TLD shows a negative exponential distribution: the OTLD in the present study shows a negative exponential distribution. If the spatial separation function is a Tanner function, then two different parameters need to be determined simultaneously. In this case, the line search algorithm can no longer be used. One alternative could be a grid search algorithm, and this new algorithm may need larger sample sizes. (iii) The sample selection is perfectly random, and no non-response and sample replacement bias is present: it is a well known fact that travel sensitivities of different socio-economic groups can be very different. If the sample is not chosen randomly or a high rate of non-response is present, then the bias may distort TLD significantly leading to an incorrect parameter estimate. An in-depth discussion of the sampling issues goes well beyond the present paper, and readers are referred to Kish (1965) and Stopher and Jones (2003).

The findings of this paper may have some professionally important implications. First of all, it reinstates the position of traditional trip distribution models (with respect to alternative OD matrix estimation techniques) for estimating a base year OD matrix with

considerably smaller budgets in contrast to orthodox belief that travel surveys may require large budgets. Secondly, with very small random samples, the prescribed quick response methodology can be used in conjunction with the OD matrix estimation techniques from link counts for the validity of the results.

However, it should be also stated that this result does not necessarily mean that small sample sizes would be sufficient for all stages of travel demand modeling. Depending on the local and planning requirements, larger sample sizes may be needed due to lack of spatially disaggregate travel and socio-economic statistics. For example, the sample size for Istanbul was intentionally kept large to establish a zonal database and a transportation information system beyond modeling purposes.

It would worth trying the same study for smaller sample sizes than 1000, but that would require an extensive amount of empirical work, and be a task for future research. Also, it will certainly be a very complementary piece of work to conduct a similar study for a TLD showing a Gamma distribution. That leads to a future research direction as well.

## Acknowledgments

The author wishes to thank to Department of Transportation of Istanbul Greater Municipality for making the survey available; to Prof. Dr. Gokmen Ergun of Bosphorus Uni. and two anonymous referees for their helpful comments and suggestions.

## References

- Cambridge Systematics, 1996. *Travel Survey Manual*. US Department of Transportation and US Environmental Protection Agency.
- Dickey, J. et al., 1983. *Metropolitan Transportation Planning*. Taylor & Francis, Bristol, PA.
- Easa, S.M., 1993. Urban trip distribution in practice. *Journal of Transportation Engineering* 119 (6), 793–815.
- Evans, A.W., 1971. The calibration of trip distribution models with exponential or similar cost functions. *Transportation Research* 5, 15–38.
- Hyman, G.M., 1969. The calibration of trip distribution models. *Environment and Planning* 1, 105–112.
- Kirby, H.R., 1979. Partial matrix techniques. *Traffic Engineering and Control* 20, 422–428.
- Kish, L., 1965. *Survey Sampling*. John Wiley & Sons, New York, NY.
- Openshaw, S., 1976. An empirical study of some spatial interaction models. *Environment and Planning A* 8, 23–41.
- Ortuzar, J.D., Willumsen, L.G., 2001. *Modeling Transport*, third ed. John Wiley & Sons, New York, NY.
- Pearson, D.F., Stover, V.G., Benson, J.D., 1974. A Procedure for Estimation of Trip Length Frequency Distributions. Texas Transport Institute, Report No.: TTI-2-10-74-17-1.
- Rose, J.G., 1975. The calibration of trip distribution models – a new philosophy. *Urban Studies* 12 (8/9), 335–338.
- Ruiter, E.R., Ben-Akiva, M.E., 1978. Disaggregate travel demand models for the San Francisco area: system structure, component models and application procedures. *Transportation Research Record* 673, 121–128.
- Stopher, P., Jones, P. (Eds.), 2003. *Transport Survey Quality and Innovation*. Pergamon, New York.
- Williams, I., 1976. A comparison of some calibration techniques for doubly constrained models with an exponential cost function. *Transportation Research* 10, 91–104.
- Wilson, A.G., 1970. *Entropy in Urban and Regional Modeling*. Pion, London.