Journal of Traffic and Transportation Engineering(English Edition)
2014,1(3):196-208

Derivation of travel demand forecasting models for low population areas: the case of Port Said Governorate, North East Egypt

Ahmed Mohamed Semeida*

Department of Civil Engineering, Faculty of Engineering, Port Said University, Port Said, Egypt

Abstract: In the last decades, there has been substantial development in modeling techniques of travel demand estimation. For low population areas the external trip estimation is important but usually neglected in travel demand modeling process. In Egypt, the researches in this field are scarce due to lack of data. Accordingly, this paper aims to identify and estimate the main variables that affect the travel demand in low population areas, and to develop models to predict them. The study focused on the Port Said Governorate in North East Egypt. A special questionnaire had been prepared in 2010 depending on interviews of passengers at basic taxi terminals in Port Said. And 2211 filled questionnaires were offering for research. To analyze the data, two modeling procedures were used. One is the multiple linear regression and the other is the generalized linear modeling (GLM) applying normal distributions. It is found that GLM procedure offers more suitable and accurate approach than the linear regression for developing number of trips. The final demand models have statistics within the acceptable regions and, also, they are conceptually reasonable. These results are so important for Egyptian highway authorities to improve the efficiency of highway transportation system in Egypt.

Key words: travel demand; linear regression; generalized linear modeling; low population areas; Port Said

1 Introduction

Transportation planning in Port Said has become more important. Although, this region has only 0.75% of total population of Egypt, but it strongly participates in national income. Therefore, it is necessary to peruse and analyze the passengers' trips from and to this significant region. The aim of the paper is to identify and select the main variables that affect the travel demand, and to develop models to predict them in Port Said, North East Egypt. The models are developed and calibrated based on two

* **Corresponding author**: Ahmed Mohamed Semeida, PhD, Assistant Professor.
E-mail: semeida77@hotmail.com.

modeling procedures. One is the multiple linear regression, and the other is the GLM applying normal distributions. All the used variables are available and will be explained in details in the section of data collection.

In the previous researches, the statistical methods and especially the linear or quasi-linear regression (simple or multiple) were widely used for travel demand forecasting. The performances of these regression models are evaluated by several techniques, e. g., the correlation coefficient, analysis of variance, factor, cluster analysis, and discriminant analysis, etc. (Stopher and McDonald 1983; Oppenheim 1995). Alfa (1986) used regression models to estimate bus running times for fixed route and fixed-schedule transit service in Winnipeg, Canada. The method has been criticized for some drawbacks. However, it is still in use for its simplicity, especially in the case where other more complex methods fail to perform satisfactorily. Koppehnan and Hirsh (1986) reviewed some of the better known models in intercity travel demand. They discussed the calibration and application contexts, the peculiarities and the characteristics of these models. Hutchinson (1993) had reviewed attempts made in Canada to model air travel demand using aggregate and disaggregate demand models. He concluded that most of the models estimated so far had very limited success in reflecting changes in the spatial socioeconomic organization and/or in the transport milieu. Varagouli et al. (2005) identified the main variables that affected the travel demand in tile intercity passenger transportation concerning the prefecture of Xanthi in Northern Greece. The models had been developed based on multiple linear regression analysis. The final demand models had statistics within the acceptable regions and were conceptually reasonable. Anderson et al. (2006) determined an interaction between small communities and nearby major cities (or highway facilities). His studies indicated that the city of interest was not an isolated island and that the economic activities in the market area surrounding a city impacted the through and external trip patterns. Han and Stone (2008) clarified the definition of nearby major city in through trip estimation and statistically approved the significant influence of a study area's economic context on patterns of through trips and, eventually, the external trips. Qian et al. (2012) developed a cost-effective method to forecast external trips from an economic point of view. Based on recent survey data, separate external trip models were developed by urban categories. The new models minimize data requirements and are easy to use. They appear transferable to other small and medium urban areas.

In Egypt, until now there are no research concerns on the external trip estimation in low population areas. There are many reasons as follows: the lack of trip data in these zones; the inattention of supervisors and researchers towards these small zones; this type of data need great efforts and finance to be collected accurately; as Egypt ranks developing country, this type of research is obscure and forgotten; the author is resident in one of these zones and realizes the economical importance of it towards national Egyptian income; Port Said Governorate under study is considered the main entrance to Suez Canal from all over the world. Thus, it is so important to study this issue in details which would improve the efficiency of Egyptian transportation system.

Many reasons convince the author to utilize these methods instead of the four-step based procedure as follows:

- The four-step procedure was used in the literature by the majority of researchers all over the world. Therefore, the innovation of another new and suitable manner is so important and beneficial.
- The main aim of this research is to investigate the impact of many factors on number of personal trips per year. So, the use of one model which introduces the relation between one dependent variable and many independent variables is necessary and helpful in this case.
- In the present research, the author concentrates on one zone only to execute the analysis. So, no need to distribute trips on other zones. The number of trips from and to this zone is needed only to complete the analysis. Hence, one single model is enough and use an indicator instead of four steps.
- The four-step procedure is valid and effective for preparing long technical reports and general stud-

ies about all Egypt zones with extended length of pages. On the other hand, the paper must concern on one zone with specific characteristics in limited number of pages. The used modeling methods verify this goal effectively.

Therefore, separate models according to trip purpose and position of Port Said in trip (origin or destination) using linear regression and GLM analysis are presented in this paper.

2 Data collection, sources, and processing

The census of Port Said is 603787 residents according to the Egyptian Central Agency for Public Mobilization and Statics (ECAPMS 2010). According to the same source, the area of Port Said is 1351 km^2. Thus, the residents of this governorate estimate only 0.75% of the total census and 0.14% of the total area of Egypt. Port Said distinctly contributes in the Egyptian national income as it has an important position in commerce, industry, education, and tourism. Consequently, it attracts and generates yearly a huge number of personal trips from and to the other governorates in Egypt. Therefore, the derivation of travel demand models in this region is so necessary and important for transportation system development in Egypt.

2.1 Data sources

There are three types of data sources discussed in the following sections.

2.1.1 *Questionnaire data*

A special questionnaire had been prepared in 2010 depending on interviews of passengers at basic taxi terminals in Port Said. The time of interviews was from 6:00 AM to 10:00 PM in two months, April and May. The results of this questionnaire consist of 2211 filled forms (interviews). This questionnaire is part of a project executed in Egypt with title "The study of preparation of inclusive transportation scheme on national level". Two main authorities participated in this project. First is the Japanese international cooperation authority (JICA), and the other is the general authority for planning transportation projects, Ministry of Transportation.

Each form concerns an interview with one person and contains the following data: trip origin governorate, trip destination governorate, time and date of trip, trip purpose, and trip frequency in the year, residence governorate and district, access mode time and cost for origin governorate, and egress mode time and cost for destination governorate.

2.1.2 *Distance data*

There are two types of distances, distances between taxi terminals (distances on rural roads), and distances between terminals and districts in origin and destination (distances on urban roads). Both distances were measured from Google earth maps with knowledge of the exact districts name and their locations on maps. These distances are so important and effective factors that affect the number of personal trips per year and represent main independent variable in the present research.

The roads network in Egypt and the location of Port Said at Suez Canal Zone are shown in Figs. 1,2, respectively. In addition, the areas of all governorates on Egypt map are indicated in Fig. 3.

2.1.3 *Personal average income*

The average personal income in Egyptian pounds (EP) per year for each governorate was obtained from ECAPMS. As these values are widely different throughout all governorates. They were used as a main variable in the present research.

2.2 The used variables

All the previous variables and their symbols are calculated for each personal trip and provided in Tab. 1. Consequently, the research used a total number of 13 variables that were divided into dependent and independent variables. The main dependent variable which presents the travel demand is Tr (number of trips per person per year). The other 12 variables represent the independent variables that affect the travel demand. The forms include 24 Egyptian governorates that represent nearly all Egyptian governorates. Therefore, this study is well comprehensive. Each governorate has a code number from 1 to 24. For example, the region canal governorates have numbers of 1, 2, and 3 for Port Said, Ismailia, and Suez, respectively. The codes of all governorates and their properties (population, area, population density, and personal average income per year) are presented in Tab. 2.

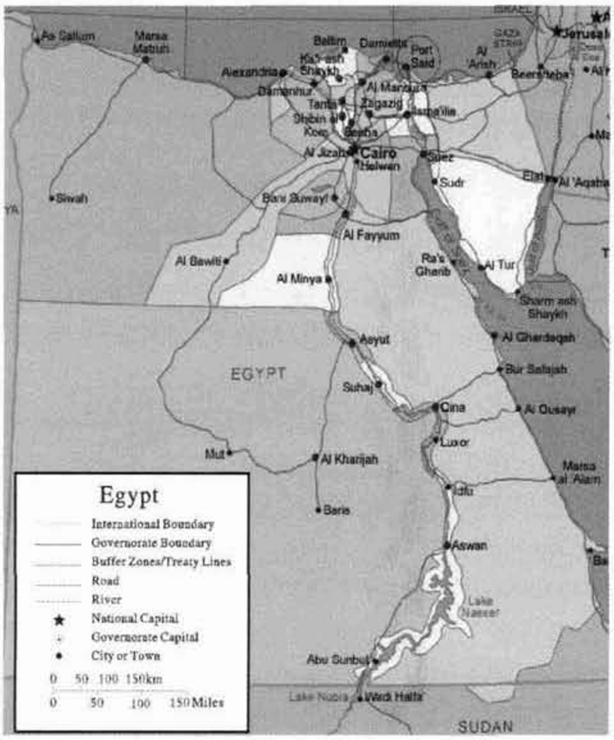

Fig. 1 Roads network on Egypt map

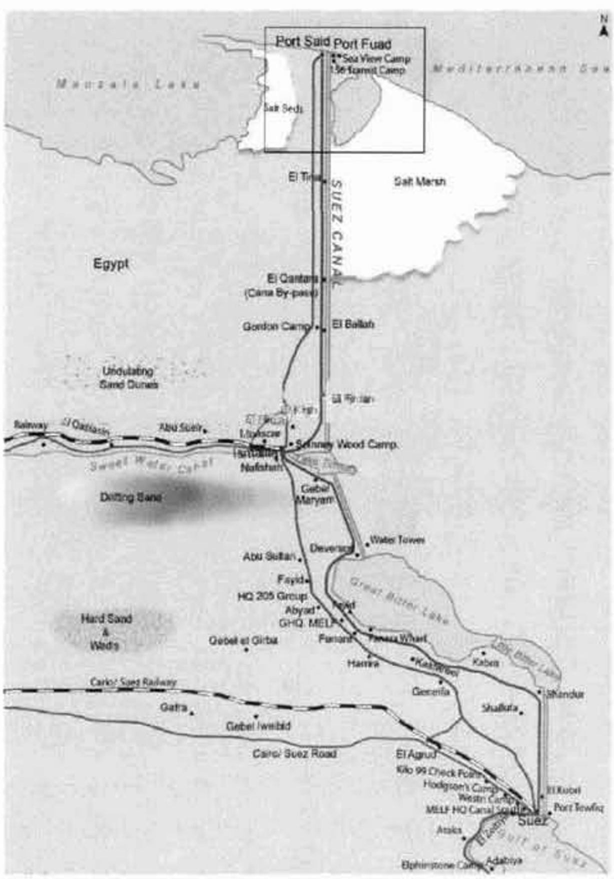

Fig. 2 Location of Port Said at Suez Canal Zone

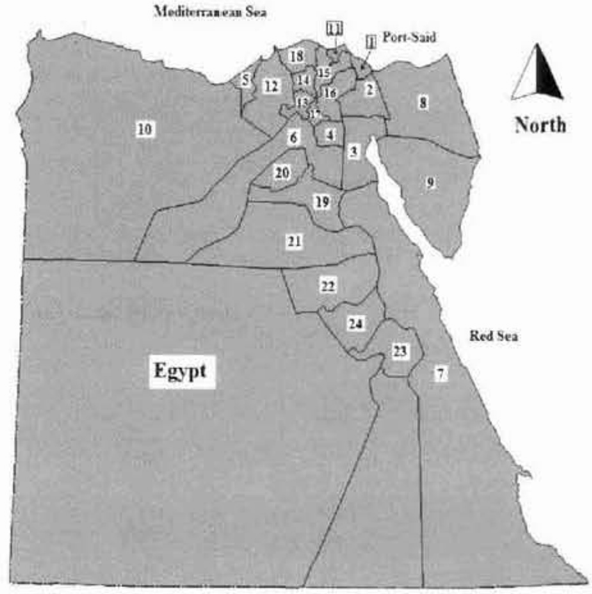

Fig. 3 Area and code numbers of all governorates on Egypt map

Tab. 1 Variables and symbols for each personal trip

Variable	Variable symbol	Max.	Min.	Avg.
1-Total distance between origin and destination in km	Dt	988	25	167
2-Distance between terminals in km (rural distance)	Dr	972	24	146
3-Population of origin governorate 2010 in person	Popo	9891700	65495	3547483
4-Population of destination governorate 2010 in person	Popd	9891700	65495	3547483
5-Area of origin governorate in km^2	Ao	203685	589	28069
6-Area of destination governorate in km^2	Ad	203685	589	28069
7-Population density of origin governorate in person/km^2	Dino	5259	1	979
8-Population density of destination governorate in person/km^2	Dind	5259	1	979
9-Average personal income for origin governorate in EP/year	Inco	29033	11392	11187
10-Average personal income for destination governorate in EP/year	Incd	29033	11392	11187
11-Total time in internal (urban) distances in minutes	Totit	190	6	38
12-Total cost in internal (urban) distances in EP	Totic	37	0	4
13-Trip frequency per year in trips/person/year	Tr	322	1	84

Tab. 2 Codes and properties of all Egyptian governorates

Code	Governorate	Population (persons)	Area (km^2)	Density (persons/km^2)	Avg. personal income (EP/year)
1	Port Said	603787	1351	446.87	26181
2	Ismailia	942832	5066	186.11	26312
3	Suez	651848	10000	65.18	28969
4	Cairo	9891700	1881	5258.63	23776
5	Alexandria	4987509	2679	1861.71	25474
6	Giza	7025279	85153	82.50	19690
7	Red Sea	301233	203685	1.42	29033
8	N. Sinai	395713	27564	14.36	22354
9	S. Sinai	65495	28438	2.30	29032
10	Matrouh	3990670	166563	23.96	26393
11	Damietta	1260930	589	2140.80	15380
12	Beheira	5254820	9119	576.24	17456
13	Minufia	3580000	2543	1407.77	16923
14	Gharbia	4400298	1942	2265.86	14577
15	Dakahlia	5551592	3459	1604.97	17980
16	Sharkia	6884000	4911	1401.75	15872
17	Qalubia	4740000	1001	4734.84	13745
18	Kafr-elshaikh	2940030	3748	784.43	18429
19	Beni-Suef	2520629	10954	230.11	12324
20	Fayoum	2712792	6068	447.01	12520
21	Minya	3796000	32279	117.60	13765
22	Asyut	6510500	25926	251.12	12805
23	Qena	3276700	10798	303.45	11712
24	Sohagc	5210500	21400	243.48	11392

Source: ECAPMS 2010.

2.3 Data processing

The main stage in data processing is the construction of origin destination matrix. The main variable in this matrix is the average number of trips per person per year (Tr) from and to Port Said. Before the access of this stage, it is necessary to distribute the forms according to trip purpose, as the trip purpose has a great effect on personal trips number per year. There are three main purposes, the first is work, the second is education, and the third is combined with different purposes from the other two purposes. Thus, the third purpose includes tour, shopping, business, and social visits.

Also, the origin and destination forms are separated. Therefore, the analysis includes a total of six travel demand models [2 (origin or destination) × 3 (work, education or other trip purposes)].

The models data with O-D pairs are presented in Tabs. 3, 4. Tab. 3 is for Port Said as origin, and Tab. 4 is for Port Said as destination.

Tab. 3 Port Said models data (Port Said as origin)

O-D pairs	Trip purpose	Tr	Dt	Dr	Popo	Popd	Ao	Ad	Dino	Dind	Inco	Incd	Totit	Totic
1→2	Work	210	80	66	603787	942832	1351	5066	447	186	26181	26312	26.1	4
1→3	Work	53	196	168	603787	651848	1351	10000	447	65	26181	28969	43.0	5
1→4	Work	90	249	220	603787	9891700	1351	1881	447	5259	26181	23776	54.3	4
1→5	Work	50	310	272	603787	4987509	1351	2679	447	1862	26181	25474	57.5	12
1→6	Work	38	251	220	603787	7025279	1351	85153	447	83	26181	19690	58.8	4
1→9	Work	4	556	536	603787	65495	1351	28438	447	2	26181	29032	35.0	1
1→11	Work	260	86	66	603787	1260930	1351	589	447	2141	26181	15380	35.4	5
1→14	Work	232	151	140	603787	4400298	1351	1942	447	2266	26181	14577	18.0	3
1→15	Work	161	113	98	603787	5551592	1351	3459	447	1605	26181	17980	24.9	5
1→16	Work	118	176	158	603787	6884000	1351	4911	447	1402	26181	15872	30.7	3
1→2	Education	115	89	77	603787	942832	1351	5066	447	186	26181	26312	22.7	4
1→3	Education	140	198	168	603787	651848	1351	10000	447	65	26181	28969	60.0	2
1→4	Education	36	244	220	603787	9891700	1351	1881	447	5259	26181	23776	43.3	5
1→6	Education	28	253	220	603787	7025279	1351	85153	447	83	26181	19690	60.0	7
1→8	Education	30	300	272	603787	395713	1351	27564	447	14	26181	22354	50.0	7
1→11	Education	76	80	66	603787	1260930	1351	589	447	2141	26181	15380	21.9	6
1→15	Education	59	111	98	603787	5551592	1351	3459	447	1605	26181	17980	21.5	4
1→16	Education	50	175	158	603787	6884000	1351	4911	447	1402	26181	15872	41.7	5
1→2	Other	6	89	74	603787	942832	1351	5066	447	186	26181	26312	26.2	3
1→3	Other	2	204	168	603787	651848	1351	10000	447	65	26181	28969	65.0	4
1→4	Other	5	243	220	603787	9891700	1351	1881	447	5259	26181	23776	41.4	6
1→5	Other	12	282	272	603787	4987509	1351	2679	447	1862	26181	25474	20.0	1
1→6	Other	1	248	220	603787	7025279	1351	85153	447	83	26181	19690	65.0	5
1→9	Other	1	546	526	603787	65495	1351	28438	447	2	26181	29032	35.0	4
1→11	Other	25	79	66	603787	1260930	1351	589	447	2141	26181	15380	41.0	3
1→14	Other	2	142	125	603787	4400298	1351	1942	447	2266	26181	14577	25.0	6
1→15	Other	5	115	98	603787	5551592	1351	3459	447	1605	26181	17980	27.5	4
1→16	Other	5	169	156	603787	6884000	1351	4911	447	1402	26181	15872	26.0	2

Tab. 4 Port Said models data (Port Said as destination)

O-D pairs	Trip purpose	Tr	Dt	Dr	Popo	Popd	Ao	Ad	Dino	Dind	Inco	Incd	Totit	Totic
2→1	Work	228	71	59	942832	603787	5066	1351	186	447	26312	26181	25.5	3
4→1	Work	37	239	220	9891700	603787	1881	1351	5259	447	23776	26181	36.3	4
5→1	Work	43	297	272	4987509	603787	2679	1351	1862	447	25474	26181	41.5	7
6→1	Work	33	255	220	7025279	603787	85153	1351	83	447	19690	26181	66.8	5
11→1	Work	213	88	66	1260930	603787	589	1351	2141	447	15380	26181	37.6	4
12→1	Work	34	242	217	5254820	603787	9119	1351	576	447	17456	26181	43.1	7
13→1	Work	23	202	178	3580000	603787	2543	1351	1408	447	16923	26181	45.0	4
14→1	Work	156	163	140	4400298	603787	1942	1351	2266	447	14577	26181	41.2	4
15→1	Work	124	120	97	5551592	603787	3459	1351	1605	447	17980	26181	40.2	5
18→1	Work	29	185	164	2940030	603787	3748	1351	784	447	18429	26181	38.3	3
2→1	Education	77	82	65	942832	603787	5066	1351	186	447	26312	26181	32.2	4
3→1	Education	56	183	168	651848	603787	10000	1351	65	447	28969	26181	25.0	4
5→1	Education	29	305	272	4987509	603787	2679	1351	1862	447	25474	26181	52.0	15
8→1	Education	14	229	204	395713	603787	27564	1351	14	447	22354	26181	50.0	6
11→1	Education	113	88	66	1260930	603787	589	1351	2141	447	15380	26181	36.8	4
12→1	Education	24	289	268	5254820	603787	9119	1351	576	447	17456	26181	36.0	4
14→1	Education	76	165	144	4400298	603787	1942	1351	2266	447	14577	26181	38.8	4
15→1	Education	58	125	95	5551592	603787	3459	1351	1605	447	17980	26181	54.8	4
18→1	Education	15	211	164	2940030	603787	3748	1351	784	447	18429	26181	70.0	10
2→1	Other	8	78	64	942832	603787	5066	1351	186	447	26312	26181	25.1	3
3→1	Other	8	179	168	651848	603787	10000	1351	65	447	28969	26181	21.7	2
4→1	Other	8	248	220	9891700	603787	1881	1351	5259	447	23776	26181	57.5	3
5→1	Other	3	294	272	4987509	603787	2679	1351	1862	447	25474	26181	35.3	9
6→1	Other	8	257	220	7025279	603787	85153	1351	83	447	19690	26181	70.0	5
8→1	Other	8	245	238	395713	603787	27564	1351	14	447	22354	26181	17.5	3
9→1	Other	2	453	436	65495	603787	28438	1351	2	447	29032	26181	25.0	13
11→1	Other	15	86	66	1260930	603787	589	1351	2141	447	15380	26181	33.1	4
12→1	Other	2	277	247	5254820	603787	9119	1351	576	447	17456	26181	55.0	5
14→1	Other	4	168	143	4400298	603787	1942	1351	2266	447	14577	26181	44.3	5
15→1	Other	7	118	97	5551592	603787	3459	1351	1605	447	17980	26181	35.0	5
16→1	Other	4	187	154	6884000	603787	4911	1351	1402	447	15872	26181	62.5	6
24→1	Other	1	726	712	5210500	603787	21400	1351	243	447	11392	26181	35.0	2

3 Modeling procedure

The procedure of travel demand prediction in the present research is divided into two main methods: multiple linear regression modeling and GLM.

3.1 Multiple linear regression modeling

Semeida (2013; 2014) used this method successfully for model operating speed at multi-lane highways on tangents and horizontal curves, respectively. Therefore, it is necessary to start modeling with this method. The collected data were used to investigate the relationships between travel demand dependent variable (Tr) and the other 12 variables that are defined before as independent variables. Simple regression is used to check the correlation coefficient (R) between the dependent variable and each of the independent variable. The independent variables that have relatively high R^2 values ($R^2 > 0.5$) are introduced into the multiple linear regression models. The form of multiple linear regression models is shown in Eq. (1).

$$\mathrm{Tr} = \beta_0 + \sum_{i=1}^{n} \beta_i x_i \tag{1}$$

where Tr is number of trips per person per year; x_i is explanatory variables from 1 to n; β_0 is regression constant; β_i is regression coefficient.

Then, stepwise regression analysis was used to select the most statistically significant independent variables with dependent variable in one model. Stepwise regression starts with no model terms. At each step, it added the most statistically significant term (the one with lowest p-value) until the addition of the next variable made no significant difference. An important assumption behind the method is that some input variables in a multiple regression do not have an important explanatory effect on the response. Stepwise regression kept only the statistically significant terms in the model.

Finally, the R^2 (coefficient of determination) and $\|\delta\|$ (infinity norm of error vector) values were calculated for each model. According to the research (Varagouli et al. 2005) the acceptable region was defined such that $\|\delta\|$ was accurate to 2 decimal digits. Several precautions were taken into consideration to ensure integrity of the model as follows (Simpson et al. 1994):

(1) The signs of the multiple linear regression coefficients should agree with the signs of the simple linear regression of the individual independent variables and agree with intuitive engineering judgment;

(2) There should be no multicollinearity among the final selected independent variables;

(3) The model with the smallest number of independent variables, minimum $\|\delta\|$, and highest R^2 value was selected.

3.2 Generalized linear modeling (GLM)

This type of modeling was used successfully in accidents prediction all over the world. Semeida (2011) used this technique and concluded important results in this context. In the present research, this model was used but in different form to fit the travel demand data. The mathematical form included all independent variables separately. This was performed by taking the natural logarithm of all these variables to produce a power relationship between trip frequency and each of the former items. It allowed to study the effect of all independent variables on yearly personal trips separately. A normal distribution with a log link function was chosen to model these data. This form takes the following shape in Eq. (2) (Varagouli et al. 2005):

$$\mathrm{Tr} = \exp(\beta_0) \times X_1^{\beta_1} \times \cdots \times X_n^{\beta_n} \tag{2}$$

This form satisfied two main conditions:

(1) This model must yield logical results (non negative). Also, at $X_i = 0$ the number of trip must be zero.

(2) The logarithmic link function that can linearize this form for the purpose of coefficient estimation should exist.

This model form was executed using the generalized linear model procedure PROC GENMOD in the SAS statistical software (SAS 2003). SAS applied the maximum log-likelihood technique to estimate the regression coefficients, standard errors, Wald Chi-squared statistics and p-values. In addition, the model with minimum $\|\delta\|$ and highest R^2 value was selected.

4 Results and discussion

The analysis includes six models as mentioned before.

Hence, the analysis details for each model are presented in the following sections.

4.1 Port Said as origin

4.1.1 *Work purpose*

(A) Linear regression model

There is one model that is statistically significant with Tr after stepwise regression using SSPS package. One variable is significant at the 5% significance level (95% confidence level). Thus, this model is as follow in Eq. (3).

$$Tr = 231.02 - 0.505 \times Dt \quad (3)$$

(B) GLM procedure

After the application of PROC GENMOD in SAS software, the best model is as follow in Eq. (4).

$$Tr = \exp(8.58) \times Dt^{-0.89} \times Popd^{0.15} \times Ad^{-0.32} \times Incd^{0.33} \times Totit^{-0.46} \times Totic^{-0.48} \quad (4)$$

4.1.2 *Education purpose*

(A) Linear regression model

The same steps are executed as the previous case. Therefore, the best model is as follow in Eq. (5).

$$Tr = 170.17 - 4.93 \times 10^{-6} \times Popd - 17.08 \times Totic \quad (5)$$

(B) GLM procedure

As earlier case, the best model is as follow in Eq. (6).

$$Tr = \exp(6.54) \times Dr^{-0.66} \times Popd^{-0.1} \times Incd^{0.34} \times Totic^{-0.7} \quad (6)$$

4.1.3 *Other purpose*

(A) Linear regression model

After stepwise regression, there was no model which ensured the minimum statistical requirements. In other word, p-value of all independent variables was more than 0.05 (95% confidence level was not verified). Also, R^2 for all models never exceeded 0.5. Therefore, no model was adopted.

(B) GLM procedure

After the application of PROC GENMOD in SAS software, the best model is as follow in Eq. (7).

$$Tr = \exp(11.14) \times Dr^{-0.7} \times Ad^{-0.6} \times Totic^{-1.04} \quad (7)$$

4.1.4 *Discussion*

Table 5 presents the parameter estimates, R^2, and $\| \delta \|$ of all models in section 4.1.

(A) Work purpose

By investigation of Tab. 5, it is found that GLM procedure gives better and more confidence model than linear model as having better R^2, and lower $| \delta |$.

The best model in Eq. (4) indicates the following conclusions:

- Tr increases with the decrease of Dt, Totit, and Totic. This result is rational;
- The positive sign of the coefficient for Popd and Incd means that the Tr increases with the increase of both. The person resident in Port Said prefers to return home directly after finishing his work avoiding the congestion and excessive prices in destination governorate. So, this generates more yearly personal trips, which is consistent with logic;
- The negative sign of the coefficient for Ad means that the Tr increases with the decrease of Ad. The lower destination area leads to less urban distance to reach work place. Therefore, the number of yearly personal trips increases. This result is acceptable.

(B) Education purpose

By investigation of Tab. 5, as the work purpose, GLM procedure produces the best model.

The best model in Eq. (6) shows the following inferences:

- Tr increases with the decreases of Dr and Totic, this result is acceptable and logical;
- The positive sign of the coefficient for Incd means that the Tr increases with the increase of Incd. The collegiate from Port Said prefers to return home directly after finishing his lectures avoiding excessive prices in destination governorate. So, this generates more yearly personal trips which is consistent with logic;
- The negative sign of the coefficient for Popd means that the Tr decreases with the increase of Popd. The collegiate from Port Said prefers to reduce his number of educational trips avoiding the congestion in transportations. This result is reasonable.

(C) Other purpose

By investigation of the best model in Eq. (7), it is found that Tr increases with the decreases of Dr, Ad, and Totic. These results are acceptable and cogent.

Tab. 5 Model estimation results (Port Said as origin)

Work purpose

Variable	Linear model Coeff.	t-stat.	p-value	GLM Coeff.	Chi-square	p-value
Constant	231.020	6.64	0.0001	8.580	1830	0.0001
Dt	−0.505	3.70	0.0060	−0.887	4455	0.0001
Dr						
Popd				0.154	851	0.0001
Ad				−0.319	2998	0.0001
Dind						
Incd				0.331	226	0.0001
Totit				−0.464	893	0.0001
Totic				−0.484	760	0.0001
R^2		0.630			0.987	
‖ δ ‖		0.305			7.53×10^{-2}	

Education purpose

Variable	Linear model Coeff.	t-stat.	p-value	GLM Coeff.	Chi-square	p-value
Constant	170.17	10.48	0.0001	6.536	256	0.0001
Dt						
Dr				−0.658	2183	0.0001
Popd	-4.93×10^{-6}	3.16	0.0250	−0.104	182	0.0001
Ad						
Dind						
Incd				0.336	94	0.0001
Totit						
Totic	−17.08	5.41	0.0030	−0.702	2290	0.0001
R^2		0.907			0.990	
‖ δ ‖		0.132			5.15×10^{-2}	

Other purpose

Variable	Linear model Coeff.	t-stat.	p-value	GLM Coeff.	Chi-square	p-value
Constant				11.143	253.00	0.0001
Dt						
Dr				−0.696	12.08	0.0005
Popd						
Ad				−0.601	50.37	0.0001
Dind						
Incd						
Totit						
Totic				−1.040	37.80	0.0001
R^2					0.979	
‖ δ ‖					9.16×10^{-2}	

4.2 Port Said as destination

4.2.1 *Work purpose*

(A) Linear regression model

There is one model that is statistically significant with Tr after stepwise regression using SSPS package. This model is as follow in Eq. (8).

$$Tr = 266.36 - 0.937 \times Dt \quad (8)$$

(B) GLM procedure

As earlier case, the best model is as follow in Eq. (9).

$$Tr = \exp(73.29) \times Dt^{-1.58} \times Popo^{0.87} \times Ao^{-0.06} \times Inco^{-4.9} \times Totit^{-7.03} \quad (9)$$

It is noticed that ‖ δ ‖ is accurate to only one decimal digit. So, it is necessary to derive more accurate models. For attainment the dataset in Tab. 4 is separated into two groups of data according to rural distances as follows (Varagouli et al. 2005).

First group with Dr less than 150 km: 2→1, 11→1, 14→1, 15→1.

Second group with Dr more than 150 km: 4→1, 5→1, 6→1, 12→1, 13→1, 18→1.

Therefore, the following two models are formulated in Eqs. (10) and (11).

$$Tr = \exp(45.62) \times Dr^{-0.6} \times Dino^{-0.73} \times Inco^{-3.33} \quad (10)$$

$$Tr = \exp(-7.04) \times Popo^{-0.16} \times Ao^{0.17} \times Inco^{1.48} \times Totit^{-0.84} \quad (11)$$

4.2.2 *Education purpose*

(A) Linear regression model

The best model is as follow in Eq. (12).

$$Tr = 114.75 - 0.34 \times Dt \quad (12)$$

(B) GLM procedure

The best model is as follow in Eq. (13).

$$Tr = \exp(6.86) \times Dt^{-0.65} \times Popo^{0.3} \times Ao^{-0.32} \times Inco^{0.34} \times Totit^{-1.25} \times Totic^{-0.16} \quad (13)$$

As the work model, the dataset in Tab. 4 is separated into two groups of data according to rural distances for obtaining better models.

First group with Dr less than 150 km: 2→1, 11→1, 14→1, 15→1.

Second group with Dr more than 150 km: 3→1, 5→1, 8→1, 12→1, 18→1.

Therefore, the following two models are formulated in Eqs. (14) and (15).

$$Tr = \exp(9.2) \times Dt^{-0.57} \times Dino^{0.23} \times Totit^{-1.02} \quad (14)$$

$$Tr = \exp(-50.09) \times Dt^{-2.84} \times Popo^{1.27} \times Ao^{0.96} \times Inco^{4.19} \quad (15)$$

4.2.3 Other purpose

(A) Linear regression model

After stepwise regression, there was no model which ensured the minimum statistical requirements. Therefore, no model was adopted.

(B) GLM procedure

The best model is indicated in Eq. (16).

$$Tr = \exp(7.61) \times Dr^{-0.39} \times Popo^{-0.12} \times Ao^{-0.18} \times Totic^{-0.38} \quad (16)$$

As the previous two cases, the dataset in Tab. 4 is separated into three groups of data according to rural distances for obtaining better models.

First group with Dr less than 100 km: 2→1, 11→1, 15→1.

Second group with Dr between 100 km and 200 km: 3→1, 14→1, 16→1.

Third group with Dr more than 200 km: 4→1, 5→1, 6→1, 8→1, 9→1, 12→1, 24→1.

Therefore, the following three models are formulated in Eqs. (17), (18) and (19).

$$Tr = \exp(8) \times Dr^{-1.76} \times Dino^{0.27} \quad (17)$$

$$Tr = \exp(-9) \times Inco^{1.07} \quad (18)$$

$$Tr = \exp(-54.97) \times Popo^{0.3} \times Ao^{0.39} \times Inco^{5.06} \times Totic^{-1.4} \quad (19)$$

4.2.4 Discussion

Table 6 presents the parameter estimates, R^2, and $\| \delta \|$ of all models in section 4.2.

(A) Work purpose

By investigation of Tab. 6, it is found that GLM procedure gives better and more confidence model than linear model as having better R^2, and lower $\| \delta \|$. The best models in Eqs. (10), (11) indicate the following conclusions:

- For Group 1 model (Eq. (10)), Tr increases with the decrease of Dr. The negative sign of the coefficient for Dino and Inco means that the Tr decreases with the increase of both variables. The person resident out of Port Said prefers to reduce his number of work trips and stays in Port Said (destination) longer time avoiding the congestion and high prices in his governorate. These results are reasonable and logical.
- For Group 2 model (Eq. (11)), Tr increases with the decrease of Totit. The negative sign of the coefficient for Popo means that the Tr decreases with the increase of this variable. The person resident out of Port Said prefers to reduce his number of work trips and stays in Port Said (destination) longer time avoiding the congestion in his governorate. In addition, the positive sign of the coefficient for Ao and Inco means that the Tr increases with the increase of both variables. These results are acceptable.

(B) Education purpose

By investigation of Tab. 6, as the work purpose, GLM procedure produces the best model. The best models in Eqs. (14), (15) show the following inferences:

- For Group 1 model (Eq. (14)), Tr increases with the decrease of Dt and Totit. In addition, Tr increases with the increase of Dino.
- For Group 2 model (Eq. (15)), Tr increases with the decrease of Dt. In addition, Tr increases with the increase of Popo, Ao, and Inco.

All the previous results are cogent and acceptable.

(C) Other purpose

By investigation of the best models in Eqs. (17)-(19), it is found that:

- For Group 1 model (Eq. (17)), Tr increases with the decrease of Dr and with the increase of Dino.
- For Group 2 model (Eq. (18)), Tr increases with the increase of Inco.
- For Group 3 model (Eq. (19)), Tr increases with the increase of Popo, Ao, and Inco. In addition, Tr increases with the decrease of Totic.

Finally, all the previous results are rational and consistent with logic.

5 Conclusions

The current paper presents identification and estimation of the main variables that affect the travel demand in low population areas. The research concerns the case of Port Said Governorate, North East Egypt. The most important conclusions of this paper are:

- The GLM models give better and more confident results than regression models in terms of predicting Tr for the all six cases in research.

Tab. 6 Model estimation results (Port Said as destination)

Work purpose

Variable	Linear model			GLM			GLM (Group 1)			GLM (Group 2)		
	Coeff.	*t*-stat.	*p*-value	Coeff.	Chi-square	*p*-value	Coeff.	Chi-square	*p*-value	Coeff.	Chi-square	*p*-value
Constant	266.400	7.09	0.0001	73.290	6391	0.0001	45.62	2643	0.0001	−7.040	60.11	0.0001
Dt	−0.937	4.97	0.0010	−1.580	4335	0.0001						
Dr							−0.60	3123	0.0001			
Popo				0.875	1800	0.0001				−0.160	16.52	0.0001
Ao				−0.060	25.5	0.0001				0.171	27.43	0.0001
Dino							−0.73	1854	0.0001			
Inco				−4.900	4785	0.0001	−3.33	1843	0.0001	1.475	178.50	0.0001
Totit				−7.030	3426	0.0001				−0.840	18.21	0.0001
Totic												
R^2		0.755			0.966			1.000			1.000	
$\lvert \delta \rvert$		0.282			0.171			2×10^{-4}			5×10^{-3}	

Education purpose

Variable	Linear model			GLM			GLM (Group 1)			GLM (Group 2)		
	Coeff.	*t*-stat.	*p*-value	Coeff.	Chi-square	*p*-value	Coeff.	Chi-square	*p*-value	Coeff.	Chi-square	*p*-value
Constant	114.75	6.15	0.0001	6.865	41.74	0.0001	9.20	3512.0	0.0001	−50.100	235.9	0.0001
Dt	−0.34	3.67	0.0080	−0.650	302.80	0.0001	−0.57	455.5	0.0001	−2.840	210.7	0.0001
Dr												
Popo				0.298	153.80	0.0001				1.270	230.4	0.0001
Ao				−0.320	183.70	0.0001				0.960	179.6	0.0001
Dino							0.23	1035.0	0.0001			
Inco				0.344	12.70	0.0004				4.189	425.7	0.0001
Totit				−1.250	388.40	0.0001	−1.02	428.0	0.0001			
Totic				−0.160	11.36	0.0008						
R^2		0.658			0.960			1.000			1.000	
$\lvert \delta \rvert$		0.251			0.118			7×10^{-5}			7×10^{-5}	

Other purpose

Variable	Linear model			GLM			GLM (Group 1)			GLM (Group 2)		
	Coeff.	*t*-stat.	*p*-value	Coeff.	Chi-square	*p*-value	Coeff.	Chi-square	*p*-value	Coeff.	Chi-square	*p*-value
Constant	7.61	80.29	0.0001	7.990	22.82	0.0001	−8.950	6.95	0.0084	−55.000	16.58	0.0001
Dt												
Dr	−0.39	15.19	0.0001	−1.760	19.65	0.0001						
Popo	−0.12	5.38	0.0203							0.298	17.21	0.0001
Ao	−0.18	9.28	0.0023							0.394	18.21	0.0001
Dino				0.275	22.08	0.0001						
Inco							1.074	10.12	0.0015	5.063	17.27	0.0001
Totit												
Totic	−0.38	9.83	0.0017							−1.400	23.84	0.0001
R^2		0.610			1.000			0.994			0.993	
$\lvert \delta \rvert$		0.358			1×10^{-4}			2.36×10^{-2}			5.6×10^{-2}	

• For all models, the negative sign of the coefficients for Dt, Dr, Totit, and Totic means that Tr permanently increases with the decrease of these variables. This conclusion is rational and consistent with logic.

• In some cases, ‖ δ ‖ values are accurate to only one decimal digit. So, it is necessary to derive more accurate models. For attainment the dataset is separated into groups of data according to rural distances until reaching more reasonable and accurate models.

• The relations between dependent and independent variables in all concluded models are reasonable, rational, and consistent with logic. In addition, the final demand models have statistics within the acceptable regions and, also, they are conceptually reasonable.

• The derived models are useful and functional for future trips forecast especially at Port Said zone. The future trips widely vary according to variation in specific independent variables as population for origin and destination zones, average personal income for origin and destination zones, and total trip cost in internal (urban) distances. Thus, the number of personal trips in future can be calculated easily by putting the main independent variables in the derived models.

• Finally, future research should be conducted to extend all aspects of this research using comprehensive field data from various governorates in order to reach more extensive modeling and analysis of travel demand. This surely improves the transportation system in Egypt. In addition, microsimulation molding procedure will be used for obtaining more accurate models.

Acknowledgments

The author acknowledges The Egyptian Central Agency for Public Mobilization and Statics (ECAPMS), for its assistance with the acquisition of Egyptian governorates property data (census, area, and annual personal income).

References

Alfa, A. S., 1986. A review of models for the temporal distribution of peak traffic demand. Transportation Research Part B: Methodo logical, 20(6): 491-499.

Anderson, M. D., Abdullah, Y. M., Gholston, S. E., et al., 2006. Development of a methodology to predict through trip rates for small communities. Journal of Urban Planning and Development, 132(2): 112-114.

ECAPMS, 2010. Egyptian central agency for public mobilization and statics. http://www.capmas.gov.eg/database.aspx.

Han, Y., Stone, J. R., 2008. Synthesized through-trip model for small and medium urban areas. Transportation Research Record, 2077: 148-155.

Hutchinson, B. G., 1993. Analyses of Canadian air travel demands. Journal of Transportation Engineering, 119(2): 301-316.

Koppelman, F. S., Hirsh, M., 1986. Intercity passenger decision making: conceptual structure and data implications. Transportation Research Record, 1085: 70-75.

Oppenheim, N., 1995. Urban travel demand modeling. John Wiley & Sons, New York.

Qian, Z., Han, Y., Stone, J. R., 2012. Forecasting external trips in small and medium cities based on local economic context. Procedia-Social and Behavioral Sciences, 43: 284-293.

SAS Institute Inc., 2003. Version 9 of the SAS system for Windows, SAS Institute Inc., Cary, NC, USA.

Semeida, A. M., 2011. Analysis and evaluation of road safety in Egypt using conventional and non-conventional modeling techniques. PhD thesis, Port Said University, Port Said.

Semeida, A. M., 2013. Impact of highway geometry and posted speed on operating speed at multilane highways in Egypt. Journal of Advanced Research, 4(6): 515-523.

Semeida, A. M., 2014. Application of artificial neural networks for operating speed prediction at horizontal curves: a case study in Egypt. Journal of Modern Transportation, 22(1): 20-29.

Simpson, A. L., Rauhut, J. B., Jordahl, P. R., et al., 1994. Sensitivity analyses for selected pavement distresses. SHRP Report No. SHRP-P-393.

Stopher, P. R., McDonald, K. G., 1983. Trip generation by cross-classification: an alternative methodology. Transportation Research Record, 944: 84-91.

Varagouli, E. G., Simos, T. E., Xeidakis, G. S., 2005. Fitting a multiple regression line to travel demand forecasting: the case of the prefecture of Xanthi, Northern Greece. Mathematical and Computer Modeuing, 42(7/8): 817-836.