Cross classification trip production model for the city of Alexandria

Mounir Mahmoud Moghazy Abdel-Aal
Transportation Eng. Dept., Faculty of Eng., Alexandria University, Alexandria, Egypt

The errors in estimating trip production are seriously carried through and may invalidate the entire transportation planning process. This paper uses the data-efficient Multiple Classification Analysis (MCA) for modeling trip production. Due to non-availability of recent travel survey for the city of Alexandria, an application of the proposed model is carried out using a recently collected data sample. A weighing procedure is applied and proved to be effective in offsetting the bias of small colleted sample. The two-dimensional MCA model shows the model sensitivity to reflect the effect of changes in socioeconomic household attributes on trip rates in the future. Different household attributes are tested and their suitability is discussed as well. The model accuracy is affected by the small sample size in terms of a wide confidence interval. A discussion is given about the sufficient sample size. An innovative validation was implemented backward to information collected 20 years ago. In addition, a simple method, to predict population cross distribution relative to the model variables, based on the available census data is introduced. The proposed model suggests that there is an indication that the average trip rate has increased by about 10% during the last two decades. More significantly, the motorized trips have largely increased while walk trips has decreased.

يؤثر الخطأ فى التنبؤ بالرحلات المتولدة تأثيراً كبيراً على عملية تخطيط النقل مما قد يؤدى إلى عدم صلاحيتها. وهذا البحث يستخدم نموذج الشرائح المتعددة وهو أحد نماذج الشرائح المتقاطعة الذى يتطلب معلومات أقل لحساب الرحلات المتولدة بأقل خطأ. ونظراً لعدم وجود إحصاءات حديثة لتنقل المواطنين فى مدينة الاسكندرية فقد تم تطبيق النموذج المقترح بإستخدام عينة من المعلومات الحديثة، ولتلافى الإنحراف المحتمل فى العينات الصغيرة تم إعطاء أوزان مختلفة لمكونات العينة. وقد أظهر التطبيق حساسية النموذج للتغير فى خصائص السكان الإجتماعية والإقتصادية وصلاحيته للتنبؤ بالرحلات المتولدة فى المستقبل. والبحث يختبر صلاحية عدد من المتغيرات المختلفة للإستخدام فى النموذج وكذلك يقترح حجم العينة المناسب للوصول إلى دقة النموذج المثالية. وقد تمت معايرة النموذج بإستخدام معلومات إحصائية متوفرة من دراسة سابقة. ويقدم البحث بالإضافة لهذا طريقة مبسطة للتنبؤ بتوزيع السكان المستقبلى على الشرائح الإجتماعية والإقتصادية المختلفة بإستخدام المعلومات المتاحة من الإحصاء القومى. وتشير نتائج هذا البحث إلى إزدياد معدل رحلات الفرد بحوالى ١٠% خلال العقدين السابقين كما يشير إلى إرتفاع كبير نسبياً فى معدل الرحلات الراكبة فى مقابل إنخفاض معدل رحلات السير.

Keywords: Transportation planning, Transportation modeling, Trip production, Cross classification model, Trip rates

1. Introduction

Transportation models, in general, are developed to assist in the formulation and evaluation of transportation plans and projects. Trip generation is the decision to travel. It is very serious to underestimate (or overestimate) trip generation since it is the first stage of the four-step modeling process. That is, errors of this stage are carried through the entire process and may invalidate work on subsequent stages of trip distribution, mode choice and trip assignment [1].

Growth factor methods are invalid. The assumption that the average trip rates will remain constant (using only population) would underestimate the trip production given the fact that the trip rates are naturally increasing with the recent changes of life styles. This is perfectly true in the developing countries with the rapid increasing rates of car ownership compared to the industrialized countries with higher rates at their peak**.** The use of the growth of variables such as income and car ownership to reflect the growth of trip rates is very crude as the effect of such variables on trip production are, in fact, so complicated to be modeled in some form of a multiplicative function. These methods are, therefore, only used in practice to predict external trips to/from an area.

The state-of-practice of trip production had two major stages started with linear regression up to the late 1960's and ended in the 1980's with various cross-classification type methods, namely: household level and person category cross-classification models [1]. In spite of the simplicity of forecasting using linear regression models, the problem of non-linearity, especially in the multiple regression analysis, reduces the effectiveness of such models. In addition, regression models tend to be good describers but not good predictors because assumptions about travel behavior may not hold, basically, because of the additive mature (multicollinearity) of the regression models.

Household-based cross-classification models find the number of trip productions as a function of household attributes (classes). The most important attributes affecting trip productions are income, car ownership, family structure (age of householder and number of workers) and family size [1]. The assumption of trip rates stability over time for a given household stratification became more acceptable. The correct application of such a model is to estimate the number of households in each class within a zone and multiply the trip rates by those numbers of households. In general, this modeling procedure leads to greater disaggregation than any other trip generation model and has the potential to provide more policy responsiveness than alternative models [2]. One critical element of such models is the data requirements, especially for models with large number of classes. Another problem is the need to predict the number of household of each class in the future.

For the city of Alexandria, Egypt, a little has been done with regard to trip production modeling. Since estimating the average trip rate for the entire city (of 1.2 trips/ inhabitant/day) as a part of TranSystem study in 1982 [3], no comprehensive travel survey, if any, has been carried out. Nevertheless, the study did not produce a behavioral model to help predicting the future trip production of the city.

There is a real need to estimate a trip production model based on recently collected data. Since large-scale surveys are prohibitively expensive, a small sampling effort is the best to be done. However, the problem of biasness of small samples can arise. This research aims to estimating a trip production model of the cross-classification type based on a small sample weighted to represent the population totals of key attributes leaving out these included in household stratification.

2. Model formulation

It has been empirically proven that household size, car ownership, and number of workers per household are good estimators of household trip production for all purposes except for home-based shopping trips [2]. Although correlated with car ownership, income has less effect on major trip purposes; e.g., work trips. The age of householder, which serves as an indication of the average age of all members of the household, seems as a good candidate explanatory variable of a trip production model. Household size, car ownership, age of householder, number of workers per household and income are the five candidate variables to be considered in this paper.

2.1. Cross-classification trip rate

2.1.1. Classical cross-classification analysis

The Cross-Classification Analysis (CCA) method is based on empirically estimating the number of trip productions as a function of household attributes. The basic assumption is that trip rates are stable over time for any given household stratification. To say nothing about its need for a large amount of data, one problem of this type of trip production models is the need to forecast the number of households of each household class in the future. Mathematically, the CCA model is given by [1]:

$$t(h) = T(h)/H(h) \qquad \forall h \,, \tag{1}$$

where:

$t(h)$ is the average number of trips made by members of households of type h,

$T(h)$ is the observed trips made by household of type h, and

$H(h)$ is the observed number of households of type h.

For a two-way classification model of M household sizes and N car ownership classes, the CCA model can be rewritten in an array format as follows:

$$t(m,n) = T(m,n) / H(m,n) \quad \forall m \in M, \forall n \in N \,, \qquad (2)$$

where (m,n) represent a household of size (m) and car ownership class (n)

To estimate a weighted model, each record (household) would have a weight different from the unity to reflect its likelihood to exist in the population. That is, number of households of each class would be, simply, the sum of the corresponding weights. Similarly, the number of trips made by a certain household class would be the weighted sum of trips. Record (household) weights are considered in calculating T (.) and H (.) as follows:

$$T(m,n) = \sum_{i \in \{n,m\}} W_i T_i \qquad \forall m \in M, \forall n \in N \,, \qquad (3)$$

$$H(m,n) = \sum_{i \in \{n,m\}} W_i \qquad \forall m \in M, \forall n \in N \,, \qquad (4)$$

where:

T_i is the trips made by household record i,

W_i is the weight of household i, and

$\{m,n\}$ is the subset of the sample belongs to the household class of size (m) and car ownership (n) class .

2.1.2. Multiple classification analysis

Multiple Classification Analysis MCA is an alternative method that overcomes the data requirement problem cited above. The mechanism of reducing the amount of data requirement is to base trip rates on the grand mean estimated over the entire sample of household and group means estimated for each row and column of the cross-classification matrix. In addition to increasing the efficiency of estimating trip rates, MCA method accommodates for estimating trip rates for household types not present in the sample. The array formulation the MCA is given by:

$$t(m,n) = \frac{\sum_{n \in N} T(m,n)}{\sum_{n \in N} H(m,n)} + \frac{\sum_{m \in M} T(m,n)}{\sum_{m \in M} H(m,n)} - \frac{\sum_{m \in M} \sum_{n \in N} T(m,n)}{\sum_{m \in M} \sum_{n \in N} H(m,n)} \quad \forall m \in M, \forall n \in N \,. \qquad (5)$$

Where the first term is the group mean for household size $m \in M$, the second term is the group mean of car ownership class $n \in N$ and the third term is the grand mean of the entire sample.

2.2. Trip production

The correct application of the model is to estimate the number of households in each class within a zone and to multiply the trip rates by those numbers of households. To estimate the aggregate Trip Production (TP) of an area, the following equation can be used:

$$TP = \sum_{m \in M} \sum_{n \in N} F(m,n) * t(m,n) \quad \forall m \in M, \forall n \in N \,, \quad (6)$$

where: $F(m,n)$ is the number of household of size (m) and car ownership class (n) in the area of interest

2.3. Average trip rate

Average trip rate for an area refers to the average trip rate per person regardless of the household class that it belongs to. The average trip rate per person is by:

$$TR = TP / POP \,, \qquad (7)$$

where:

TR is the average trip rate per person, and

POP is the total population of the area.

Substituting form eq. (7) into (6), one gets:

$$TR = \sum_{m \in M} \sum_{n \in N} a(m,n) * t(m,n) \ \forall m \in M, \forall n \in N \ , \quad (8)$$

where: $a(m,n)$ is the Percentage of household of size (m) and car ownership class (n) in the area of interest; $a(m,n) = F(m,n) / POP$

3. Sample description

A travel survey of 500 households of the city of Alexandria, Egypt, was intended in 2002 by the author. The data was collected during the spring and autumn of that year. Only 340 households were successfully surveyed, out of which 170 households had complete records of household, person and trip information. The interviewed households were asked, among other questions, to state their household attributes: HouseHold Size (HHSize), number of Cars Owned by the HouseHold (HHCO), HouseHold annual income (HHInc), Age of the head of HouseHold (HHAge) and number of Workers in the HouseHold (Wkr/HH).

3.1. Sample statistics

In order to isolate the effect of each household attribute on the trip production, a correlation analysis was performed on the sample. The results are shown in table 1.

It is easy to notice that all considered attributes are positively correlated with trip production. The first four attributes have higher correlation while the fifth attribute, household income, has less correlation. It is worth noting that both car ownership and income have higher positive correlation with private car trips. Although correlated with car ownership, income has less effect on trip production than on mode choice. In general, car ownership has more explanatory power than income, as the correlation rates indicate.

The sample aggregate statistics are shown in table 2 in comparison with the population statistics for the city of Alexandria in 2002: the year of sample collection. The population statistics were based on data available from the Central Agency of People Mobilization and Statistics (CAPMAS) 1996 Census [4], CAPMAS statistical books of 2000 and 2001 [5,6] and the Ministry of Foreign Trade Monthly Digest of 2002 and 2003 [7,8]. It is important to emphasize here, that for the HHAge, Wkr/HH, and HHInc, the population statistics for the city of Alexandria were substituted by the national statistics. Table 2 also illustrates the sample bias in terms of the relative error between the sample and the population means of different attributes.

3.2. Sample weighing

In order to eliminate the effect of the sample bias, each sample record was factored using a weight to reflect its likelihood of occurrence in the population. Three weights associated with the three household attributes, HHAge, Wkr/HH, and HHInc were assigned to each record such that the corresponding weighted sample statistics comply exactly with those of the population. The record combined weight was calculated as the product of the three weights. As expected, although each weight would individually make the corresponding statistic fits its population

Table 1
Sample correlation between household attributes and trip production

HouseHold trips	HouseHold attributes				
	HHSize	HHCO	HHAge	Wkr/HH	HHInc
Trip production (total trips)	0.39	0.23	0.23	0.22	0.11
public transportation trips	0.13	-0.16	0.10	0.07	-0.12
Private car trips	0.11	0.60	0.18	0.11	0.33

Table 2
Comparison between sample and population statistics for the year 2002

	HouseHold attributes				
	HHSize	HHCO	HHAge	Wkr/HH	HHInc[a]
Sample	4.18	217	51.30	1.61	25188
Population	4.10	74	43.60	1.14	21789
Weighted sample	3.82	169	42.60	1.07	22926
Relative error (sample)			17.7%	41.2%	15.6%
Relative error (weighted sample)			-2.3%	-6.1%	5.2%

[a] : yearly income in LE

counterpart, the combined weight did not have the same perfect effect on the weighted sample. This is clear in terms of the little residual relative error that the weighted sample still has as shown in table 2. These small residual errors manifest the effectiveness of the weighing procedure relative to the three considered attributes. It is worth noting that no factoring associated with the first two attributes has been performed. These two attributes, HHSize, and HHCO, are chosen to be used in the model stratification.

4. Evaluation of model estimation results

Estimation involves finding the values of the parameters which make the observed data more likely under the model specification. The best set of parameters is defined by examining certain goodness-of-fit measures. These measures generally have well-known statistical properties which in turn allow confidence limits to be built around model parameters and predictions [8]. In our case, the use of mean trip rate for each class has lead to implement the Central Limit Theory (CLT) [3]. According to the CLT, a class mean is a normally distributed random variable with mean equals to the sample mean and standard deviation equals to the sample standard deviation divided by the square root of the number of observations (number of households) used to estimate the mean trip rate of that class.

4.1. Model stratification

As mentioned above (section 3.2), household size and household car ownership are chosen as the model stratification variables. There are two reasons for their choice. The first is the high correlation they have with trip production. The second is the easiness of forecasting their expected values and distributions in the future. The car ownership is easy to be estimated as the number of registered cars of each city is available in the CAPMAS statistical year book for the preceding year. The population of the city is one of the easiest aggregate attributes to be projected, having a good guess of the annual growth rate.

The household size attribute is a basic attribute of the decennial census at the city level. Although the change of the household size wouldn't be available between two censuses, no significant change in the household size is expected in such a short period of time. However, the number of households in the city of Alexandria can be projected based on available historical data.

The proposed model will be initially stratified based on household size of 1,2,3,4,5 and 6+ persons per household and number of cars owned of 0, 1 and 2+ cars per household with a total of 18 classes.

4.2. Model estimation results

The CCA and MCA models, formulated above, were estimated using the raw sampled data. Then, the same two models were estimated in a more compact stratification, i.e., less number of classes. Finally, the compact MCA model was estimated using the weighted sample. The reason of estimating the weighted model is to evaluate the effect of sample bias on the model behavior and the effectiveness of the weighing technique in eliminating that bias. The estimated cross-classification trip production models are shown in tables 3, 4 and 5. Tables 3-a through 3-c show the number of households observed of each household class, CCA model and MCA model, respectively.

As can be seen, the number of households surveyed, of different classes, ranges from 2 to 26. It is clear that all household classes have less than the traditional 30 observations required by statisticians to estimate average trip rates. The improbable case of households with one person and more than one car are not existing in the data. Although it is more likely, the case of households with one person and exactly one car is not represented in the surveyed households. It is worth noting, however, that the MCA model produces trip rates even for empty cells.

In addition, some counterintuitive progressions of trip rates are present in the case of CCA model. An example is the decrease in trip rate values for 0 cars/households when increasing household size from 3 to 4 and from 5 to 6. Another counterintuitive progression can be detected in the unchanged trip rates for household size of 2 persons when increasing cars/household from 1 to 2+. These counterintuitive progressions, fig.1-a, which may have arisen by the problem of small sample size, are eliminated in the case of MCA model, fig.1-b.

However, the trip rates of households with 3 and 4 persons are indifferent for various car ownership classes of both models, as shown in figs. 1-a and 1-b. Trip rates of households with 5 and 6+ persons are almost indifferent as well. To overcome the problem of cells with counterintuitive progression and almost indifferent trip rates, a more compact stratification of the household size was used to estimate these models. The proposed models will be stratified based on household size of 1,2,3-4,5+ persons per household. Table 4 illustrates the result of estimating this compact (12-classes) model. As can be seen, the number of households surveyed in different classes ranges from 2 to 43. While only one counterintuitive progression is found in the CCA model, no counterintuitive progression is found in the MCA model.

It is important to emphasize that the compact model is, theoretically, more aggregate in forecasting trip rates for households in the combined classes. However, it is more efficient, i.e., more accurate, to estimate mean trip rates with larger number of observations (households).

Fig. 2 shows the behavior of the compact MCA model relative to the HHSize. The model is smooth with no counterintuitive progression and is monotonically increasing with the increase of both household size and car ownership. From this point on, only the MCA model is considered.

Similar results of the weighted MCA model are given in tables 5. As expected, the weighted sample produced less overall trip rates. That is because the original sample was biased towards household attributes with higher positive correlation with trip production such as income, number of workers and age of householder.

4.3. Trip production and average trip rates prediction

To predict the base year aggregate trip production (or average trip rates) of an area, the number (percentages) of households of

Table 3
Cross classification trip production models (unweighted)

HHCO / HHSize	0	1	2+	Total
1	6	0	0	6
2	5	2	2	9
3	9	17	6	32
4	14	26	10	50
5	19	23	10	52
6+	5	11	5	21
Total	58	79	33	170

(a) Sample

HHCO / HHSize	0	1	2+	Trip Rate
1	3.7			3.7
2	5.8	6.0	6.0	5.9
3	8.8	8.2	11.3	9.0
4	8.6	8.4	11.3	9.1
5	11.4	12.1	12.5	11.9
6+	10.0	12.7	13.4	12.2
Trip Rate	8.9	10.0	11.7	9.95

(b) CCA

HHCO / HHSize	0	1	2+	Trip Rate
1	2.6	3.7	5.4	3.7
2	4.8	5.9	7.6	5.9
3	7.9	9.0	10.7	9.0
4	8.0	9.1	10.8	9.1
5	10.9	12.0	13.6	11.9
6+	11.2	12.3	14.0	12.2
Trip Rate	8.9	10.0	11.7	9.95

(c) MCA

Table 4
Compact cross classification trip production models (unweighted)

HHCO / HHSize	0	1	2+	Total
1	6	0	0	6
2	5	2	2	9
3,4	23	43	16	82
5+	24	34	15	73
Total	58	79	33	170

(a) Sample

HHCO / HHSize	0	1	2+	Trip Rate
1	3.7			3.7
2	5.8	6.0	6.0	5.9
3,4	8.7	8.3	11.3	9.0
5+	11.1	12.3	12.8	12.0
Trip Rate	8.9	10.0	11.7	9.95

(b) CCA

HHCO / HHSize	0	1	2+	Trip Rate
1	2.6	3.7	5.4	3.7
2	4.9	5.9	7.6	5.9
3,4	8.0	9.1	10.7	9.0
5+	11.0	12.1	13.7	12.0
Trip Rate	8.9	10.0	11.7	9.95

(c) MCA

Table 5
Compact MCA trip production models (weighted)

HHCO / HHSize	0	1	2+	Total
1	16	0	0	16
2	9	1	2	12
3,4	29	49	12	90
5+	17	33	4	53
Total	70	83	17	170

(a) Sample

HHCO / HHSize	0	1	2+	Trip Rate
1	2.1	4.0	6.7	3.5
2	3.9	5.8	8.5	5.3
3,4	6.3	8.2	11.0	7.8
5+	8.9	10.7	13.5	10.3
Trip Rate	6.6	8.5	11.2	7.98

(b) MCA

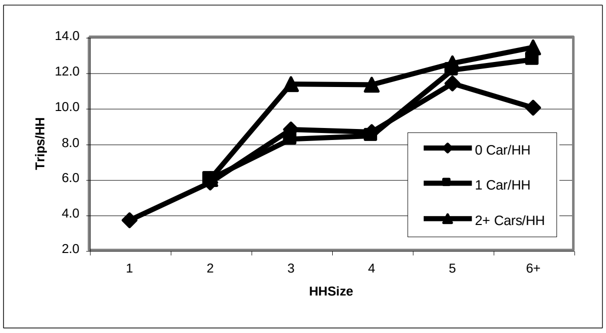

Fig. 1-a: CCA trip production model (unweighted).

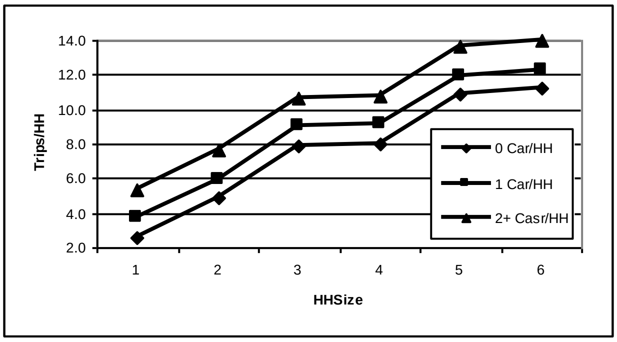

Fig. 1-b MCA trip production model (unweighted).

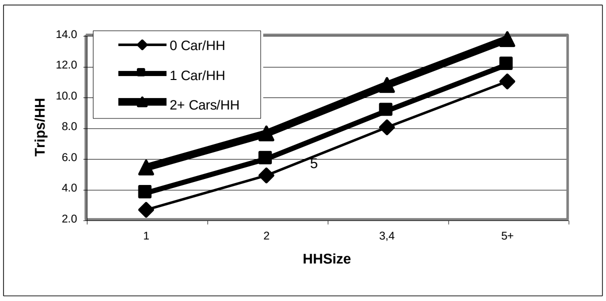

Fig. 2. Relationship between average household trips and HHSize of compact MCA trip production model (unweighted).

each household class needs to be estimated. To accomplish this task, the marginal distribution of the stratification variables; namely HHSize and HHCO, need to be known.

In terms of HHCO, the distribution of the household car ownership was available in the 1996 census [4]. This distribution corresponded to a total of 185,682 private cars owned by 3.32 million Alexandrians was updated for the year 2002 using simple extrapolation (see table 6). The registered cars were projected for the year 2002 to be 273,821 private cars based on annual growth rate of 6.7%. Similarly, the population was projected using an annual growth rate of 1.87%. The aggregate car ownership (Cars/1000 inhabitants) was estimated at 74 in the year 2002. This car ownership rate was used to update the HHCO distribution.

Similarly, the distribution of household size was available in the 1996 census [4]. This distribution corresponded to an average household size of 4.15 persons. The average household size was estimated to be 4.10 in 2002. This estimation was based on projecting number of households using an annual growth rate of .5% of the total population. The decrease in average household size was reflected by manipulating the percentages of household sizes of 5+ persons (see table 6). It is important to mention that the percentage of large households of sizes 5+ persons was kept at its original value of 41%.

For the year 1982, both HHSize and HHCO marginal distributions were defined based on the aggregate car ownership and average household size for that year (see Table 6). The HHCO distribution was based on the 1996 distribution updated using the aggregate car ownership of 37 cars/1000 inhabitants for 1982. The HHSize also stemmed from the 1996 distribution updated to reflect the average household size of 4.75 persons for 1982.

The number of cars found in the 1982 O-D survey by TranSystem [3], i.e., those owned and used by members of the households, was much lower than the number of registered cars. Part of this difference was due to the company cars and cars registered but not present in Alexandria and, therefore, unutilized by household members. This particular aspect was confirmed by observing car flows in the city sections and cars parked overnight in the city streets as well [3]. The expected ratio between actually usable and registered cars in Alexandria was found to be about 57 %. The number of cars observed by the 1996 census [4] composes a ratio of about 35% of the registered cars in the same year. Thus, the HHCO marginal distributions to be used in estimating the population cross distribution was based on the assumption that the cars usable by the household members are about 50% of the registered cars.

Table 6
Estimating HHSize & HHCO marginal distributions of Alexandria

Year variable	1982	1986	1996	2000	Average annual Growth rate	2002 (Projected)
Cars owned	97836		185682	240566	6.69%	273821
Population (million)	2.65	2.90	3.32	3.58	1.87%	3.71
# Households	558431	642699	799755		0.50%[a]	906293
Cars/HH	0.175		0.232			0.302
Car ownership/1000 inhabitants	37		56			74
HHCO						
0 Car (%)	82.80		77.2			70.31
1 Car (%)	16.83		22.3			29.02
2+ Cars (%)	0.38		0.5			0.67
Average household size	4.75	4.51	4.15			4.10
HHSize						
1 (%)	5		7.5			7.5
2 (%)	9		13.5			13.5
3(%)	14.0		15.8			15.8
4(%)	20.0		22.2			22.2
5(%)	21.0		18.9			21.9
6+(%)	31.0		22.1			19.1

[a] : this percentage is the annul growth in number of households as a function of total population

The marginal distributions of the two basic variables were used to estimate the cross distribution of the population using the Iterative Proportional Fitting (IPF) technique. For more information about the theoretical basis of this technique see Beckman et al. [9]. The cross distribution was seeded with the sample information. The resulted cross table of the city of Alexandria in 2002 based on HHSize and HHCO is given in table 7-a. Similar process was performed to estimate the cross table of 1982 which is also shown in table 7-b.

Using the estimated cross classification model and the population cross table of 2002, average trip rates were calculated for the MCA model. Table 8 shows the resulted trip rates (trips/person/day) of the model for 2002 in terms of motorized trips, walk trips and total trips. A comparison between the model trip rates and those trip rates estimated by TranSystem study of 1982 is also given. The MCA model total trip rate is higher (by 43%) than that of TranSystem. Looking more closely will reveal that while the motorized trip rate is much higher (by 86%) than that of Transystem while the walk trip rate is less (by 18%).

The estimated MCA model can be used in forecasting zonal trip production by projecting the cross distribution of zones at any level and applying the model trip rates. This trip production model along with trip attraction model is the basis for the next step of the urban transportation planning process: trip distribution.

By examining the MCA model, shown in table 5, the following observation is noticeable. Although the household mean trip rates increase as the household size increases, the trip rates increase is not proportional to the household size increase. Using the sample average household sizes of 3.6 and 5.8 persons/HH for the household groups of 3-4 and+, respectively, the individual trip rates are obtained. It is interesting that the individual trip rates are decreasing with the increase of the household size for all car ownership groups. This is illustrated in fig. 3. The significance of this finding is that one should expect the average trip rate to decrease with the increase of the average household size. In other words, trip production is expected to increase for the same population with smaller households. It is important to notice that the individual trip rate of households with 1 person and 2+ cars is very high (6.7 trips/person/day). This is not significant since the existence of such a household is very rare and, consequently, would not alter the average trip rates (or trip production).

Table 7
Marginal and cross distributions (%) of Alexandria

HHCO HHSize	0	1	2+	Total
1	7.5	0.0	0.0	7.5
2	12.8	0.7	0.0	13.5
3 &4	30.4	7.4	0.1	38.0
5+	34.5	6.4	0.2	41.0
Total	85.2	14.5	0.3	100.0

a) 2002

HHCO HHSize	0	1	2+	Total
1	5.0	0.0	0.0	5.0
2	8.8	0.2	0.0	9.0
3 &4	30.2	3.7	0.1	34.0
5+	47.4	4.4	0.1	52.0
Total	91.4	8.4	0.2	100.0

b) 1982

Table 8
Comparison between estimated MCA model and TranSystem trip rates (trips/person/day)

Mode	Motorized	Walk	Total	Walk (%)
MCA model (2002)	1.30	0.41	1.71	24.1
TranSystem (1982)	0.70	0.50	1.20	41.7
Relative error (%)	85.9	-17.6	42.8	-42.3

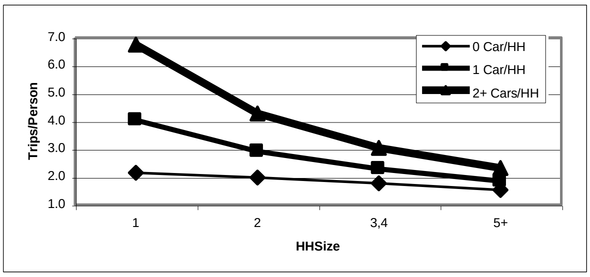

Fig. 3. Relationship between individual trip rates and HHSize.

4.4. Goodness-of-Fit

Table 9 shows the standard error (*Se*) of each class mean (weighted and unweighted) for both CCA and MCA models. The confidence intervals at 95% confidence level can be constructed for each class as follows:

$$CI = Mean \pm Z_{\alpha} * Se \,, \qquad (9)$$

where: Z_{α} is the value of standard normal variate for confidence level $1-\alpha$.

The confidence interval, which is bounded by the lower bound (minimum) and upper bound (maximum) trip rate, should include the correct rate with 95% probability. It is worth noting that the standard errors of the MCA model are smaller than those of the CCA model for almost all (but two) classes. This means, in general, that the CI is smaller for the MCA model than that for the CCA model. It is important to emphasis that the fact that the standard errors are larger for the weighted MCA model than for the unweighted model doesn't mean that the weighted model is inferior. That is because the unweighted model was proven to be biased (see section 3.1). It is fair to state that the weighted model is not *as good* in terms of goodness-of-fit but is, indeed, *unbiased*. By increasing the sample size, the unweighted (biased) model would simply collapse into an unbiased model without increasing its standard errors.

To evaluate the goodness-of-fit for the estimated average trip rates, the average standard error for the weighted MCA model was calculated with a value of 0.28 trips/person/day. Substituting in eq. (9), the CI of the estimated MCA model was constructed at 95% confidence level in terms of lower and upper bound trip rates for 2002. Table 10 illustrates this confidence interval.

To find the sample size required to reduce the standard error of the mean trip rate, the admissible error (e_a) has to be determined. Then, the required sample size, at the same accuracy level, can be calculated as a function of the current sample size as follows [10]:

Table 9
Comparison between standard errors of CCA and MCA (trips/person/day)

		Class standard errors									Average standard errors		
		CCA			MCA			se(MCA)/se(CCA)			CCA	MCA	se(MCA)/se(CCA)
	HHCO HHSize	0	1	2+	0	1	2+	0	1	2+			
Unweighted	1	0.1	0.0	0.0	0.1	0.1	0.1	1.1	N.A	N.A			
	2	0.3	0.5	0.2	0.1	0.1	0.1	0.4	0.2	0.5	0.23	0.12	0.52
	3 &4	0.2	0.1	0.2	0.1	0.1	0.1	0.5	0.9	0.6			
	5+	0.3	0.3	0.4	0.1	0.1	0.1	0.5	0.5	0.4			
Weighted	1	0.3	0.0	0.0	0.2	0.3	0.3	0.9	N.A	N.A			
	2	0.7	0.7	0.2	0.3	0.3	0.3	0.4	0.5	1.3	0.59	0.28	0.48
	3 &4	0.3	0.6	0.6	0.3	0.3	0.3	0.8	0.5	0.5			
	5+	0.8	0.6	2.4	0.3	0.3	0.3	0.4	0.6	0.1			

$$n_r \geq (Z_\alpha * Se)^2 / e_a{}^2 * n_c \,, \tag{10}$$

where: n_r is the required sample size to satisfy the admissible error: e_a and

n_c is the current sample size.

For 10% error of the estimated mean trip rate, i.e., .17 trips/day/person, of the same model specification, a sample of about 1800 observations is required. Let us assume that, with increasing sample size, the biased model will collapse into an unbiased model keeping its standard error at the original low level of 0.12 trips/person/day. In this case the required sample size will decrease to less than 400 observations. Of course, a larger sample size is required for models with multi variable stratification. The expected increase in the sample size is directly proportional to the number of classes. A similar calculation of the required sample size for the CCA model produced a range between 700 and 4300 observations. This difference in the required sample size between MCA and CCA models is consistent with the basic definition of the MCA model (see section 2.2.).

5. Model validation

Because the large majority of transportation models are built on the basis of cross-sectional data, there has been a tendency to interpret model validation exclusively in terms of goodness-of-fit achieved between observed behavior and base year predictions. Although this is a necessary, it is not sufficient, condition for model validation. Validation requires comparing the model predictions with information not used in the process of model estimation. This has been demonstrated by number of cases where model predictions were compared with observed behavior in before-and-after studies [1].

It is clear that this is not possible in our case. Therefore, alternative backward validation, in which the model predictions for year 1982 will be compared with trip rates observed in the same year available in TranSystem [3], is proposed in this paper. To perform this backward validation, the cross table of 1982 was used along with the estimated MCA model. Table 11 illustrates this comparison.

The validated model was obtained by dividing class trip rates of the estimated MCA model (table 5) by a calibration factor equals to the ratio between model predicted total trip rate for 1982 and the observed total trip rate of the TranSystem. The calibration factor was found to be 1.3. The validated MCA trip production model is given in table 12.

Table 13 shows the resulted trip rates (trips/person/day) of the validated cross classification model for 2002. From table 13, it easy to notice that the average trip rates

Table 10
Confidence interval of the estimated MCA model trip rates for 2002 (trips/person/day)

Mode	Motorized	Walk	Total
MCA Model (2002)	1.30	0.41	1.71
Model Lower Bound	0.88	0.28	1.15
Model Upper Bound	1.73	0.55	2.27

Table 11
Comparison between model predictions and observed trip rates (1982)

Mode	Motorized	Walk	Total	Walk (%)
Model prediction (1982)	1.17	0.38	1.56	24.6
Observed (TranSystem)	0.70	0.50	1.20	41.7

Table 12
Validated MCA trip production model

HHCO HHSize	0	1	2+	Trip Rate
1	1.7	3.1	5.2	2.7
2	3.0	4.5	6.6	4.1
3,4	4.9	6.4	8.5	6.0
5+	6.8	8.3	10.4	7.9
Trip Rate	5.1	6.5	8.6	6.2

predicted by the validated MCA model are within the confidence interval boundaries of the estimated MCA model. It is important to emphasis that the expected total trip rate (of the validated MCA model) is only 10% larger than the trip rate observed in 1982. More importantly, the expected motorized trip rates is larger than that of 1982 by more than 43%, while the walk trip rate is lower by about 36%.

6. Conclusions

This paper serves as a framework for a more comprehensive sampling effort with larger sample size. The cross-classification modeling procedure illustrates the potential provided by the MCA model. The model specification procedure shows that household size and car ownership have the highest explanatory power among all other household attributes. Other household attributes such as age of householder and number of workers per household have the potential to be good estimators although they are more difficult to project in the future than household size and car ownership.

The MCA model is data-efficient due to its dependence on the entire sample of households in estimating trip rates of different household classes. The compact model form (with less number household classes) is more efficient in estimating trip rates although more aggregate in forecasting trip productions. Moreover, the weighing procedure applied to the sample is effective in offsetting the bias of the small at the expenses of larger estimation error. A larger sample size between 400 and 1800 observations is required to minimize the standard error for the same model specification. Even larger sample would be required for models with detailed attribute classes and more attributes.

The model is sensitive and can reflect changes in socioeconomic household attributes in terms of its clear trendency to increase relative to the considered attributes. Also, the model is suitable to forecast trip production at zonal level and average trip rates as well. Interestingly, the model application shows that the average household size is inversely proportional to the average trip rate. That is, the recent tendency of smaller household size will result in more trips relative to population. Coupled with the rapid increase in the car ownership, average trip rate is expected to go even higher. Additionally, a simple, yet effective, method of estimating the cross distribution of the household attributes based on the available census data, is introduced. Finally, there is a considerable indication that the total trip rate has increased in 2002 by about 10% since 1982. More significantly, motorized trips have increased by over 40% while walk trips has decreased by over 35%. This puts a real burden on the transportation networks.

Table 13
Model validation

Model	Trips rate			Relative error (%)		
	Motorized	Walk	Total	Motorized	Walk	Total
MCA model (2002)	1.30	0.41	1.71	85.9	-17.5	42.8
Model lower Bound	0.88	0.28	1.15	25.2	-44.5	-3.9
Model upper Bound	1.73	0.55	2.27	146.7	9.4	89.5
Validated MCA model (2002)	1.00	0.32	1.32	43.6	-36.3	10.3
TranSystem (1982)	0.70	0.50	1.20			

References

[1] J. Ortuzar, D. de and L.G. Willumsem, Modeling Transport, John Wiley & Sons Ltd, England, Second Edition, Dec. (1996).
[2] P.R. Stopher, and K.G. McDonald, Trip Generation By Cross Classification: an Alternative Methodology, Transportation Research Record 944, Transportation Research Board, Washington, D.C., pp. 84-91 (1983).
[3] TRANSYSTEM, Public Transportation Study- City of Alexandria, Alexandria (1982).
[4] Central Agency of People Mobilization and Statistics (CAPMAS), 1996 Census, April (1999).
[5] Central Agency of People Mobilization and Statistics (CAPMAS), The Statistical Year Book, pp. 1993-1999 (2000).
[6] Central Agency of People Mobilization and Statistics (CAPMAS), The Statistical Year Book 2000 (2001).
[7] Ministry of Foreign Trade, Monthly Economic Digest, April (2002).
[8] Ministry of Foreign Trade, Monthly Economic Digest, May (2003).
[9] R.J. Beckman, K.A. Baggerly, and M.D. McKay, Creating Synthetic Baseline Populations, Transportation Research-A, Vol.30 (6) pp. 415-429 (1996).
[10] W. Mendenhall, D.D. Wackerly R.L. and Scheaffer, Mathematical Statistics with Applications, PWS-KENT Publishing Company, Boston, 4th Edition (1990).

Received December 31, 2003
Accepted January 22, 2004