İSTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

MEASURING THE INCREASE IN URBAN MOTORIZED PASSENGER MOBILITY IN THE CASE OF DECREASE IN TRAVEL TIME

Phd THESIS

Enver Cenan İNCE

Department of Urban and Regional Planning

Urban and Regional Planning Program

DECEMBER 2019

İSTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

MEASURING THE INCREASE IN URBAN MOTORIZED PASSENGER MOBILITY IN THE CASE OF DECREASE IN TRAVEL TIME

Phd THESIS

Enver Cenan İNCE
(502152814)

Department of Urban and Regional Planning

Urban and Regional Planning Program

Thesis Advisor: Prof. Dr. Hüseyin Murat ÇELİK

DECEMBER 2019

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

MOTORLU TAŞITLAR ARACILIĞIYLA GERÇEKLEŞTİRİLEN KENTSEL HAREKETLİLİK ARTIŞININ ÖLÇÜLMESİ

DOKTORA TEZİ

Enver Cenan İNCE
(502152814)

Şehir ve Bölge Planlaması Anabilim Dalı

Şehir ve Bölge Planlama Programı

Tez Danışmanı: Prof. Dr. Hüseyin Murat ÇELİK

ARALIK 2019

Enver Cenan İNCE, a Ph.D. student of İTU Graduate School of Science Engineering and Technology student ID 502152814, successfully defended the thesis entitled "MEASURING THE INCREASE IN URBAN MOTORIZED PASSENGER MOBILITY IN THE CASE OF DECREASE IN TRAVEL TIME", which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor : **Prof. Dr. Hüseyin Murat ÇELİK** ....
İstanbul Technical University

Jury Members : **Prof. Dr. Hilmi Berk ÇELİKOĞLU** ....
İstanbul Technical University

Prof. Dr. Aliye Ahu AKGÜN ....
İstanbul Technical University

Assoc. Prof. Dr. İbrahim DEMİR ....
Yıldız Technical University

Assoc. Prof. Dr. Kevser İ. ÜSTÜNDAĞ ....
Mimar Sinan Fine Arts University

Date of Submission : 14.11.2019
Date of Defense : 03.12.2019

To my father,

FOREWORD

Firstly, I want to give many thanks to my advisor Prof. Dr. Hüseyin Murat ÇELİK for his vital helps and his contribution to this thesis. In addition, I want to give many thanks to my family, namely Ozan İNCE (my brother), Sevim İNCE (my mother), and Remzi İNCE (my father) for their patience and contributions throughout my life involving the Phd process. Lastly, my dear father, who was a retired Turkish teacher, has unfortunately died this year and his efforts on my painful education process was excellent. I dedicate this thesis to him and I will not forget him till I die.

December 2019

Enver Cenan İNCE
(City Planner)

TABLE OF CONTENTS

Page

ABBREVIATIONS

Auto	: Automobile
CBD	: Central Business District
Coef.	: Coefficient
Cov	: Covariance
FFT	: Free Flow Travel Time
GMM	: Generalized Method of Moments
GSEM	: Generalized Simultaneous Equations Model
HBS	: Home Based School trip
HBW	: Home Based Work trip
hh	: household
hhsize	: household size
IV	: Instrumental Variable
IVPRM	: Instrumental Variable Poisson Regression Model
IVZTPRM	: Instrumental Variable Zero Truncated Poisson Regression Model
Km	: Kilometer
L	: Linear function
LR	: Likelihood Ratio
LRM	: Linear Regression Model
ME	: Marginal Elasticity
motorfft	: motorized free flow travel time
ML	: Maximum Likelihood
MSE	: Mean Squared Error
NBRM	: Negative Binomial Regression Model
O/D	: Origin/Destination
OLS	: Ordinary Least Squares
PRM	: Poisson Regression Model
PSU	: Primary Sampling Units
Root MSE	: Root of the Mean Squared Error
SEM	: Simultaneous Equations Model
SSM	: Sample Selection Model
SSPRM	: Sample Selection Poisson Regression Model
Std. Err.	: Standard Error
TAZ	: Traffic Analysis Zone
TL	: Turkish Lira
TSTT	: Total System Travel Time
3SLS	: Three Stages Least Squares Regression Model
2SLS	: Two Stages Least Squares Regression Model
VMT	: Vehicle Miles Travelled
ZTM	: Zero Truncated Model

SYMBOLS

chi2(1) : chi-square value for degrees of freedom 1
_cons : constant value of the model
Cov (x,y) : The function standing for finding the covariance between x and y
d.f. : degree of freedom
_d : dummy variable of the asserted binary case
∂y∂x : partial derivative function of y with respect to x.
e : exponential funtion
E[x] : The function standing for finding the expected value of the random variable x
E[y|x] : The function standing for finding the expected value of the random variable y, given that x
ε : Disturbance Term
exp (x) : exponential function of x (e^x)
Γ(∫) : Gamma integral function for ∫.
$\int f(x)dx$: Integral function of x (f(x)) with respect to x
λ : Lambda (represents the expected value such as mean)
H_0 : Null hypothesis
L : Linear
Ln : Natural Logarithm
μ : Model Residual
ρ : probit
Prob (x) : Probability function of the random variable x
r (x,y) : coefficient of correlation between the variables x and y
R-squared : coefficient of determination
S-W test : Shapiro - Wilk test
Φ (.) : Probit function
$\sum_{i=1}^{n}(Xi)$: Sigma function summing each observation i (from 1 to n) for X
θ : Rate parameter
$\bar{x}$: mean value of x
Var : Variance
Var (x_i) : The function standing for finding variance of the random variable x_i

LIST OF TABLES

Page

LIST OF FIGURES

Page

MESURING THE INCREASE IN URBAN MOTORIZED PASSENGER MOBILITY IN THE CASE OF DECREASE IN TRAVEL TIME

SUMMARY

Unlike any previous researches of urban passenger mobility demand reference to the travel demand behaviors of the individuals, this thesis firstly proposes a measurement focusing specifically on the interrelation between the travel times of the individuals and the number of motorized trips they exhibit in a day. It is the first time in the literature that the related measurement focuses on the additional number of daily motorized trips -instead of focusing on measuring vehicle miles traveled- as a result of decrease in daily motorized travel time.

This research has been developed with pure individuals based cross-section data, which is gathered via 2006 household Origion- Destination (O/D) survey of İstanbul metropolitan area. The survey was conducted by the Department of the Transportation of the Metropolitan Municipality of İstanbul. In this survey, a total of 264,000 passengers -belonging to 72,000 households- were interviewed face to face, and a total of 356,000 daily trips were recorded between 451 Traffic Analysis Zones (TAZs) defining 203,401 distinct movements. In addition, the response rate of this field survey is 80 %. That type of pure individuals based data collection process would have made sampling errors dramatically decrease owing to the direct usage of individuals based data (without any sampling replacement) instead of the household based ones. In this sense, ***respondent's fatigue*** would have been eliminated, which would refer to the prevailing problem in travel demand models in literature. That is to say, any selected respondent in the household, who would most probably be the household head, would produce biased data due to the case called respondent's fatigue in that the selected respondent would not know or remember some amounts of trips of all members of the related hosehold during the process of long-lasting travel surveys. Hereby, for large samples, the data would be dramatically biased, so this study would be one of the rare studies in literature, taking this into consideration. Secondly, the new way of measuring induced urban passenger mobility demand has been proposed via grasping the coefficient of the marginal effect between the "*daily number of motorized trips*" and "*daily motorized travel time*" of each passenger. Herein, the variable called total daily motorized travel time has been taken as the major component of the generalized cost function as an explicit proxy variable. In an other words, unlike most researches on urban passenger mobility demand, this thesis does not introduce exogeneous measures of accessibility or generalized cost of travel, but instead uses the survey data on reported daily travel times to approximate each individual's generalized cost of travel.

Reference to the methodological framework of this research, three technical obstacles have been encountered, namely the non-linearity of number of daily motorized trips per passenger (as a count variable), excess amounts of zero observations in the number of daily motorized trips per passenger, and endogeneity of motorized travel time of each passenger. If non of these technical obstacles did not

occur, classical Linear Regression Model (LRM) would be implemented on the main dependent variable called daily number of trips exhibited by each motorized passenger. On the other hand, in the context of these three technical obstacles, there have been asserted a number of models as benchmarks, namely Poisson Regression Model (PRM), Negative Binomial Regression Model (NBRM), Sample Selection Model (SSM), Sample Selection Poisson Regression Model (SSPRM), Instrumental Variable Poisson Regression Model (IVPRM), Instrumental Variable Zero Truncated Poisson Regression Model (IVZTPRM), Generalized Simultaneous Equations Model (GSEM), and Two Stages Least Squares Model (2SLS). Each of these model structures has been proposed as a benchmark for others reference to the related three technical obstacles so as to find the most convenient model structure for such a research. In this sense, it is also to be mentioned that there has been no clear explanation in literature for a system of equation in that one equation is linear, while the second one is non-linear in the case that one of the dependent variables is endogeneous. Thus, the methodological effort of the thesis refers to exploring possible ways to deal with such a case.

The first technical obstacle with refers to the count nature of the main dependent variable of the research, called daily number of motorized trips per passenger, necessitates to take count models into account. In this context, PRM and NBRM come into considerations so as to model this dependent variable. Furthermore, for the second technical obstacle -related to the excess amounts of zero observations in the daily number of the motorized trips of a passenger- SSM, SSPRM and Zero Truncated Model (ZTM) structures come into considerations. Besides, for the third technical obstacle –related to the endogeneity of motorized travel time of each passenger- IVPRM, IVZTPRM, GSEM, and 2SLS model structures come into prominence. Selection among all these models is directly related to the three technical obstacles with reference to two main variables. In other words, the most convenient model structure, modelling the inter-relationship between daily number of trips and daily travel time of passengers, is expected to be able to cope with all these three technical obstacles together.

On one hand, all these models are able to cope with the count nature of the main dependent variable, called daily number of motorized trips of a passenger. On the other hand, excess zero observations in the counts of the daily trips of a passenger are not able to be tackled with by PRM, NBRM, and IVPRM structures. Furthermore, with reference to the endogeneity of daily travel time of passengers, none of the single index models, namely PRM, NBRM, and SSPRM is able to produce consistent and efficient estimations. At this juncture, multi-equations model structures such as IVPRM, IVZTPRM, GSEM, and 2SLS are preferred. In the light of these eliminations ,according to the related technical obstacles, three model structures, namely IVZTPRM, GSEM, and 2SLS have been remained. But, IVZTPRM has also been eliminated among these, since the assumption of equidispersion for the dependent variable of the Poisson Regression Model has been failed. Herein, it is meant that the PRM structure statistically assumes that mean and variance of the dependent variable (number of daily motorized trips per passenger) are equal to each other. On the other hand, according to the equidispersion test for the daily counts of motorized trips per passenger, the mean and variance of this variable are not equal to each other, which have been detailly exhibited in the chapter of the thesis called Model Results. Hence, only GSEM and 2SLS have been remained as the optional convenient models for the research of thesis. Non-linear structure of elasticity estimate of these models might further allow someone to estimate the spatial variation of generative impact of induced urban passenger mobility and to integrate it into the trip generation models since it is possible to account for the individual characteristics in the estimation of elasticities as long as researchers have disaggregated spatial data.

According to the results of GSEM, a ten minute decrease in average motorized travel time (26 % decrease in travel time) makes daily number of motorized trips increase by 1.2 % per passenger, which refers to 261,250 more motorized daily trips - with refers to the 174,167 more motorized vehicles in the daily traffic- in addition to the 21 million total daily trips (14 million motorized vehicles in a day) in İstanbul. On the other hand, with refers to the results of 2SLS, the same amount of decrease in average motorized travel time makes daily number of motorized trips increase by 11.4 % per passenger, which refers to 1.19 million more motorized daily trips in total – with reference to 793,333 more motorized vehicles in the daily traffic- in addition to the 21 million daily motorized trips (14 million daily motorized vehicles) in İstanbul. Herein, 2SLS gives much higher marginal effect estimation when compared to the one of GSEM. Such a difference between the related marginal elasticity estimations of these two optional models would be caused by the difference in their model structures, which would have made the travel time sensitivities of passengers decrease significantly in the case of GSEM when compared to the one of 2SLS. The selection among these two model structures would be a kind of ***state of art*** for researchers, which requires more similar future studies.

The potential multiplication effect of all these empirical findings of the thesis are able to be explained by three frameworks. The first is that these marginal elasticity estimations would be able to be integrated into the classical four stages travel demand models, namely trip generation stage, trip distribution stage, modal split stage, and network assignment. In this context, the number of daily motorized trips would be the outcome of the first stage called trip generation models, while the daily motorized travel time would be the outcome of the last stage called network assignment, since the network assignment stage systematically produce total system travel times (TSTT) by its nature. Hereby, the static nature of the classical travel demand models would be transformed into more iterative framework.

Secondly, subsequent to detection of the prominent factors generating daily motorized trips in the case of İstanbul according to the related model results, these prominent factors are able to be benefited in formulating any travel demand management policy in a similar developing country, which would partly be different from the developed ones.

Lastly, the empirical findings of this thesis would strengthen the basis of cost & benefit analysis of any transportation project in urban scale, since the travel time refers to a kind of proxy in measuring travel cost per passenger.

MOTORLU TAŞITLAR ARACILIĞIYLA GERÇEKLEŞTİRİLEN KENTSEL HAREKETLİLİK ARTIŞININ ÖLÇÜLMESİ

ÖZET

Ulaşım talep yönetimi politikalarına yönelik deneysel tartışmalar kapsamında yapılan kışkırtılmış ulaşım talep modellemesine yönelik yazın çalışmalarının hemen hemen hepsinde, "motorlu taşıtlarla yapılan ortalama yolculuk süresi kısaldıkça, yapılan toplam yolculuk mesafesi artar" hipotezi test edilegelmiştir. Öte yandan, bu tür araştırmaların hiçbirinde, kışkırtılan yolculuk taleplerinin miktarı ve dağılımı ile ilgili bir ele alış benimsenmemiştir. Bir diğer ifadeyle, önceki çalışmalardan farklı olarak, "yolculuk süresindeki tekil değişimin, toplam yapılan yolculuk mesafelerine etkisi" odaklı bir yaklaşımdan ziyade, "bireyler bazında gerçekleştirilen günlük yolculuk süreleri ile günlük yolculuk sayıları arasındaki etkileşimin incelenmesi" gereksinimi ortaya çıkmış olup, söz konusu gereksinim ekseninde oluşan motivasyonla bu doktora tez çalışması üretilmiştir.

İlgili araştırma sorusu çerçevesinde kullanılan veriler, İstanbul metropolitan alanı sınırları içinde 2006 yılında hane halkları bazında yapılan yolculuk anketleri ve saha araştırmaları üzerinden elde edilmiştir. Söz konusu çalışma, İstanbul Büyükşehir Belediyesi'ne bağlı Ulaşım Departmanı'nca yürütülmüştür. Çalışmada kullanılan örneklem büyüklüğü 90.000 hane, örneklem oranı ise % 3 şeklindedir. Bu şekilde 450 adet trafik analiz bölgesi kapsamında 72.000 adet hane halkına ve toplamda 264.000 bireye yüz yüze görüşmeler üzerinden yarı yapılandırılmış derinlemesine anket görüşmeleri uygulanmış olup, toplamda 356.000 günlük yolculuk, yolcukların başlangıç ve bitiş bilgilerinin de dâhil olduğu detaylarıyla kaydedilmiştir. Söz konusu saha araştırması anketlerine yönelik cevaplanma oranı ise % 80 'dir.

Bireyler bazında toplanan bu tür toplulaştırılmamış bir veri seti ile çalışma olanağının sağlanması, denek yorgunlukları kaynaklı örneklem hatalarının ve yanıltıcı verilerin en aza indirilmesinde hayati önem taşımaktadır. Bu noktada, hane halkı düzeyinde yapılan araştırmalara kıyasla çok daha etkili ve gerçeği yansıtan verilerin oluşturulması olanağı artmaktadır. Şöyle ki, hane halkı düzeyinde yapılan araştırmalarda her bir hane içerisinden seçilen denek, hanenin diğer üyelerinin gerçekleştirdikleri günlük yolculukları hatırlamayabilir ve/veya tamamlanması uzun süren anketler boyunca yorulmalarını takiben yanıltıcı bilgiler verebilir. Bu bağlamda, tüm İstanbul için üretilen büyük bir veri seti ile çalışılması durumunda meydana gelebilecek toplulaştırılmış ölçüm ve gözlem hatalarının oldukça büyük değerlere işaret etmesi kaçınılmaz olacaktır. Ulaşım talep modelleri yazınında yaygın olarak kullanılan hane halkı bazında üretilen verilerin aksine, bu tez çalışması kapsamında bireyler bazında üretilen veriler üzerinden çalışılmasının sebebi budur. Ek olarak, tez çalışmasına konu olan saha çalışması kapsamında seçilen örneklem birimleri için kesinlikle ikame yapılmaması ilkesi benimsenmiştir.

Motorlu taşıtlarla gerçekleştirilen kentsel hareketliliğe yönelik kışkırtılmış talebin ölçülmesine yönelik olarak ortaya konulan araştırma çerçevesinde iki temel değişken

tanımlanmıştır. Bunlar: her bir yolcu bazındaki günlük toplam motorlu yolculuk sayısı ve her bir yolcu bazındaki günlük toplam yolculuk süresidir. Bu noktada, yolculuk süresi değişkeni, her bir yolcuya yönelik genelleştirilmiş yolculuk maliyeti fonksiyonunun ana bileşeni olarak ele alınmıştır. Bir diğer ifadeyle, yolculuk sürelerindeki değişimin, günlük motorlu taşıt yolculukları üzerindeki tekil etkilerinin hesaplanması hedeflenmiştir. Böylesi bir araştırma çerçevesinde ise, ekonometrik model yaklaşımları bağlamında üç temel problemle karşılaşılmıştır. Bunlar: günlük motorlu yolculuk sayısı verisinin normal dağılmayan bir sayım verisi olması, günlük yolculuk sayıları verisinin çok sayıda sıfır gözlemi içermesi ve günlük yolculuk süresi değişkeninin içsel bir değişken olma özelliği göstermesi şeklindedir. Bu bağlamda, sırasıyla Poisson Bağlanım Modeli, Negatif Binom Dağılımlı Bağlanım Modeli, Örneklem Seçimli Bağlanım Modeli, Örneklem Seçimli Poisson Bağlanım Modeli, Araç Değişkenli Poisson Bağlanım Modeli, Araç Değişkenli Sıfır Gözlemlerden Arındırılmış Poisson Bağlanım Modeli, Genelleştirilmiş Eşanlı Denklem Sistemleri Modeli ve İki Aşamalı En Küçük Kareler Bağlanımı Modeli ortaya konulmuştur.

Bireyler bazındaki günlük yolculuk sayısı verisinin normal dağılmayan bir sayım verisi olması, ekonometri yazınında sayım verisi modelleri olarak öne çıkan Poisson Bağlanım Modeli ve Negatif Binom Dağılımlı Bağlanım Modeli yapılarını gündeme getirmiştir. Günlük yolculuk sayıları verisi içerisindeki yaygın sıfır gözlemler dolayısıyla ise Örneklem Seçimli Bağlanım Modeli, Örneklem Seçimli Poisson Bağlanım Modeli ve Sıfır Gözlemlerden Arındırılmış Model yapıları öne çıkmıştır. Yolculuk süresi verisinin, yolculuk sayıları modellerine yönelik içsel bir veri yapısı sergiliyor oluşu ise, Araç Değişkenli Poisson Bağlanım Modeli, Araç Değişkenli Sıfır Gözlemlerden Arındırılmış Poisson Bağlanım Modeli, Genelleştirilmiş Eşanlı Denklem Sistemleri Modeli ve İki Aşamalı En Küçük Kareler Bağlanımı Modeli yapılarını öne çıkarmıştır.

Araştırma çerçevesinde bahsi geçen iki temel değişken arasındaki tekil etkinin tahmin edilmesine yönelik ortaya konulan bir model yapısının, söz konusu üç problemle de aynı anda başedebilecek tutarlı ve anlamlı sonuçlar üretmesi beklenmektedir. Bu noktada, bir yandan, ortaya konulan tüm model yapılarının, normal dağılmayan bir sayım verisi olan günlük yolculuk sayılarının modellenmesinde kullanabileceği açıkça söylenebilir. Öte yandan, günlük yolculuk sayıları verisi içerisindeki sıfır gözlemler sorunsalıyla (ekonometrik model yaklaşımı açısından) baş edebilmesi mümkün olmayan modeller grubunda ise Poisson Bağlanım Modeli, Negatif Binom Dağılımlı Bağlanım Modeli ve Araç Değişkenli Poisson Bağlanım Modeli yer almaktadır. Ek olarak, tekli indeks modeller grubunda olan Poisson Bağlanım Modeli, Negatif Binom Dağılımlı Bağlanım Modeli ve Örneklem Seçimli Poisson Bağlanım Modeli yapılarının hiçbiri, yolculuk süreleri değişkeninin içselliği ile baş edememektedir. Dolayısıyla, çok değişkenli model yapıları grubunda yer alan Araç Değişkenli Poisson Bağlanım Modeli, Araç Değişkenli Sıfır Gözlemlerden Arındırılmış Poisson Bağlanım Modeli, Genelleştirilmiş Eşanlı Denklem Sistemleri Modeli ve İki Aşamalı En Küçük Kareler Bağlanımı Modeli yapıları, yolculuk sürelerinin içselliğine yönelik uygun model yapıları olarak değerlendirilmiştir.

Söz konusu üç teknik problemin aynı anda çözümlenebileceği modeller tartışmasında ise, yukarıda değinilenler ışığında, Araç Değişkenli Sıfır Gözlemlerden Arındırılmış Poisson Bağlanım Modeli, Genelleştirilmiş Eşanlı Denklem Sistemleri Modeli ve İki Aşamalı En Küçük Kareler Bağlanımı Modeli yapıları öne çıkmaktadır. Ancak, Araç

Değişkenli Sıfır Gözlemlerden Arındırılmış Poisson Bağlanım Modeli de, ana bağımlı değişkene (günlük yolculuk sayıları) yönelik yapılan ortalama ve varyans değerlerinin eşit olduğuna yönelik varsayımın çürütülmesi dolayısıyla elenmiştir. Sonuç olarak, geriye kalan Genelleştirilmiş Eşanlı Denklem Sistemleri Modeli ve İki Aşamalı En Küçük Kareler Bağlanımı Modeli, söz konusu araştırma çerçevesine en uygun iki model yapısı olarak öne çıkmışlardır.

Genelleştirilmiş Eşanlı Denklem Sistemleri Modeli sonuçlarına göre, günlük yolculuk sürelerindeki 10 dakikalık bir azalma, tüm İstanbul için günlük mevcut toplam 21 milyon motorlu taşıt yolculuklarına ek olarak günlük 261 bin 250 ilave motorlu taşıt yolculuğun, ve her bir yolcu için % 1,2 yolculuk artışına referansla, yapılmasına sebebiyet vermektedir. Bu noktada, bireyler bazındaki günlük ortalama yolculuk sürelerindeki olası 10 dakikalık bir azalmanın, günlük mevcut 14 milyon motorlu taşıt kullanımına (21 milyonluk toplam günlük yolculuğa referansla) ek olarak 174 bin 167 motorlu taşıtın günlük motorlu yolculuk trafiği sayımlarına eklenmesi sonucunu doğuracağı tahmin edilmiştir. Öte yandan, İki Aşamalı En Küçük Kareler Bağlanım Modeli sonuçlarına göre ise, yolculuk sürelerindeki aynı miktarda olan azalma (10 dakika), tüm İstanbul için günlük 1 milyon 19 bin ek (21 milyon yolculuğa ek olarak) motorlu taşıt yolculuğu (her bir yolcu için % 11,4 yolculuk artışı) yapılmasına neden olmaktadır. Söz konusu bulgu ise, günlük ilave 793 bin 333 ilave motorlu taşıtın (14 milyon günlük motorlu taşıta ek olarak) günlük motorlu taşıt trafiği sayımlarına ekleneceğine işaret etmektedir. Motorlu yolculukları gerçekleştiren yolcuların yolculuk sürelerine bağlı yolculuk yapma hassasiyetlerinin ölçüldüğü tekil etki tahminlerinin, birbirlerine opsiyonel olarak ileri sürülebilen söz konusu iki model yapıları içindeki farklılıklardan kaynaklanıyor olabilir. İlgili model yapıları arasından seçim yapılabilmesi için ise, farklı ülkelerin farklı illerine yönelik daha fazla deneysel çalışmaya ihtiyaç duyulduğu açıktır. Bu tez çalışması, böylesi bir çerçevenin başlatılması gibi bir misyonu üstlenmiştir.

Tez kapsamında ortaya konulan söz konusu deneysel bulgular ışığında, ulaşım planlaması ve yolculuk talep modelleri yazınına yönelik üç temel çarpan etkisi ileri sürülebilir. Bunlardan ilki, günlük yolculuk sürelerinin günlük yolculuk sayıları üzerindeki tekil etki tahminlerinin, klasik dört aşamalı yolculuk talep tahmini modellerine entegre edilebilmesiyle ilgilidir. Burada, klasik dört aşamalı modellerin ilk ayağı olan yolculuk üretimlerinin modellenmesi aşaması, doğrudan günlük yolculuk sayıları tahminleriyle ilişkilendirilebilir. Klasik dört aşamalı modellerin dördüncü ve son adımı olan ağ ataması aşamasının matematiksel çıktısı olan toplam sistem yolculuk süresi değişkeni ise, doğrudan günlük yolculuk süreleri verisiyle ilişkilidir. Bu bağlamda, tek yönlü ilerleyen klasik dört aşamalı yolculuk talebi tahmin modelleri, dördüncü aşama sonrasında elde edilecek olan toplam sistem yolculuk süresi (günlük yolculuk sürelerine referansla) tahminlerinin birinci aşamanın bir çıktısı olan günlük yolculuk sayıları tahminlerini tekrardan etkileyecek ve bu işlem iki değişken arasındaki optimum denge yakalanana dek bir döngü olarak devam edecektir. Böylelikle, statik ve tek yönlü ilerleyen klasik dört aşamalı yolculuk talebi tahmin modelleri, döngüsel ve çok yönlü bir yapıya evrilebilecektir.

İkinci olarak, "yolculuk yapma potansiyeli olan bir bireyin harcadığı günlük yolculuk süresindeki bir birimlik bir azalmanın, ilgili birey bazında günlük kaç adet yolculuğu tetikleyeceği" şeklinde kurgulanan bir araştırma sorusuna yönelik deneysel bulgular, ulaşım talep yönetimi politikalarının formüle edilmesi, izlenilmesi ve değerlendirilmesi süreçlerinde kullanılabilecek olmaları açısından tezin ikinci çarpan etkisi olara ifade edilebilir. Bu noktada, yolculuk sayılarını en çok etkileyen parametreler belirlenerek, söz konusu politikalar bu parametreler üzerinden geliştirilebilecek ve kentsel ölçekteki motorlu taşıtlar aracılığıyla gerçekleştirilen hareketliliğin en aza indirilmesine yönelik performans göstergeleri de yine aynı parametrelerin zaman içerisindeki değişimleri üzerinden tanımlanabilecektir. Tezin üçüncü ve son çarpan etkisi ise, söz konusu deneysel bulguların, herhangi bir ulaşım projesine yönelik fayda ve maliyet analizlerine entegre edilebilmeleri olarak ifade edilebilir. Şöyle ki, yolculuk süreleri, her bir bireyin yolculuk maliyetlerinin bir ölçümü olarak ele alınabilir ve bireyler bazında yapılacak olan zamanın para değeri araştırmalarını takiben yolculuk süreleri modellenebilir. Akabinde ise, herhangi bir ulaşım projesi/yatırımı sonrası oluşabilecek yolculuk süresi değişimleri tahmini üzerinden günlük yolculuk sayılarına olan etkiler modellenebilir. Günlük yolculuk sayıları değişimi üzerinden ise, her bir motorlu ek taşıt yolculuğunun çevresel etkileri (gaz emisyonu, hava kirliliği, vs) irdelenebilir. Böylelikle, ulaşım projelerine yönelik fayda ve maliyet analizlerine farklı boyutlar kazandırılabilecektir.

1. INTRODUCTION

After many years of continuing investment in highway network to solve traffic congestion in urban areas, transportation professionals came to the idea that the efforts to solve the traffic congestion by constructing new highways might be a futile effort, since these new highways are re-congested in a relatively short period of time, confirming the contention that "you cannot build your way out of traffic congestion" Downs (1992). According to the theory, the newly created capacities to solve an existing congestion problem stimulate the suppressed demands and cause shifts in time, route and mode of daily travels, as called "triple divergence" by Downs (1992). Emergence of the suppressed demand, accompanying with these convergences as a consequence of the constructing a new highway facility, is called as "induced demand".

Induced demand may be composed of several parts. While some part of the induced demand comes from diversion of the existing demand, some other parts are newly generated traffic. Newly generated traffics have two main forms. One is the release of suppressed demand, and the other comes from new urban development around improved facilities. While divergence and suppressed demand effects are short term immediate effects, developmental effects are generally realized in the long term.

Since traffic congestion is observed on highway network, the induced demand measurements in the literature are concentrated on the measurement of the highway distance with respect to travel time. The obtained elasticity is deemed reflecting the people's willing to travel as travel time decreases. These measurements may have several deficiencies: first of all, these measurements may reflect a partial urban equilibrium when especially done in facility based or corridor based measurement. Area-wide measurements, however, may implicate urban development effect and suffer in isolating individuals' elasticity to travel more or farther. Cervero (2002) isolated the urban development effect and showed us that an important part of the induced demand comes from urban development.

These studies have proved enough so far that people are willing to make more vehicle miles in their car trips as travel time decreases. This increase can come from either making longer trips for their existing trips by shifting the origin/destination of the trip, or making additional trips not needed or suppressed previously. That part of the research has not make a clear measurement on this issue yet. Besides, the literature is dominated by the research from the western cities, where private car is the dominating transportation mode. However, in the developing part of the world, people are more dependent on public transit and their response to change in travel time will certainly affect the number of ridership, and may result in an eventual change in cost & benefit analysis of public transit investment. In that sense, effect of induced demand not only for highway transport, but also for total trip generation gains importance, all of which reveal one of the significant contributions to literature for the case of İstanbul.

Induced demand literature is dominated by the studies from the western countries, (mainly from the United States) because of a high rate of automobile dependency for urban travels, and consequent traffic congestion in these countries. Even though the sensitivity to travel time is intuitively valid for all travel modes, the literature is mainly devoted to the car transportation, since traffic congestion may have the highest externality. However, if public transportation is dependent on rubber-tired transit, travel sensitivities may also become important leading to increased transit demand, eventual worsening traffic congestion. In any case, beyond partial urban equilibrium, it would certainly be very useful to have a system-wide elasticity to include the induced demand effect in urban transportation investment, that is constituting the baseline of the motivation of this thesis.

As Cervero stated “until trip generation techniques adequately account for latent trips, the traffic assignment step adequately captures route shifts, and dynamic feedback loops are created to account for land use shifts spurred by new road, the art and science of travel demand forecasting will fall far short of the ideal. Progress is sorely needed on this front” (Cervero, 2002:18). The intention of this research is an attempt to measure induced demand into trip generation, using total travel times spent in daily trips with their personal characteristics to suggest some progress on this front.

The supply of new transportation infrastructure, with respect to the related travel demand of the individuals, constitutes a type of strategy, which is not sustainable. Therefore, the considerations of “travel demand management policies”, asserting a kind of control mechanism to the travel demand of the individuals, have been widely taken into consideration. In this sense, most of the studies have diverted their empirical efforts in exploring the marginal effect of travel time on vehicle miles travelled (VMT) so as to discuss the minimization of the VMTs as much as possible. On the other hand, such measurements have not been able to isolate the “generative part” and “redistributive part” of the induced travel demand considerations. In other words, there exists an explicit requirement of explaining the inter-relationship between “the number of daily trips” and “ daily travel time” specifically, that has not been carried out up to now in literature. Herein, via capturing the marginal elasticity of “number of daily trips (count variable)” with respect to the “daily travel time spent by the individuals (Gaussian distributed continious variable)”, is concretely able to highlight the empirical basis for “travel demand management policies”. At that point, the “total system travel time (TSTT)”, which is an outcome of the fourth stage of travel demand models (network assignment stage), refers to the “daily travel time of individuals”, while the “daily number of trips” intuitively refers to the “trip generation stage (first stage)” of travel demand models. Such a kind of iterative process indicates that an amount of decrease in the “daily travel time of individuals” induces an amount of increase in the “number of daily trips”, which in turn does effect the “daily travel time”. Such an iteration, firtsly, makes the conventional four stage travel demand modelling processes much more realistic. Secondly, it proposes a kind of individuals based disaggragated modelling framework for transportation planners. Thirdly, such that iterative mechanism offers us to compare alternative transportation investments with regards to their related outcomes referring to different marginal elasticity coefficients of “number of daily trips”with respect to the “daily travel time” of the individuals. Herein, according to these different marginal elasticity coefficients, any potential new transportation investment would refer to an amount of “time loss/gain” when compared to others, which in turn re-affect the first stage (trip generation stage) of conventional travel demand modelling process, all of which are able to empirically highlight the backgrounding mechanism of the “cost & benefit analysis” of such investments with regards especially to the “travel demand management policies”. In other words, such these measurements propose policy makers a kind of control mechanism with regards to any transportation investment via their related effects on "daily induced travel demand" so as to minimize the number of daily trips as much as possible, all of which have been referring to the "main motivation of this thesis".

1.1 Purpose of Thesis

In literature, as asserted in detail in the "literature review" part of the study, the related studies have proved enough so far that passengers are willing to make more vehicle miles in their car trips as travel time decreases. This increase can come from either making longer trips for their existing trips by shifting the origin/destination of the trip, or making additional trips not needed or suppressed previously. That part of the research has not been highlighted yet, which underlies one of the unique aspects of this study in that it is aimed to investigate the interrelationship between "number of trips (instead of vehicle miles traveled)" and "travel time" via traveler based (disaggragated) data analysis in the case of İstanbul. In other words, the hypothesis, stating that "the less amount of daily travel time spent by a passenger, the more number of daily motorized trips is carried out by the traveler, which reciprocally causes an amount of increase in the daily travel time of this passenger." has been tested. Such a formulation of the "research question" would require "newest methodological framework". In this sense, currently available methods of econometric analyses reveal deficiencies in measuring the system-wide elasticity between the dependent endogeneous variables, namely "number of trips" and " travel time consmued" by each passenger, which affect each other reciprocally. Herein, the one (number of trips) is non-linear count variable by its nature, while the other (travel time) is a type of Gaussian distributed continious variable. If both were referred to the types of linearly distributed continious variables, the empirical interrelationship between the two would be easily modelled via classical "simultaneous equations models (SEM)", leaded by 2SLS (2 Stages of Least Squares estimation) or by 3SLS (3 Stages of Least Squares estimation) techniques, with consistently defined related "instrumental variables". But, for the case of this study, as mentioned above, the model structures would differ from the classical ones in that the one is non-linear count model, while the other is Gaussian distributed continious variable, which would require newest methodological perspective, such as "path analysis" conducted via "Generalized Simultaneous Equations Model (GSEM)". Moreover, post-estimation calculations (i.e., system-wide elasticity calculations, etc.) have been exhibited subsequent to testing such these models with the underlying mathematical explanations and logic asserted behind these models. All these indications have been constituting another unique aspect of this thesis.

Furthermore, the vitality of measuring motorized traffic with reference to the improvement of transportation infrastructure in urban spaces would mostly come from the concern of minimizing (or managing) motorized traffic so as to enhance the level of social benefits of urban societies via stimulating the usage of green modes of transport, which mostly refer to the non-motorized trips such as cycling, walking, etc., all of which stand for the concern of minimizing the level of gas emissions in urban spaces. In this sense, the thesis aims to measure the motorized traffic and then to constitute a baseline to monitor any travel demand management policies, regarding the motorized traffic flows in urban spaces. In addition, these considerations have been asserted in the thesis so as to make policy makers see the urgency of discouraging enlargement of new urban roads for motorized traffic, since each new motorized route induces extra motorized travel demand. All these aims are directly related to understanding the mechanism of motorized traffic of passengers on the urban scale and then to manage it. Herein, measuring induced motorized travel demand via the changes in average daily motorized travel times of the passengers would explicitly be able to feed the empirical baseline of designating and monitoring the performances of any travel demand management policy in urban spaces. In another words, any new improvement of transportation infrastructure would be able to make the average daily motorized travel time of individuals decrease, which would cause dramatic increase in the average number of daily motorized trips of these individuals. In this context, the comparison of the potential changes in motorized travel times -as the result of developing new optional transportation projects- would be added to the cost & benefit analyses of these related projects so as to select the one with optimum social benefit. Herein, it is referred to the related project's contribution to the benefit of urban society with reference to the inter-relationship between daily motorized travel times and the number of daily motorized trips. It is meant that as the average daily motorized travel time decreases, more number of motorized trips would be exhibited by each individual due to the fact that travel time is a kind of disutility for passengers.

On the other hand, it is not meant that the worst project with highest travel times would be selected within the process of selecting any transportation project among its options. Instead, it is here meant that the reciprocal relationship between the number of motorized trips and motorized travel times is to be investigated in that the decrease in travel times would induce new motorized trips, while these new induced motorized trips would make the decreased travel times enormously increase again. This reciprocal relationship would refer to a kind of recursive relationship between these two parameters with refers to the selection of potentially optimum transport project in the urban scale. From the view of traditional four stages travel demand forecasting models, the number of motorized trips generated is estimated in the first stage called trip generation stage, while the total system travel time- with reference to the total motorized travel time- is estimated in the fourth stage called network assignment. The findings of this study would be able to make this modelling process more recursive in that after predicting the total system travel time via the fourth stage called network assignment, then the first stage called trip generation is to be tested again till the optimization between travel times and number of motorized trips generated is grasped. Such this iterative process would be able to be implemented within the classical travel demand forecasting models and then the selection of transportation project (or bundle of projects) would be carried out according to the findings of this iterative process, which is able to be integrated into the travel demand models. Such this consideration would refer to a kind of milestone for the considerations of travel demand models in literature, which would refer to another multiplication effect of this thesis to the related literature.

Besides, such an optimizing concern would refer to a minimum-minimum problem (minimization as an optimization problem) in that the transport project (or a bundle of transport projects) in urban space is to be selected systematically with reference to the minimum travel times and to minimum number of motorized trips in a day as much as possible. To put it in a different way, the urgency of spreading non-motorized green modes of trips within urban spaces would be able to be justified via the findings of this study in that for motorized trips, it is explicitly seen that each new improvement or development of transportation facility would be able to induce new motorized travel demand resulting in a kind of vicious circle for urban transportation policy makers. This thesis just aims to initiate such these considerations in literature. Obviously, other than İstanbul, more measuring is required and the findings of the study should be improved via numerous urban areas with new studies.

1.2 Literature Review

The notion of induced travel demand refers mainly to two frameworks: 'diversion of the existing demand' and 'newly generated traffic'. In addition, the concept of 'newly generated traffic' refers to two sub-forms, namely 'release of the suppressed demand' and 'newly generated traffic with regards to the urban development effects'. That is to say, while the diversion of the existing demand and release of the suppressed demand refer to the short-run effects, newly generated traffic with regards to the urban development effects refers to the long-run effects reference to literature of induced urban passenger mobility demand.

The literature of induced urban passenger mobility demand mostly refers to the interrelationship between the Vehicle Miles Travelled (VMT) as the main dependent variable and the total travel time as the main independent variable. In other words, the most of the empirical studies of induced urban passenger mobility demand have focused on measuring the marginal effect of travel time on VMT. On the other hand, such these researches exhibit some deficiencies.

The first deficiency would be due to the different approaches in defining the spatial resolutions of the study area within the related studies. Herein, the ones, conducting the facility based (a neighbourhood unit with its surrounding) or corridor based (along a highway route) analyses, produce partial urban equilibrium marginal elasticities instead of system-wide urban equilibriums. Secondly, even though some other studies in literaure, reference to the urban space as a spatial unity in their related analyses, have been able to carry out the related system-wide marginal elasticities, these coefficients would still refer to the biased results due to the aggregated data structures of their models. That is to say, the travel survey data, disregarding the behavioral units (individuals based), may produce biased results that are far away from the reality.

The examples for the facility or corridor based studies were the ones, conducted by Pells (1989), Hansen et al. (1993), Kroes et al. (1996), Luk & Chung (1997), and Mokhtarian et al. (2000), while the area-wide studies involve the ones of Hansen & Huang (1997), Noland & Cowart (2000), Fulton et al. (2000), Cervero & Hansen (2002), Cervero (2003), Silva & Costa (2007), Ozuysal & Tanyel (2008), Holcombe & Williams (2010), Hymel, Small & Dender (2010), Melo, Graham, & Canavan (2012), and Vos & Witlox (2013). The facility or corridor based studies have mostly adopted the methodological frameworks, namely "growth comparison analysis" and "matched pair analysis" so as to grasp the related marginal elasticities in the percent form. On the other hand, the area-wide studies mostly involve the econometric models, such as Ordinary Least Squares (OLS) regression models, auto-regressive models, and travel demand models so as to get the marginal elasticity coefficient of travel time with regards to the related measures of VMT.

According to the findings, firstly, the facility or corridor based studies reveal that it has been possible to seperate the middle-run and long-run effects of changing travel time on the related VMT measures. The related marginal elasticity coefficients change between 0.15 and 0.30 for the four years time horizon; 0.30 and 0.40 for the ten years time horizon; 0.40 and 0.60 for the sixteen years time horizon (Pells ,1989; Hansen et al., 1993; Kroes et al., 1996; Luk & Chung, 1997; Mokhtarian et al., 2000). On the other hand, according to the findings of the area-wide studies, the related short-run marginal elasticity coefficient varies from 0.30 to 0.50 for the county level, while it falls between 0.54 and 0.61 for the metropolitan region scale (Cervero, 2002:4).

In addition to the differences in spatial resolution of the related studies, as it has previously been stated, the related models of these studies are able to be grouped into two: "aggregated models" and "disaggregated models". In this sense, the level of aggregation refers both to data gathering structure (whether conducting individual scale field surveys or not) and to the related model structures. In almost all these studies, the VMT has been defined as the main dependent variable. On the other hand, within the studies, involving aggregated time-series econometric models, the related independent variables are defined as the lane-miles additions with several time lagged variables and geographical variables, while within the disaggregated ones, the independent variables mostly refer to the "total travel time" and "average travel speed" in addition to the individuals based socio-economic variables. Moreover, the functional form of log-linear model specification has mostly been selected in such studies so as to grasp the related marginal elasticity coefficients. Within almost all the related studies, reference to both aggregated data and aggregated models, the findings would exhibit enourmously increasing aggregated estimation errors due to both data gathering processes and generalized functional forms. In addition, such estimation errors would increase as the study area is spatially expanded. In this context, the behavioral units (individuals based) based data gathering and modelling approaches are required so as to minimize the related estimation errors. At this juncture, the study of Barr (2000) attracts the cares as an interesting example. In that study, the households based field survey data has been gathered for the United States on the national scale. The models of this study, carried out via the methodological framework of cross-sectional data analysis, refer to the "logarithm of the VMT per household" as the main dependent variable, while the households based socio-economic variables have been defined as the independent variables (Barr, 2000). Furthermore, the related models have been stratified according to the spatial sizes of the related metropolitan regions, located in the United States. On the other hand, the related results of the study indicate that there has not existed any statistically significant differences in the related marginal elasticity coefficient estimations reference to these spatial size based stratifications (Barr, 2000).

In addition to the aggregated estimation errors of data & model structures, there exists another source of error in measuring the induced travel demand. This source of error is able to be defined as disregarding the "reciprocal relationship" between the dependent and main independent variables. In this sense, the main dependent variable, called Vehicle Miles Traveled (VMT), might exhibit a kind of simultaneous relationship with one of the preliminary independent variables, called "lane miles additions". That is to say, an increase in the total length of lanes via the lane mile additions would make VMT increase, indeed that increase in VMT would also make the travel demand be induced by the new lane miles additions. Disregard of such a relationship in designation of the related models would make the level of estimation errors enormously increase. To illustrate, such studies, asserting such a related simultaneity effect, involve the one of Noland & Cowart (2000) and the one of

Cervero & Hansen (2002). In the first example, the related simultaneity effect is coped with via the addition of “instrumental variables”, which theoretically justify the interrelationship between VMT and lane miles additions, while in the second example, the problem of simultaneity has been coped with via the two-stages least squares (2SLS) simultaneous equations model structure (Noland & Cowart, 2000; Cervero & Hansen, 2002).

There are also further examples in literature, tackling with the problems of endogeneity and simultaneity in a deeper manner. To illustrate, the study of Cervero (2003) asserted four simultaneous equations in the models reference to the dependent variables, namely “urban development”, “lane miles growth”, “VMT”, and “travel speed” (Cervero, 2003). In other words, Cervero developed four different equations with respect to these dependent variables, all of which refer to the “simultaneous relationship” between each other. According to the findings of this study, the related marginal elasticity coefficients of the related dependent variables exhibit higher amounts, when the related dependent variables are modelled without regarding the cases of simultaneity (Cervero, 2002:15; Cervero, 2003). That is to say, according to the results, the simultaneous relationship between the related dependent variables makes the related marginal elasticity coefficients decrease when compared to the ones of independently modelled dependent variables due to the reciprocal relationships of these variables.

Lastly, in some other studies, taking the measurement of induced urban passenger mobility demand into account, it has been investigated that whether the level of traffic congestion constitutes a statistically significant variance on the estimated marginal elasticity coefficients reference to the measures of induced travel demand or not. In this context, according to the findings of the study of Hymel, Small & Dender (2010), the level of traffic congestion creates a statistically significant variance on the induced travel demand estimations in a negative direction (as it is theoretically expected), which increases as the level of income of the passengers increase (Hymel, Small & Dender, 2010). On the other hand, according to the empirical findings of the study, carried out by Noland & Cowart (2000), the variance on induced travel demand estimations, that has been created by the level of traffic congestion, does not exhibit statistically significant measures (Noland & Cowart, 2000).

In the light of all these views, the literature of induced urban passenger mobility demand measurements are able to be seperated into three main methodological categories: “aggregated data collection procedure versus disaggregated data collection procedure”, “facility or corridor based studies versus area-wide studies”, and “single index model structures versus simultaneous equations model structures”. These methodologies based categories are also able to be stated via the headings, namely “data structure approach”, “spatial resolution approach”, and “model structure approach”, respectively.

In the light of these categoizations derived from the literature review, it is explicitly able to be concluded that the requirements for further researches on the measurement of induced urban passenger mobility demand might be exhibited as in the followings:

- Instead of the classical investigations on the marginal effect of change in travel time on VMT, the new researches -taking the marginal effect of travel time on specifically the number of trips into account- would be welcome. Via the findings of such these new studies, the travel demand management policies would be able to be assessed, according to their performance measures within a much clearer manner.
- Reference to the potential research question of such future studies, taking the “number daily motorized trips” and “daily motorized travel time” in the core as the main dependent variables, a type of convenient simultaneous equations model structure is to be developed.
- Reference to the data collection procedures, the disaggregated type of approaches are to be adopted in that the related field travel survey studies are to be conducted with regards to the behavioral units, namely individuals. Such kind of data collection approach is expected to make the aggregated estimation errors dramatically decrease.
- So as to grasp system-wide marginal elasticity coefficients, according to the asserted research designs, the spatial resolutions of the related studies should refer to “area-wide” approach, instead of the ones reference to the facility or corridor based approaches. Otherwise, the related estimated marginal elasticity coefficients would refer to the concept of “partial urban equilibrium”, which would explicitly fall short to highlight the practical sides of urban scale travel demand management policies.

1.3 Hypothesis

The hypothesis of the thesis is able to be explained as “the less amount of daily travel time spent by a passenger, the more number of daily motorized trips is carried out by the traveler, which reciprocally causes an amount of increase in the daily travel time of this passenger.”.

2. STUDY AREA, DATA & MATERIALS

This chapter of the thesis involves the detailed explanation of the study area, pilot study area, sampling design, sampling method, and data & materials.

2.1 Study Area

The research area spatially refers to an urban scale for İstanbul as a whole. The population of İstanbul in 2005 was around 10,500,000 and the administrative borders of the city covers 5,400 km^2. İstanbul was the capital of the Byzantine and the Ottoman Empires. Even though the capital moved to Ankara after the Republic of Turkey in 1923, the city has sustained her economical supremacy over the country. After the 1950, when the high rate of urbanization started in the country, İstanbul was the main destination of internal migration. Today, the city carries the 17 % of national population while the administrative area of the city is only 7 per ten thousand of the country. Furthermore, the city includes approximately 34 % of national manufacturing and 35 % of national financial employments, and approximately 44 % of foreign trade of the country comes from İstanbul.

İstanbul is located around the Bosphorus, the water strait separating the Asia and Europe. The historical core, which is the central business district, is circled area in Figure 2.1. Urban fabric starts from the south coasts and expands into the upper north forest areas and valuable agricultural lands. These forest areas include eight potable water dams, host many environmentally sensitive areas, and establish life support systems of the metropolitan area. With this fragile geography, a sustainable population of the city is around 16 million while the trends tend to 22 million implying very serious future environmental challenges.

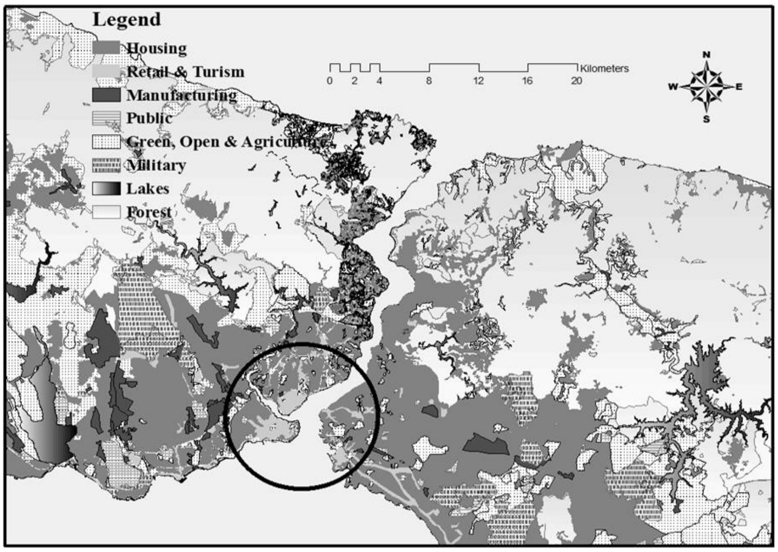

Figure 2.1 : Land use in the İstanbul Metropolitan Area (İBB, 2007).

The city is connected with suspension bridges, and the main destination of morning commuting is towards the Central Business District (CBD), and to the European rim since the European rim accommodates 65 % of population and 72 % of all employment. This unbalanced distribution of population and employment would have been caused by the lack of an extensive rail transit network aggravate traffic congestion in the city especially for the continent crossing. There are 21 million daily trips in the metropolitan area, and half of them are motorized as Figure 2.2 reveals 6 % of all trips make continent crossing. Approximately, 300,000 of them were carried by the ferries while 1,000,000 trips use the bridges.

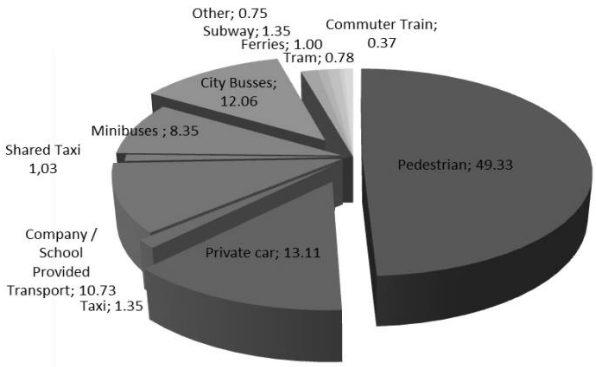

Figure 2.2 : Modal split of İstanbul Metropolitan Area (İBB, 2007).

The share of rail transit was only around 3 %, and marine transit around 1 %. Approximately 15 % of the whole trips were by private car, and the remaining 32 % was by rubber-tired public transit. Even though the share of private car and the trip rate are lower than the most western countries, the level of congestion in İstanbul is very high. There are three important reasons for such level of congestion. The first is that the city is highly dense especially in and around the CBD. The second is that the public transit system of the city heavily depends on rubber-tired public transit worsening the traffic congestion and air-pollution. The third is that feeder arterials of the bridges are used for intra-rim traffic and when they are congested by the continent crossing traffic, the intra-rim traffic gets to be congested in the both sides of the city.

There are some 147 km of rail network with different capacities within the city currently. When ongoing rail projects are completed, there will be approximately 270 km of core rail network in İstanbul (see Figure 2.3). Even though all is completed without any delay, the rail network will be lower than those of world's comparable sized metropolitan areas. The late transportation network plan states that it is possible to extent the rail network to 550 km if enough level of resources is raised.

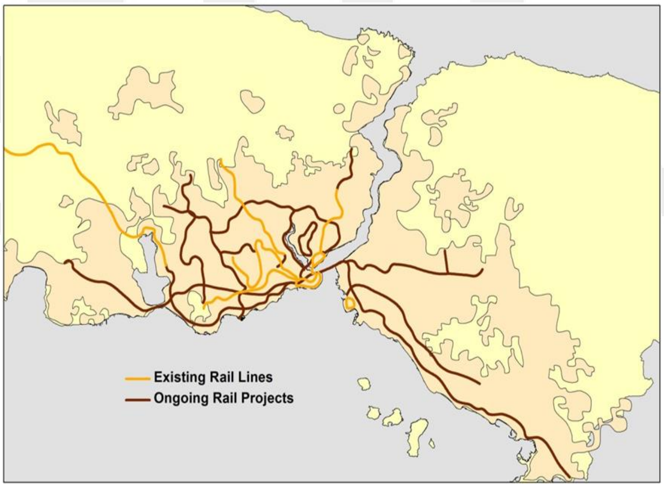

Figure 2.3 : Railway network of İstanbul (İBB, 2007).

2.2 Pilot Study Area

The pilot study aims to (i) detect the parameters of the probability distribution function of the trip generation rates in İstanbul Metropolitan Area (with regards to the coefficient of correlation), (ii) compare the operational effects of differently designed questionnaires in the field research, and (iii) designate the research instruments.

In the study, there have been carried out five different methods for conducting field survey with refers to the questionnaires; namely, questionnaires with face to face interaction, questionnaires with followed face to face interaction, questionnaires with followed by telephones for face to face interaction, questionnaires with computer aided telephone interaction by just one call, and questionnaire with followed by computer aided telephone interaction. In this sense, each one was conducted with 300 respondents, which makes 1,500 respondents in total.

To begin with, in the face to face interaction, it was made an appointment for each selected household especially in the mornings for conducting the questionnaires especially in the evening hours in that the related respondents will be at their homes so as to make the response rate increase. Secondly, the questionnaires with followed face to face interaction was splitted up into two. Herein, the first day was allocated to collect the socio-economic data for each member of the related household via submitting questionnaires to the members of the households and via explaining the details for their answers so as to make them answer the questions for their following daily trips of the following day. As the following day ended, the related household was visited on the next day to check their answers to the questionnaires and to help them if there were some unanswered questions.

Thirdly, in the questionnaires with followed by telephones for face to face interaction, the related information of the individuals in the household was recorded by directly the technical workers of the field survey. Then, this household is informed that they will be called by telephone to record information about their daily trips, which will be written in the questionnaire sheets that were enough copied for each member of the related household. Hence, all their related information for their daily trips was collected by phone calls instead of collecting the answered records of their daily trips by another visit.

Fourthly, in the questionnaire with computer aided telephone interaction by just one call, each selected household is called by telephone so as to collect the related data for their daily trips for the present members of the household at the time of the phone-call. The related data was recorded by the computer technology.

Fifthly, in the questionnaire with followed by computer aided telephone interaction, more than one telephone call were realized, which makes it different from the previous method. In the first telephone call, the socio-economic data for each member of the related household was collected, and then the household was informed that they will be called by telephone to collect data about their daily trips in the following day (with refers to the household members who are above 6 years old). Afterwards, the second telephone call was carried out so as to collect the related data for their daily trips, which was systematically recorded by the computers.

The face to face questionnaires were conducted to randomly selected 45 sampling units of different counties that each involves 20 households. The remaining 600 questionnaires were conducted by computer aided telephone interaction and each sampling unit was randomly selected among the whole list of recorded households (listed by their telephone numbers). According to the related records, a full list of required answers to the questionnaires were given by all the members of just 218 households (among 1,500 households), while only some members of 90 households gave full considerations for their answers (the remaining members of these households did not respond to these questionnaires). Furthermore, 408 households gave only their socio-economic information, the remaining trip data based questions were not answered by them, while 784 households were not able to be contacted.

The detailed response rates per each method has been revealed as in the Table 2.1. As it is explicitly seen from the table, the most efficient method, with refers to the response rates, is the face to face interaction.

Table 2.1 : The response rates per each method of conducting field survey by questionnaires (Genar, 2006, p. 24).

		Questionnaire Completed	Half of the Questionnaire Completed	No one at home	Questionnaire was postponed	Appointment was rejected	Total
Face to face interaction	**Frequency**	119	61	86	-	54	320
	%	37.2	19.1	26.9	-	16.9	100
With followed face to face interaction	**Frequency**	48	103	86	-	43	280
	%	17.1	36.8	30.7	-	15.4	100
With followed by telephones for face to face interaction	**Frequency**	23	155	77	-	37	300
	%	7.7	51.7	25.7	-	12.3	100
With computer aided telephone interaction by just one call	**Frequency**	14	272	-	1	14	302
	%	4.6	90.1	-	0.3	4.6	100
With followed by computer aided telephone interaction	**Frequency**	14	262	-	1	21	298
	%	4.7	87.9	-	0.3	7.0	100
Total	**Frequency**	218	853	249	11	169	1500
	%	14.5	56.9	16.6	0.7	11.3	100.00

In the light of these observations, as mentioned before, there have been collected full list of trip data from 734 individuals belonging to 218 households in total in this pilot survey. These 734 individuals carried out 1,127 thousand daily trips in total. According to this, gross mean number of the daily trips per person is 1.54, while the standard deviation of this ratio is 0.048. In this sense, the coefficient of variation for total number of trips is equal to 0.85. So, with regards to the 95 % confidence level (0.05 % error), the sample size is 1,11 thousand individuals. When considering the average sample size, which is equal to 3.60, fully completed questionnaries with approximately 360 household units will be statistically enough for such this research.

2.3 Sampling Design

One of the prioritized tasks, after pilot study, is to determine for the sample size. In this sense, according to previous three different household surveys, the sampling rates were 0.08 % (with 1,2 thousand households), 0.16 % (with 2,4 thousand households), and 0.42 % (with 11,795 thousand households) respectively in the years 1985, 1987, and 1997. As discussed above, for a basic variable called trip ratio, it would be required at least 1000 individuals within an urban scale in İstanbul. On the other hand, the sample size is to be much larger especially for representing the frequency distribution of the travel distance and related trip matrix on behalf of estimating travel demand functions for modal split ratios of trip distributions significantly. Besides, as the sample size increases, the sampling error tends to decrease dramatically. Therefore, the sample size, which is to be as higher as possible in such a research, will be upper limited by the research budget constraints.

In the light of these considerations, with reference to the related market research and to the budget research of the İstanbul Metropolitan Municipality, it was seen that it would be possible to conduct such a research with 90,000 households, which refers to a sampling ratio of 3 %. This ratio has been the highest one for the related household researches in İstanbul up to now.

2.4 Sampling Method

Since the study area refers to İstanbul as an urban spatial unit as a whole, it involves 33 counties and related 987 districts. The main motivation behind sampling method is able to be stated as grasping the spatial variation of the travel demand patterns in

İstanbul. This variation would be grasped by district based simple random sampling, but such an approach would require a sampling frame, which refers to the full list of buildings with the related address based full records of households in İstanbul. Herein, even though these statistics would be theoretically able to be collected from the Turkish Statistical Institute, such these statistics would not have been updated reliably since İstanbul is a kind of rapidly (and endlessly) developed huge city. Therefore, the self-weighted, two stages random cluster sampling has been adopted as the convenient sample selection method.

2.5 Data and Materials

Data used in this study comes from the 2006 Household Origin-Destination (O/D) Survey of İstanbul Metropolitan Area. This survey was coordinated by the advisor of this thesis in 2006, and the related cost of such this sophisticated survey was approximately 3,200 million Turkish Lira in that year, revealing that updating such a big database would not be economically feasible, which also refers to the preliminary research limitation of this thesis. The survey was conducted by the Department of Transportation of the Metropolitan Municipality of the İstanbul. The sample size in this survey was 90,000 households, and sampling rate was 3 %. Samples were chosen by two-stage random cluster sampling. At the first stage, approximately 4,000 primary sampling units (PSU) were randomly drawn from the latest building list of the Metropolitan Municipality (*see* Figure 2.4).

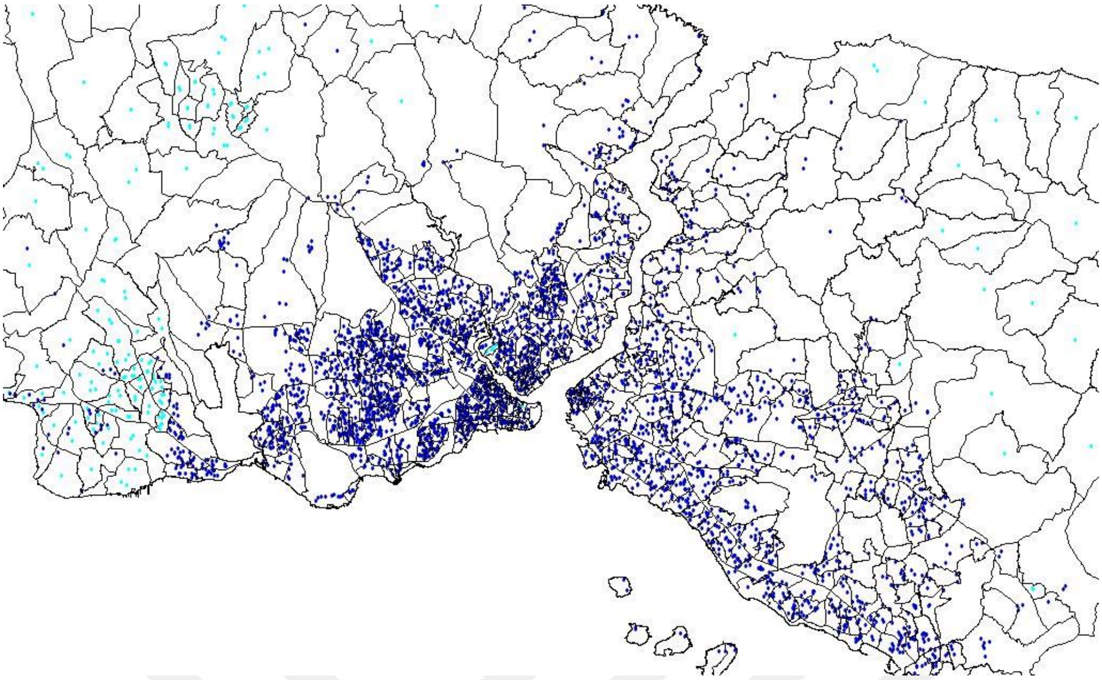

Figure 2.4 : The spatial distribution of the starting points of the primary sampling units (İBB, 2007).

In the second stage, 30 households out of the 90 recorded households around the PSU were surveyed systematically as the secondary sampling units in 450 traffic analysis zones. Each address of primary sampling units constitutes the starting point for each secondary sampling unit. Beginning with each of this address (from primary sampling units) as the starting point for the secondary sampling units, each neighborhood unit -involving the related recorded buildings- was walked clockwise and each housing unit in this spatial frame is recorded in a list. Afterwards, each currently used household unit is enumerated from 1 to 90. Then, a number from from 1 to 3 was selected among these random numbers per each housing unit that will be visited in this neighborhood unit. This selection continued till all 30 visiting housing units among 90 units were selected. Herein, no sampling replacement was allowed. The walking rule to select each visiting unit among the selected three ones is represented in the Figure 2.5.

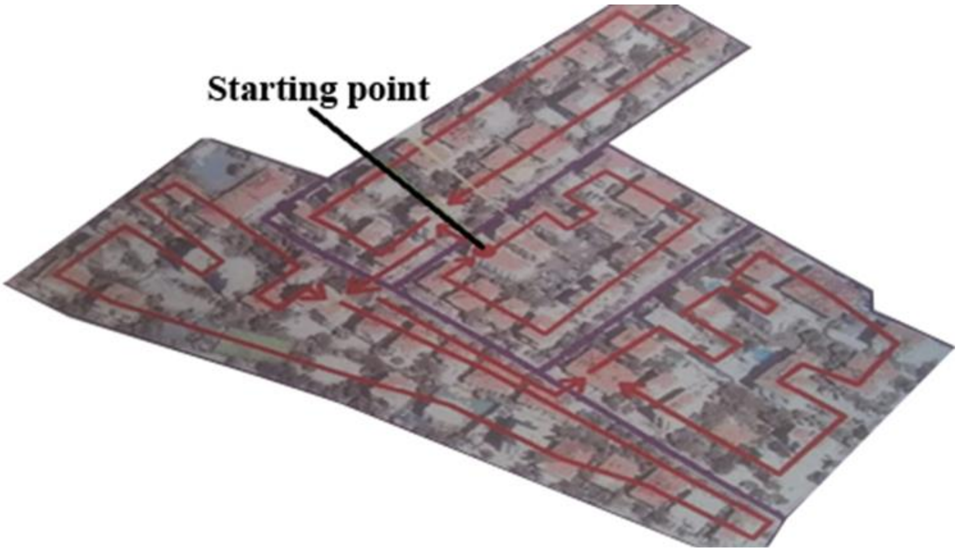

Figure 2.5 : Walking rule to designate the secondary sampling units (İBB, 2007).

Once the samples were selected, no sample replacement was allowed to avoid non-sampling errors. In this way, 80 % unit response rate was achieved. In 72,000 households, a total of 264,000 passengers were interviewed face to face, and a total of 356,000 daily trips were recorded between 451 Origin-Destination pairs defining 203,401 distinct movements. All the related database has been stored in the software called ***Transcad 4.8*** subsequent to processing raw data of field survey by the software called ***SAS 15.0.0***. The Zone system of İstanbul is given in Figure 2.6 and Figure 2.7.

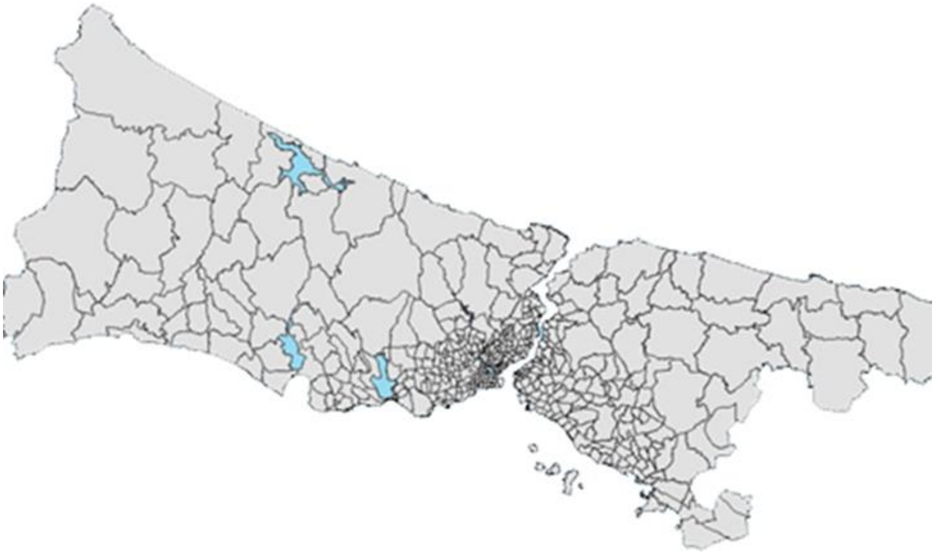

Figure 2.6 : Traffic Analysis Zones of İstanbul Metropolitan Area (İBB, 2007).

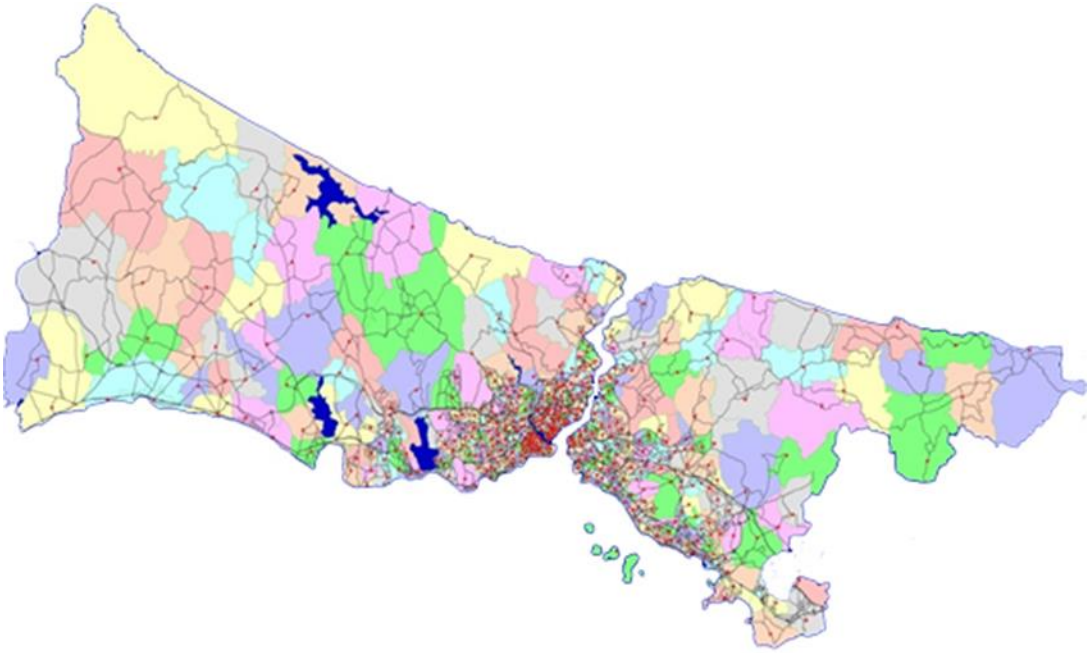

Figure 2.7 : Traffic Analysis Zones of İstanbul Metropolitan Area and highway network (İBB, 2007).

Set of explanatory variables included the negative of total motorized travel times (motor_time) spent by an individual in daily motorized trips, and personal & family characteristics. In addition, the motorized free flow travel time (motorfft)- as a purely disaggregated intermediate traffic parameter- has been asserted especially in the GSEM. Travel time variables are the key variables, and their elasticities with reference to total number of motorized trips are assumed to reflect the willingness to travel more depending on a reduction in daily motorized travel times. The other explanatory variables are dummy variable specifying sex (male_d), dummy variable specifying if the individual is household head (hh_head_d), dummy variable asserting whether the respondent exhibited at least one motorized trip in the stated day or not (mobility_dummy), household disposable income in thousand Turkish liras (hh_income), number of the private cars in the family (auto_number), size of the household (hhsize), age of individual (age), dummy variables if individual has realized at least one home based work trip(s) or not (hbw_d) or home based work trip(s) (hbs_d) respectively, and year of schooling the individual attended (schooling_year) (Table 2.2).

Table 2.2 : Variable definitions.

Variable	**Variable Definition**	**Measurement Unit**
motor_y	number of motorized trips carried out by an individual traveller, in a randomly given day (within 24 hours) .	trip numbers.
motor_time	amount of time (mn.) spent by an individual traveller, through motorized trips in any given day (within 24 hours) .	minutes.
motor_distance	amount of distance (km.) spent by an individual traveller, through motorized trips in any given day (within 24 hours) .	kilometer (km).
male_d	dummy variable. "1", if traveler is male, "0" , otherwise.	0 or 1.
hh_head_d	dummy variable. "1" : if the traveler is the household head. "0" : otherwise.	0 or 1.
hh_income	household disposable income per month (TL)	TL (turkish lira).
auto_number	number of the private cars in the family	number of cars.
hhsize	number of people belonging to the related household.	number of people in household.
age	age of the individual	number of years.
hbw_d	dummy variable."1" ,If the individual traveler has realized at least one home based work trip(s) in the given day, "0" , otherwise.	0 or 1.
hbs_d	dummy variable."1" ,If the individual traveler has realized at least one home based school trip(s) in the given day, "0" , otherwise.	0 or 1.
schooling_year	number of years of schooling that the individual has attended.	number of years.
mobility_dummy	dummy variable whether the individual traveler selects motorized travel in the given day (1) or not (0).	0 or 1.
motorfft	is the traffic volume parameter, indicating that, if there were only one travelling unit, the hypothetic time of the travel (reference to that of between the related origin and destination) , experienced by the traveller, with regards to the related free flow speed.	minutes.

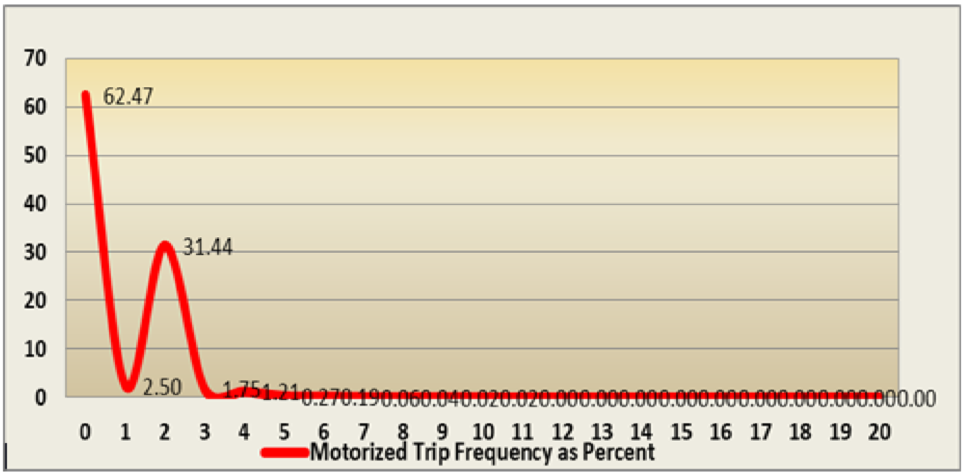

Figure 2.8 : Motorized trip frequencies.

All people older than 16 are included in this study and the sample size included 194,000 people. The means and the standard deviations of all variables used in the models are presented in Table 2.3. Total motorized number of daily trips made by an individual (motor_y) is the dependent variable in the models. The frequency distributions of motorized trips are given in Figure 2.8. Zero trips have the highest frequency in motorized trips. Number of the trips was obtained from the 2006 household survey while trip distances are the system produced shortest paths between the O/D pairs.

Table 2.3 : Descriptive statistics of variables.

Variable	mean	st. dv.	Variable	mean	st. dv.
motor_y	0.79	1.12	**auto_number**	0.39	0.58
Ln_motor_y	0.41	0.55	**hhsize**	3.92	1.08
motor_time	37.85	66.53	**age**	37.83	15.45
motor_distance	8.48	16.96	**hbw_d**	0.34	0.47
male_d	0.5	0.5	**hbs_d**	0.05	0.23
hh_head_d	0.36	0.48	**schooling_year**	7.35	3.87
hh_income	1.057	1.28	**mobility_dummy**	0.37	0.484
motorfft	0.63	1.11			

3. METHODOLOGY

In this chapter, subsequent to the discussions on the nature of the indicated variables of the research, all the aserted model structures have detailly been discussed in that whether they would be appropriate for the nature of the empirical research of this thesis or not. In this sense, Poisson Regression Model (PRM), Negative Binomial Regression Model (NBRM), Sample Selection Model (SSM), Sample Selection Poisson Regression Model (SSPRM), Instrumental Variable Poisson Regression Model (IVPRM), Instrumental Variable Zero Truncated Poisson Regression Model (IVZTPRM), Generalized Simultaneous Equations Model (GSEM), and Two Stages Least Squares (2SLS) Regression Model have been widely discussed.

3.1 Model Discussions: Towards GSEM

According to the research question of the thesis, as mentioned before, the interrelation between the "number of daily trips of a passenger" and "the daily motorized travel time" of the related passenger are required to be modelled. Such an investigation requires a modelling framework, that is able to cope with the followings respectively;

i. Non-linear nature of "number of daily trips of an individual" as a count variable .

ii. Excess-zero observations in "number of daily motorized trips of an individual (motor_y)".

iii. Endogeneity of the "daily motorized travel time" of the related individual.

To begin with, according to the first requirement (i), non-linear model structures come into agenda. In other words, so as to model the count variable, namely "number of daily trips of an individual (motor_y)", the convenient count models, that have non-linear nature in theirselves, come into considerations. Within that categorization, Poisson Regression Model (PRM) and Negative Binomial Regression Model (NBRM) are the leading model structures (Green, 2007; Cameron & Trivedi, 2005).

To begin with, the Poisson Regression Model (PRM) is the most basic form of the count models. According to the poisson model, the random variable $\boldsymbol{y_i}$ shows a poisson distribution, and mean of this distribution is $\boldsymbol{\lambda_i}$ as indicated in equation 3.1.

$$Prob\ (Y_i = y_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} \quad (3.1)$$

$$y_i = 0, 1, 2, \ldots.$$

The mean of the distribution, $\boldsymbol{\lambda_i}$, is explained by a set of variables, $\boldsymbol{x_i}$, the formulation to estimating model parameters (except for the dummy ones) is the log-linear model (equation 3.2).

$$\ln \lambda_i = \beta' x_i \quad (3.2)$$

The basic assumption of this model is the equidispersion, meaning that conditional mean and conditional variance are equal (equation 3.3).

$$E[y_i\,|x_i] = Var[y_i|x_i] = \lambda_i = e^{\beta' x_i} \quad (3.3)$$

The elasticity with respect to any given variable is nonlinear, and can either be estimated at the variable means or as the mean of individual elasticities in the sample (equation 3.4).

$$\frac{\partial E[y_i\,|x_i]}{\partial x_i} = \lambda_i \beta = \lambda_i e^{\beta' x_i} \quad (3.4)$$

The PRM is nonlinear and maximum likelihood can be used for parameter estimation as a mathematical simplicity (equation 3.5).

$$\ln L = \sum_{i=1}^{n}[-\lambda_i + y_i\,\beta' x_i - \ln y_i!] \quad (3.5)$$

Equidispersion implicitly assumes that “the formula for the probability of an occurrence is a deterministic function of the explanatory variables –it is not allowed to differ between otherwise- identical individuals” (Kennedy, 1998, p. 247). However, this assumption is relaxed by introducing an “unobserved heterogeneity” effect into the conditional mean, called “scale variable”. This leads to a different model, the NBRM, in which the conditional variance is larger than conditional mean as revealed in equation 3.6.

$$P(y_i|x_i) = \frac{\Gamma(\theta+y_i)}{\Gamma(\theta)\Gamma(y_i+1)} r_i^{y}(1-r_i)^{\theta} \text{ ,where } \quad r_i = \frac{\lambda_i}{\lambda_i + \theta_i} \quad (3.6)$$

Conditional mean of this distribution is $\boldsymbol{\lambda_i}$ and conditional variance $\boldsymbol{\lambda_i(1 + (1/\theta)\lambda_i)}$. The elasticities of the NBRM are still estimated as asserted above (Green, 2007).

This fact implies that it is important to test for overdispersion if you use the PRM. Even with the correct specification of the mean structure, estimates from the PRM, when there is overdispersion, is inefficient with standard errors that are biased downwards (Long, 1997, p.236). Several tests are suggested for overdispersion (Green, 2003; 2007) without estimating a NBRM. Since the PRM and the NBRM are nested, the log-likelihood of the NBRM needs to be improved over the PRM in case overdispersion is present, and this can be checked by a log-likelihood ratio (LR) test (equation 3.7):

$$LR = 2 \times (\ln L_{NBRM} - \ln L_{PRM}) \tag{3.7}$$

LR shows a chi-square distribution, and any value larger than critical threshold with two degrees of freedom favors the NBRM.

The case of "overdispersion" in count data is able to exist due to unobserved heterogeneity, in which the events are seriously independent. Therefore, the rate parameter, with reference to the conditional mean, becomes to behave as a random variable itself, which requires further modelling approaches, namely mixed modelling approach (Cameron & Trivedi, 2005). At this point, "Negative Binomial Regression Model (NBRM)", which is taken into consideration as the specific kind of "Mixture Modelling Approach", comes into considerations (Cameron & Trivedi, 2005).

Modelling total number of daily motorized trips via negative binomial regression model, which is a specific kind of parametric count regression models, requires a special kind of care. Herein, the indication is that "mean is not equal to the variance" might be due to unobserved heterogeneity. Therefore, the conditional mean -as a rate parameter itself- is taken as a random variable with refers to the baseline of the mixture approach. In this context, the Negative Binomial Regression Model (NBRM) is taken into consideration as a special kind of mixture models (equation 3.8). In this framework;

$$f(y|\lambda)= \frac{\exp(-\lambda)*(\lambda^{y})}{y!}$$

$$h(y|\mu\ , a)=\int f(y|\mu, v) * g(v|a)dv \tag{3.8}$$

Herein, the function $h(y|\mu,a)$ refers to the marginal density of y, where $\lambda=\mu*v$ in that μ is a deterministic function of y, $v>0$ with density function $g(v|a)$ -reference to mixing distribution function- with unknown parameter ***a***. Also, ***f(y|λ)*** is a "poisson density function" and as revealed in the equation 3.9;

$$g(v)=(v^{ʃ-1} \times e^{-vʃ} \times ʃ^{ʃ}) / \Gamma(ʃ)$$

where v, ʃ>0, is the "gamma density function"

with E[v]=1 and Var(v) = 1/ʃ

$$E[v] = 1 \text{ ,since } E[\lambda|\mu]=\mu \qquad (3.9)$$

Thus, "negative binomial" is able to be obtained as a "mixture density" (equation 3.10);

$$h(y|\mu, ʃ) = \int_0^{\infty} \frac{e^{\mu v}(\mu v^{y})}{y!} * \frac{v^{ʃ} e^{-vʃ}(ʃ^{ʃ})}{\Gamma(ʃ)} dv$$

$$= \frac{\Gamma(a^{-1}+y)}{\Gamma(a^{-1})\Gamma(y+1)} * \left(\frac{a^{-1}}{a^{-1}+\mu}\right)^{a\wedge(-1)} * \left(\frac{\mu}{\mu+a^{-1}}\right)^{y} \qquad (3.10)$$

Herein, ***Γ(.)*** denotes the "gamma integral", specifying a "factorial for integer argument" (Cameron & Trivedi, 2005). Also, all the related equations asserted for NBRM refer to the two moments, which is indicated as equation 3.11;

$$E[y|\mu,a]=\mu$$

$$Var(y|\mu,a) = \mu(1+a\mu) \qquad (3.11)$$

Herein, the requirement for the variables, called "motorfft & mobility_dummy", to be the instrumental variables, refers that (equation 3.12);

$$Cov(mobility\_dummy, \varepsilon 1)=0,$$

$$Cov(motorfft, \varepsilon 1)=0 \qquad (3.12)$$

Here, ***ε1*** is the model residual reference to the endogeneous variable (Cameron & Trivedi, 2005). It is needless to say that there are many variations of such models to improve the estimation efficiency (please *see* Long, 1997; Cameron & Trivedi, 2005; Green, 2007 and Winkelmann, 2008 for details). Application of these models takes place in many diversified areas: crime analysis, disease occurrence, doctor visits, occupational injuries, software faults, accident analysis and prevention, manufacturing defects to name the few. On the other hand, these models are only capable of dealing with "non-linear nature of the main dependent variable (motor_y)". In other words, such these models are not able to cope with the second (ii) and third (iii) requirements, that are asserted above with reference to "excess-zero obervations in daily number of trips" and "endogeneity of travel time".

According to the second requirement (ii), the dependent variable, called “number of motorized trips per traveler per day (motor_y)”, includes so many zeros as observations, which refers to the problem of “excess zero observations” in econometrics. In other words, it has been required further modelling approaches with reference to the problem of “preponderance of excess zero observations” in the variable of “motor_y”. The related models, dealing with also “excess zero observations”, are “Zero Truncated Models” (Cameron & Trivedi, 2005). In this context, the assumption is that that the related data occurs only over the range of the response varriable (Cameron & Trivedi, 2005). In other words, zero counts are not taken into consideration in that they are eliminated from the models, that is why such these models are called “Zero Truncated (or Left Truncated) Models” (Cameron & Trivedi, 2005). On the other hand, that scope (zero-truncated models) is only able to deal with the first and the second requirements (i & ii), asserted above, reference to the “non-linearity” and “the problem of excess zero observations” of the daily amount of trips (motor_y). On the other hand, the endogeneity of “motor_time” (iii) still remains as a problem, since this model may have another important specification problem probably causing ***endogeneity bias***: the dependent variable and the key independent variable (i.e. total of reported travel times in minutes spent in these motorized travels) might have causal relationship. In other words, dependent variable is determined by an explanatory variable in a way as the explanatory variable is also determined by the dependent variable in turn. In such situations, since the error term is correlated with the dependent variable(s), the conventional methods produce biased parameter estimates and standard errors, leading to invalidation of the core results.

In principle, endogeneity bias is a form of omitted variables bias, and Mokhtarian and Cao (2008) summarizes seven different techniques to deal with endogenity problems: (i) *direct questioning,* (ii) *statistical control,* (iii) *instrumental variables model,* (iv) *sample selection model*, (v) *joint discrete choice model* (vi) *cross-sectional structural equations,* and (vii) *longitudinal models.* Concerning our model, only three of them, namely instrumental variables, sample selection model, and structural equation models seem meaningful for the endogeneity problem in modelling “number of daily trips” of an individual. Herein, time standing instrumental variables estimator would be an alternative to eliminate the bias coming from omitted variables, and these omitted variables would be the network information in our case. However, the discrete nature of our dependent variable, and difficulty of obtaining a good set of instrumental variables have made the sample selection model a better alternative to deal with endogeneity problem. Besides, the structure of our data, selection of trip making or not-making, makes Sample Selection Model (SSM) a viable alternative.

SSM is a two-stage model. In the first stage, a selection equation is specified as a binary outcome. That was trip making or not making in our case with respect to socio demographic variables (equation 3.13);

$$z_i^* = (w_i' \times \alpha) + e_i, \\ z_i = 0 \quad if \quad z_i^* \leq 0 \quad and \quad z_i = 1 \quad if \quad z_i^* > 0 \tag{3.13}$$

This selection equation is generally estimated as a binary probit model. In the second stage, an outcome equation is estimated. Independent variables included were total of travel times in daily trips in addition to socio demographic variables (equation 3.14);

$$y_i^* = x_i' + \mu_i, \\ y_i = y_i^* \quad if \quad z_i = 1;\ y_i\ not\,observed\,if\ z_i = 0 \tag{3.14}$$

In this formulation two errors, ***u*** and ***e*** "are assumed to have correlation ρ…the two sets of explanatory variables, ***w*** and ***x***, need not be disjoint, and, indeed, in some empirical applications, they are identical" (Breen, 1996). As Breen (1996) and Mokhtarian & Cao (2008) stated, the correlation of these two error terms is attributed to the omission of explanatory variables. Heckman's (1979) estimator is recommended as an unbiased alternative for parameter estimation in SSM.

Even though Sample Selection Model can eliminate endogeneity bias, the bias that come from a discrete dependent variable still might exist in our models. However, Terza (1998) elaborated on count data models with Sample Selection Poisson Regression Model (SSPRM) that may eliminate a significant portion of above mentioned estimator and specification bias. Furthermore, Terza used the derived model to estimate a model's parameters of the daily trip frequency (the same dependent variable as ours) versus a set of explanatory variables. Just like our model, the model had an endogeneity problem with one of the explanatory variables, namely number of cars a person owns. The model proved a successful estimator under the given constraints. For formal treatments of all these models and their marginal effects can be found extensively in Maddala (1983), Breen (1996), Terza (1998) and

Green (2003). To provide a comparative evaluation opportunity, all the mentioned models above are presented. In this way, number of motorized daily trips of an individual has been modelled. In this sense, in addition to the Poisson Regression Model (PRM) and Negative Binomial Regression Model (NBRM) as benchmarking modelling structures, Sample Selection Model (SSM) and Sample Selection Poisson Regression Model (SSPRM) have been asserted. *STATA-15* software package was used for the estimation of these models. STATA-15 was also included built-in estimators to calculate the marginal effects of all variables at variables' means, as the results of all these are exhibited in the next chapter of the thesis called Model Results.

In addition to the PRM, NBRM, SSM, and SSPRM structures, also nonlinear Generalized Method of Moments (GMM) with instruments model structure has been asserted. In this sense, in the case that one of the main dependent variables is linear, while the other is non-linear (or count) variable, an alternative approach would be GMM with instruments (Cameron & Trivedi, 2005). Herein, it would be implemented by obtaining fitted values for the endogeneous regressor and performing convenient nonlinear regression, such as Poisson regression and then doing regular linear Instrumental Variable (IV) via using the fitted value as the instrument for the count variable (Cameron & Trivedi, 2005).

For the concept of GMM with instruments, the model disturbance term would refer to the following when considering the nonlinear regression model, in which the error term may be additive or nonadditive (equation 3.15);

$$\mu_i = r\,(Y_i, X_i, \beta) \tag{3.15}$$

where $\boldsymbol{Y_i}$ represents the main dependent variable, $\boldsymbol{X_i}$ stands for the related explanatory variables, and $\boldsymbol{\beta}$ represents the parameter estimation for each explanatory variable in the nonlinear model with reference to additive error by the special case (equation 3.16);

$$\mu_i = Y_i - g(X_i, \beta) \tag{3.16}$$

where ***g(.)*** is a kind of specified function and the following conditional moment is expected to be observed if the number of instruments $\boldsymbol{(z_i)}$ is at least equal to the number of explanatory variables (Cameron & Trivedi, 2005, p.193) (equation 3.17);

$$E[\mu_i \mid z_i] = 0 \quad (3.17)$$

In the context of nonlinear GMM with instruments, two models, namely Instrumental Variable Poisson Regression Model (IVPRM) and Instrumental Variable Zero Truncated Poisson Regression Model (IVZTPRM) have been indicated. For the Instrumental Variable Poisson Regression Model (IVPRM), the GMM model structure with instruments has been defined as in the following (equation 3.18);

$$(\text{motor\_y}) = f\,(X_i),$$
$$(\text{motor\_time}) = g\,(Z_i), \quad (3.18)$$

where $\boldsymbol{X_i}$ stands for the explanatory variables, which are male_d, hh_head_d, hh_income, auto_number, hhsize, age, schooling_year, hbw_d, and hbs_d, while $\boldsymbol{Z_i}$ represents the instruments called motor_distance, motorfft, and mobility dummy. In addition, in the equations above, ***f*** represents the poisson regression function for the count variable called motor_y (daily number of motorized trips per passenger) and ***g*** stands for Instrumental Variable (IV) linear function for the endogeneous variable called daily motorized travel time per passenger (motor_time) with refers to the instruments.

Furthermore, in the Instrumental Variable Zero Truncated Poisson Regression Model (IVZTPRM), an additional condition has been asserted when compared to the one of IVPRM structure. This condition is directly related to the elimination of zero counts from the daily number of motorized trips per passenger (motor_y) due to getting rid of the problem called preponderance of zero counts. Then, the IVZTPRM gets the form as revealed in equation 3.19;

$$(\text{motor\_y} \mid \text{motor\_y} > 0) = f\,(X_i),$$
$$\text{motor\_time} = g\,(Z_i \mid \text{mobility\_dummy}=1), \quad (3.19)$$

Therefore, so as to guarantee that number of daily motorized trips per passenger is non-zero, it is given that the dummy variable asserting that whether the passenger carries out at least one motorized trip in a given day (mobility_dummy) or not is equal to ***1*** in the IVZTPRM structure. By adding this condition, the non-motorized individuals are systematically eliminated from the system, which makes number of daily motorized trips per passenger non-zero.

On the other hand, all the model structures (except for IVZTPRM), asserted up to here, have not been able to tackle with the technical constraints, namely ”non-linear nature of daily number of trips”, “excess zero observations in daily trips”, and “endogeneity of daily travel times” at the same time. In this sense, PRM and NBRM, as count models, are able to deal with only the non-linearity of daily number of trips. Besides, SSM structure is only able to tackle with the excess amount of zero observations in daily trips. On the other hand, SSPRM can overcome both non-linear nature of daily trips and excess amount of zero observations in daily trips, while the endogeneity of daily travel time has not been able to be overcome by this model structure. Furthermore, IVPRM is not able to overcome the excess amount of zero observations in daily number of trips, while it can tackle with the non-linearity of daily amounts of trips and endogeneity of daily motorized travel time. Lastly, IVZTPRM structure seems to be able to overcome all these three technical obstacles called ”non-linear nature of daily number of trips”, “excess zero observations in daily trips”, and “endogeneity of daily travel times”. On the other hand, IVZTPRM is also to be eliminated due to the failure of equidispersion assumption of PRM structure in it, which asserts that the mean and variance of the main dependent variable (number of daily motorized trips) is equal to each other. This failure is explicitly able to be validated by the Log Likelihood Ratio (LR) test (*see* equation 3.7), the result of which is exhibited in the following part of the thesis called Model Results.

Hence, it is required further newest modelling approaches, which have not been highlighted yet by any of these model structures. In fact, the intuitive effort, exhibited up to here, comes from the intention of justifying one of the final model selections; namely, “path analysis” carried out via “Generalized Structural Equations Model (GSEM)” (see Table 3.1). In this model, the dependent variable is the total number of motorized trips done by an individual within 24 hours. The main explanatory variable is the negative of total motorized travel time spent in minutes for these daily trips, since travel time defines a disutility. The remaining explanatory variables are the personal and family characteristics, as explained in the data section. Furthermore, there is also one more dependent variable, coded as “mobility_dummy”, with its explanatory variables, based on individuals’ socio-economic characteristics. At this juncture, since the term “mobility_dummy” refers to the binary variable (see Table 2.3), it refers to a probability structure as a dependent variable. In the light of these views, the related GSEM model structure seems to be in the form, as asserted in Table 3.1 and in Figure 3.1.

Table 3.1 : Designation of the GSEM structure[1].

Dependent Variable	Function of Exploratory Variables	Function Family	Function Link
motor_y	f (motor_s, X_i)	negative binomial mean distribution	logarithm
motor_time	g (estimated_motor_y[2])	normal distribution (Gaussian Family)	identitiy link
mobility_dummy	Φ (X_i)	probit function family	
motorfft	h (motor_distance, motor_time, mobility_dummy, X_i)	negative binomial mean distribution	logarithm

[1] Where, ***f*** is the negative binomial distribution function with the conditional mean function: ***exp(Xeβe)***, ***g*** is the gaussian distributed linear function with the conditional mean function: $\boldsymbol{g(X_L\beta_L)}$, $\boldsymbol{\Phi}$ is the probit function with the conditional mean function : $\boldsymbol{\Phi(X_p\beta_p)}$, ***h*** is the gaussian distributed linear function with the conditional mean function: $\boldsymbol{h(X_L\beta_L)}$, and $\boldsymbol{X_i}$ is vector of covariates, defining individual based socio-economic variables in the GSEM system, in that; $\boldsymbol{X_i}$ **:**c (male_d , hh_head_d , hh_income , auto_number , hhsize ,age , schooling_year, hbw_d, hbs_d), where ***c*** is the vector notation, and the abbreviations of ***e***, ***L***, and ***p*** refer respectively to exponential, linear and probit functional forms.

[2] ***estimated_motor_y*** is the estimated value from the result of negative binomial regression model (NBRM) and generated as the observed variable so as to provide the recursive relationship between "motor_y" and "motor_time". In GSEM, like the logic of 2SLS, it is possible to run the model, so as to solve the endogeneity and simultaneity problems, all of which are with non-linear relatonships within the GSEM design. Here, the ***estimated_motor_y*** is used as another observed variable for the explanation of ***motor_time***. Herein, this is realized by obtaining the fitted values for the endogeneous regressor via convenient nonlinear regression model such as count regression model on all the instruments subsequent to linear IV using this fitted value as the instrument for the count reference to the Basmann's approach (Cameron & Trivedi, 2005).

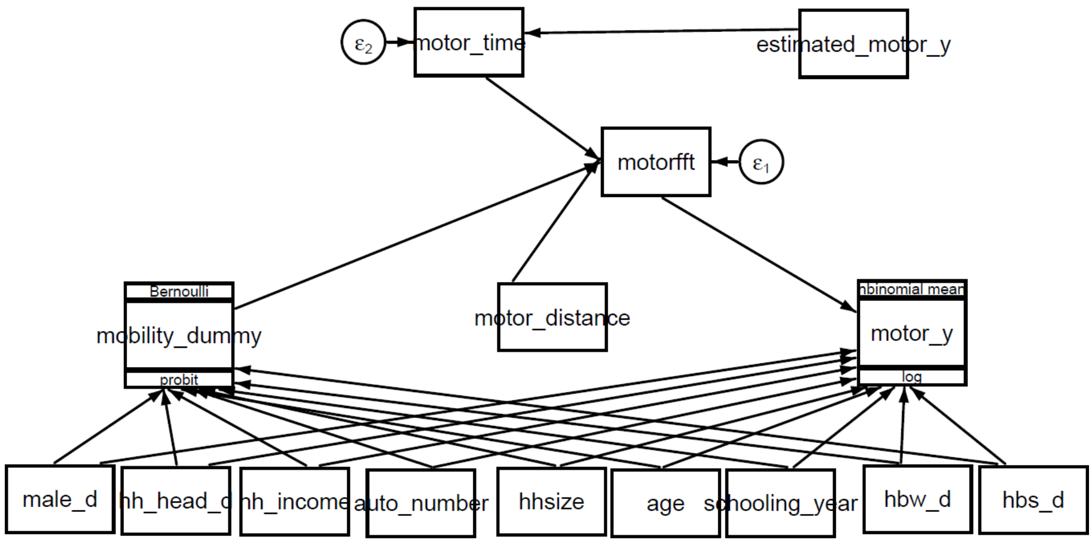

Figure 3.1 : GSEM path structure.

Within the GSEM structure, revealed in Table 3.1 and Figure 3.1, the variable called "motor_time" is the sum of the daily travel time of the related individual in that there may have occurred a kind of aggregation error, hence would misleadingly affect the probability of the individual to travel in a given day (and also affect the number of daily trips of her) positively. Hence, it has been existed the requirement to define a purely disaggregated intermediate variable, such as "motorfft (free flow travel time of the motorized trip)" so as to apprehend the inter-relationship between "daily number of trips (motor_y)" and "daily travel time (motor_time)" of the individual". Herein, the "total daily motorized travel time of the individual (motor_time)" directly affects the "free flow travel time (motorfft)", and this "motorfft" affects "daily number of motorized trips (motor_y)" of this individual, and hence, this "motor_y" also affects the "daily travel time (motor_time)" in turn. In an other words, it has been used an intermediate variable, called "motorfft" in explaining the inter-relationship between "daily number of trips of the individual (motor_y)" and "daily total motorized travel time, beared by the individual (motor_time)". This cycle of the relationships makes us perceive the inter-relationship between the variables, called "motor_y" and "motor_time" deeper, which is compatible with also the induced travel demand theory.

Furthermore, following the model structure design, the calculation of the related marginal elasticity coefficients come into our agenda, with regards to our research question. Via that regard, there have been able to be defined three main methodological frameworks in calculating the marginal elasticities after the related non-linear models (Cameron & Trivedi, 2005):

a) Estimating averages of Marginal Elasticity (ME) per each individuals.

b) Calculating ME at $\mathbf{X} = \bar{\mathbf{x}}$ (at means) , that computer programs generally carry out this.

c) Marginal Elasticity estimation at $\mathbf{X} = \mathbf{X^*}$, where $\boldsymbol{X^*}$ is a specific value that is theoretically meaningful.

With helps of our model, which has been involving a type of stochastic model nature, the estimations of the related marginal elasticity coefficients intuitively gets the form as revealed in equation 3.20;

$$\frac{\frac{1}{N} * \left(\sum_{i=1}^{i=N} \frac{\partial \hat{E}(motor\_yi|Xi,motor\_time\_i)}{\partial motor\_time\_i} * \sum_{i=1}^{i=N} \frac{\partial \hat{E}(motor\_time\_i|Xi,\ motor\_yi,\ motorfft)}{\partial Xi} \right)}{\frac{1}{N} * \left(\sum_{i=1}^{i=N} \frac{\partial \hat{E}(motor\_time\_i|Xi,motor\_time\_i,motorfft)}{\partial motor\_yi} * \sum_{i=1}^{i=N} \frac{\partial \hat{E}(motor\_yi|Xi,\ motor\_time\_i)}{\partial Xi} \right)} = \frac{\partial(\mathbf{motor\_y})}{\partial(\mathbf{motor\_s})} \quad (3.20)$$

In the light of all these views, the GSEM model seems to be able to cope with all the technical requirements; namely, "non-linearity of motor_y", "excess zero problem in motor_y", and "endogeneity of motor_time". Hence, the GSEM would be one of the most convenient models for the research framework of this thesis, producing unbiased marginal elasticity estimations.

3.2 Model Discussions: Towards 2SLS

In addition to the justification of the GSEM structure, one more model structure has been designated in this thesis. In this sense, as mentioned before, one of the main problems has been that there exists the potential case of endogeneity of the variable called daily travel time (motor_time) in modelling number of daily trips of the individuals (motor_y) when modelling disaggregated level of trip making. In this context, the endogeneity of the daily travel time would necessitate further methodological approach such as simultaneous equations model (SEM) structure with mainly refers to two stages least squares (2SLS) models in addition to the GSEM structure.

Generally, related dependent & endogeneous variables are taken into consideration as Gaussian distributed continious variables within classical SEM structure (Cameron & Trivedi, 2005, pp.101-102; Gujarati, 2003, pp.717-730). In this thesis, main dependent variable has been defined as the number of daily motorized trips of a passenger, while the daily motorized travel time of this passenger has been defined as endogeneous variable. In this context, endogeneous variable called daily motorized travel time per passenger (motor_time) is explicitly Gaussian distributed variable[3]. On the other hand, daily number of motorized trips is a kind of count variable, which would be hardly distributed Gaussian especially for the small samples. Herein, although the linear regression theory (involving 2SLS) would not require the dependent variable(s) to be normally distributed, the observations for daily motorized trip makings (motor_y) in this study comes from 194,000 observations from passengers. This makes ones assume that the dependent variable called number of daily motorized trips per passenger is asymptotically Gaussian distributed[4] according to the central limit theorem (Hill, 1998; Wilson, Voorhis, & Morgan, 2007; Hogg, Tanis, & Zimmerman, 2014). At this juncture, with helps of adding the constant ***1*** and then conducting the log transformation of the trip making data, it would be able to be coped with both the discrete nature (Figure 3.2) and with the zero observations of the variable (Lachin, et al., 2011; Lee, Guldmann & Rabenau, 2018). This explicitly highlights that these zero observations are still zero,

---

[3] *Shapiro-Wilk test* (Shapiro & Wilk, 1965) for normality of univariate called *motor_time* in STATA 15 by the code: "*swilk motor_time*", Result: W= 0.89531, V= 4929,489, z= 24.001, and Prob>z=0.00000, and for large sample size (n=194,000) & for 99 % confidence level the critical $W_{\alpha=0.01}$=0.930, which states that by 99 % confidence level, we ACCEPT H_0*:"motor_time is normally distributed"* since $W_{\alpha=0.01}= 0.930>W=0.89531$ (calculated W value is less than the critical W value). Decision: *motor_time is normally distributed with probability 0.99.*

[4] *Shapiro-Wilk (S-W) test* for normality of univariate called *motor_y* in STATA 15 by the code: "*swilk motor_y*", Result: W= 0.97837, V= 1018,617, z= 19.550, and Prob>z=0.00000, and for large sample size (n=194,000) & for 99 % confidence level the critical $W_{\alpha=0.01}$=0.930, which states that by 99 % confidence level, we REJECT H_0*:"motor_y is normally distributed"* since $W=0.97837>W_{\alpha=0.01}= 0.930$ (calculated W value is higher than the critical W value). Decision: *motor_y is not normally distributed variable with probability 0.99*. Furthermore, same S-W test for normality of *Ln_motor_y* by the code: "*swilk Ln_motor_y*", reveal that: W= 0.97006, V= 1410,001, z= 20.468, and Prob>z=0.00000, and for large sample size & for 99 % confidence level the critical $W_{\alpha=0.01}$=0.930, which explicitly states that by 99 % confidence level, we again REJECT H_0*:"Ln_motor_y is normally distributed", but this time the calculated W value is smaller for Ln_motor_y when compared to the one of motor_y.* This reveals that Ln_motor_y is nearer to the normal distribution when compared to motor_y (*Figure 3.2*), that is why within the related 2SLS, the main dependent variable has been asserted as Ln_motor_y instead of motor_y. On the other hand, although both motor_y & Ln_motor_y are not normally distributed according to the S-W test, it is explicitly able to be assumed that both variables are asymptotically normally distributed according to central limit theorem (*see* Hill, 1998; Wilson, Voorhis, & Morgan, 2007; Hogg, Tanis, & Zimmerman, 2014).

which does not give harm to these observations. Hereby, log transformation of the nonlinear variable and forming the related model structure as the combinations of these linear equations is a type of model structure, which is called two stages least squares (2SLS) model. This approach would be both pragmatic and efficient in the related estimates.

3.2.1 Designation of the 2SLS Model Structure

According to the research question of this paper, the inter-relationship between *number of daily motorized trips exhibited by an individual (motor_y)* and *daily motorized travel time beared by this individual (motor_time)* has been investigated. Such a relationship requires a kind of special care, since daily travel time (motor_time) would be a kind of endogeneous variable in explaining the main dependent variable called daily amount of trips (motor_y). In other words, it would strongly be probable that daily motorized travel time (motor_time) affects the number of daily motorized trips (motor_y), while the number of daily motorized trips also re-affects the daily motorized travel time (motor_time) reciprocally.

If the endogeneity of daily travel time (motor_time) were not on the carpet, the related ordinary least squares (OLS) model structure would take the form as revealed in equation 3.21;

motor_y = ƛ + Ƀ1*(motor_time) + Ƀ2*(male_d) + Ƀ3*(hh_head_d) +
Ƀ4*(hh_income) + Ƀ5*(auto_number) + Ƀ6*(hhsize) + Ƀ7*(age) +
Ƀ8*(schooling_year) + Ƀ9*(hbw_d)
+ Ƀ10*(hbs_d) + μ (3.21)

On the other hand, the dependent variable of this paper (motor_y) would not be a classical Gaussian distributed continious variable, which makes this model form (equation 3.21) less preferable. As mentioned before, even though the normality for the dependent variable would not strictly be required in the linear regression theory (involving 2SLS model structures), the sample size of this study (194,000 observations) explicitly meets the requirements of making the assumption that the dependent variable is normally distributed. However, logarithm transformation impelemented to dependent variables after adding constant ***k*** would overcome the zero observations in these variables and this would make the coefficient of determination of such models increase significantly (Lachin, et al., 2011; Lee, Guldmann & Rabenau, 2018). In this sense, the operation of taking the natural logarithm of the dependent variable would be implemented subsequent to adding constant ***1*** to each observation of this variable. Hereby, the zero observations of the dependent variable (motor_y) begin to behave as nonzero observations via adding ***1*** to each. In addition, the dependent variable called number of daily motorized trips would begin to behave as qualitatively more Gaussian distributed variable by such this mathematical transformation when compared to its original form (*see* Figure 3.2). Hereby, the related OLS model would take its new form as;

$$\ln(motor\_y+1) = \text{ʒ} + \text{Ъ}_1*(motor\_time) + \text{Ъ}_2*(male\_d) + \text{Ъ}_3*(hh\_head\_d) + \text{Ъ}_4*(hh\_income) + \text{Ъ}_5*(auto\_number) + \text{Ъ}_6*(hhsize) + \text{Ъ}_7*(age) + \text{Ъ}_8*(schooling\_year) + \text{Ъ}_9*(hbw\_d) + \text{Ъ}_{10}*(hbs\_d) + \text{ụ} \quad (3.22)$$

,where ***ʒ*** is the second model's regression constant, ***ụ*** is the residual of the second model structure, and ***Ъi*** is the coefficient of the variable ***i***, where i = 1,2,3,…,10, for the second structure of the OLS model.

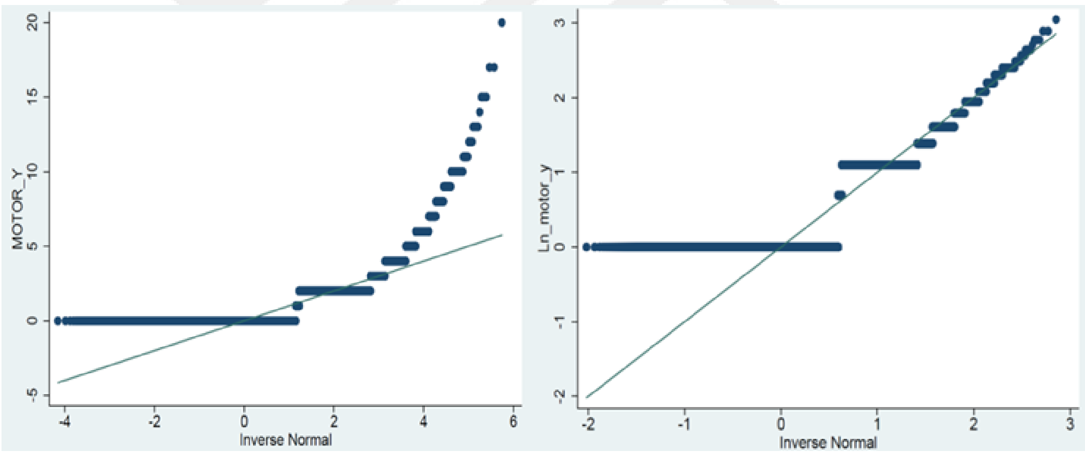

Figure 3.2 : Normal quantile plot of motor_y (in left) and Normal quantile plot of Ln((motor_y)+1) (in right).

On the other hand, both OLS model structures (*see* equations 3.21 & 3.22) would most probably produce inefficient and biased estimations for the related parameters, since the classical OLS theory assumes that the dependent variable has non-reciprocal relationship with its explanatory variables (Gujarati, 2003, pp.724-725). In other words, the *endogeneity of daily travel time (motor_time)* would not be able to be tackled with by these related ordinary least squares (OLS) model structures, which makes two stages least squares (2SLS) model structure come into considerations (Gujarati, 2003, pp. 770-774; Cameron & Trivedi, 2005; Dalgleish, et al., 2007, pp.

101-102). Before designating the 2SLS model structure, the endogeneity test is to be implemented to the variable called daily motorized travel time (motor_time) as it has been realized and exhibited in the following part.

3.2.2 Endogeneity test of the daily motorized travel time

The ***endogeneity test*** for the potentially endogeneous variable (motor_time) is required to be run before designing the related model structure. Strictly speaking, if endogeneity of *daily motorized travel time per individual* does not exist, the estimations of ordinary least squares (OLS) regression will be both consistent and efficient. On the other hand, if the related endogeneity exists, then the estimations of OLS will be biased in that there will be observed neither efficient nor consistent estimates (Kennedy, 1998; Gujarati, 2003; Cameron & Trivedi, 2005; Dalgleish, et al., 2007). In this context, the Hausman tests, namely Wu- Hausman test & Durbin-Wu- Hausman test have been realized in the study. According to the results of these tests (*see* Table 3.2), it is explicitly seen that daily motorized travel time of the individual (motor_time) is an endogeneous variable (please *see* Cameron & Trivedi, 2005, pp. 271-272 for details).

Table 3.2 : Tests of endogeneity of daily motorized travel time of an individual (motor_time) in 2SLS[5].

Tests of endogeneity	
Ho: variables are exogenous	
Durbin (score) chi2(1)	105348 (p = 0.0000)
Wu-Hausman F(1,193241)	231587 (p = 0.0000)
Result: ***variable Y_2 (motor_time) is strongly endogeneous variable (Reject Ho).***	

In pursuit of validation for the endogeneity of the variable called daily motorized travel time (motor_time), it (motor_time) has to be defined by the convenient instrumental variables so as to run the related 2SLS structure. In this sense, there have been defined two instrumental variables called *motor_distance* and *mobility_dummy*. These two instrumental variables would theoretically be justifiable in explaining the endogeneous variable called daily motorized travel time. The first instrumental variable called total daily motorized travel distance in kilometer (motor_distance) would be an ambitious variable in explaning daily motorized travel time of an individual in that the related variables together define the average daily

[5] STATA code to run this test is:" ***(estat endog motor_time)*"*.**

motorized travel speed. The second instrumental variable called *mobility_dummy* would be a kind of explicit explanation for an individual in that whether she has selected to travel at least once in a day or not. That is to say, this variable would be another ambitious variable as an instrumental variable in explaining the daily motorized travel time. In addition, these instrumental variables called *motor_distance* and *mobility_dummy* are also expected to satisfy the following conditions (see equation 3.24 below) (Cameron & Trivedi, 2005, p.100);

i. Cov (motor_distance, μ_2) = 0 and Cov (mobility_dummy, μ_2) = 0 ,

ii. Coefficient of correlations = r1(motor_distance, motor_time) and r2(mobility_dummy, motor_time) should be high as much as possible, which would empricially be between 0,7 and 1,0.

The related calculations have been realized so as to make ones check that whether the technical requirements asserted above (***i*** and ***ii***) have been satisfied or not. According to results, the coefficient of correlations between instrumental variables and daily travel time (motor_time) [6] are equal to 0.7397 and 0.7339, respectively for *motor_distance* & *mobility_dummy*. In addition, the covariance between motor_distance & mobility_dummy with model residuals are[7] 5.4e-11 and 2.3e-10, which are almost zero. As a result, the instrumental variables called *motor_distance* and *mobility_dummy* satisfy the conditions above as explicit exogeneous variables.

3.2.3 Formulation of the 2SLS Model

According the research question of the paper, the marginal effect of travel demand with respect to the one unit change in daily travel time is intended to be grasped via 2SLS model structure. In this sense, the model formulation has initially been based on the following equations (equation 3.23 and equation 3.24);

---

[6] STATA code*s are: "correlate motor_distance motor_time" and "correlate mobility_dummy motor_time",* respectively.

[7] The results of this explicitly indicates the zero covariance between the instrumental variables & model residuals (μ_2) coming from the equation 3.24**.** STATA code for these: "*correlate motor_distance μ2, covariance*" & "*correlate mobility_dummy μ_2, covariance"*. Herein, μ_2 is derived by the code: "*predict μ_2, residuals*" following OLS on motor_time (according to equation 3.24) by the code: "*regress motor_time motor_y motor_distance mobility_dummy*".

$$Ln((motor\_y)+1)) = \alpha_1 + \beta_{1.1}*(motor\_time) + \delta_{1.1}*(male\_d) + \delta_{1.2}*(hh\_head\_d)$$

$$+\delta_{1.3}*(hh\_income)$$

$$+\delta_{1.4}*(auto\_number) + \delta_{1.5}*(hhsize) + \delta_{1.6}*(age) + \delta_{1.7}*(schooling\_year) +$$

$$\delta_{1.8}*(hbw\_d) + \delta_{1.9}*(hbs\_d) + \mu_1 \qquad (3.23)$$

while the other equation;

$$motor\_time = \alpha_2 + \beta_{2.1}*Ln((motor\_y)+1) + \delta_{2.1}*(motor\_distance)$$

$$+\delta_{2.2}*(mobility\_dummy) + \mu_2 \qquad (3.24)$$

where $\boldsymbol{\mu_1}$ refers to first equation's (equation 3.23) residuals; $\boldsymbol{\mu_2}$ stands for second equation's (equation 3.24) residuals, and ***i.j*** for each coefficient represents ***i***=1,2 respectively for first & second equations of the 2SLS system asserted above (equations 3.23 & 3.24), ***j***=1,2,.. stands for each explanatory variable $\boldsymbol{\alpha_1}$ & $\boldsymbol{\alpha_2}$ represents the related regression constants. In equations 3.23 and 3.24, *Ln((motor_y)+1))* is the main dependent variable and *motor_time* is the endogeneous variable, while the remainings are the exogeneous variables. In addition, *motor_distance* & *mobility_dummy are* the instrumental variables in modelling the endogeneous variable called motor_time.

3.2.4 Identification of the 2SLS Model

There are two basic rules for checking any simultaneous equations system in that whether it is identified or not. The first rule is called *order rule* while the other is called *rank rule*. To begin with, according to the order condition, it is intuitively stated that the number of instrumental variables used in the 2SLS model has to be at least equal to the number of endogeneous variables (Gujarati, 2003, pp. 739-746; Cameron & Trivedi, 2005, p.100; Dalgleish, et al., 2007). In the 2SLS model, there have been defined two instrumental variables, namely motor_distance and mobility_dummy, while there have been asserted two endogeneous variables called log transformed motor_y and motor_time. In other words, the number of instrumental variables and the number of endogeneous variables are exactly same in 2SLS model, which would partly reveal that this 2SLS system is a canditate of being just identified according to the *order condition* (Dalgleish et al., 2007). On the other hand, with regards to further technical check for the order condition, our simultaneous equations system seems to be *overidentified*. In this sense, let us code the number of predetermined variables in the model (equation 3.23 & 3.24) as ***K***. Then, this ***K*** is explicitly equal to ***11***, because it includes the intercepts of the model. Moreover, let us code the number of predetermined variables as ***k***, which excludes the intercepts in the given model (equations 3.23 & 3.24). Then, this ***k*** is equal to ***9***. Lastly, let us code the number of endogeneous variables in 2SLS system (equations 3.23 & 3.24) as ***m***, which is equal to ***2***. Afterwards, the related two calculations with regards to ***K***, ***k***, and ***m*** reveal that (equation 3.25);

$$K\text{-}k = 11 - 9 = 2 \text{ and } m\text{-}1 = 2\text{-}1 = 1, \text{ so } K\text{-}k = 2 > m\text{-}1 = 1 \qquad (3.25)$$

With respect to the results shown in equation 3.25, the equation 3.23 of 2SLS system seems explicitly to be overidentified in accordance with the order condition (Gujarati, 2003, p.748). Furthermore, for the other equation of the 2SLS system (equation 3.24), the related calculations reveal that (equation 3.26);

$$K = 11,\ k = 2,\ m = 2. \text{ Then, } K\text{-}k = 11\text{-}2 = 9 > m\text{-}1 = 1 \qquad (3.26)$$

So, with regards to the equation 3.26, equation 3.24 of 2SLS is also overidentified according to the order rule. As a result, it is explicitly able to be stated that the two stages least squares (2SLS) regression equations system would be overidentified according to the order rule.

On the other hand, the order rule is necessary but not sufficient rule to be sure that the related equations system is identified or not. In this context, the second control for identification, namely ***rank condition*** is also required to be checked (Gujarati, 2003, pp. 750-753; Dalgleish, et al., 2007). Herein, small coefficient of correlations (equal to approximately zero) of the instrumental variables with the remaining explanatory variables of 2SLS are required (Table 3.3).

Table 3.3 : Partial and semipartial correlations of instrumental variables with the exogeneous variables (X_i).

Exogeneous Variable (X_i)	**Partial Corr.** r (X_i, motor_distance)	**Partial Corr.** r (X_i, mobility_dummy)
male_d	0.0366	0.0198
hh_head_d	0.0351	0.0419
hh_income	0.0199	0.0260
auto_number	0.0815	0.0867
hhsize	-0.0257	-0.0377
age	0.0152	0.0112
schooling_year	0.0927	0.1263
hbw_d	0.2976	0.4626
hbs_d	0.1191	0.2223

As it is seen from Table 3.3, the related partial correlations of the instrumental variables with the related exogeneous variables are approximately zero, which signals the satisfaction of the full rank condition in addition to the order condition. In the light of both order and rank rules, the simultaneous equations system called 2SLS is able to be stated as an explicit type of overidentified system.

Reference to all these indications, 2SLS model structure (other than GSEM structure) would also be another convenient model for the research framework of this thesis. Because, it seems to be able to tackle with all the technical requirements, namely, non-linearity of daily trip makings (motor_y), excess zero observations in daily travel times, and endogeneity of daily travel times. Furthermore, the designated 2SLS of the thesis proves that it is a kind of over-identified system with explicit endogeneity of daily travel time (motor_y) and convenient exogeneous instrumental variables for this endogeneity.

3.3 Towards Selection of the Convenient Models: GSEM & 2SLS

In the light of the discussions on model structures, GSEM and 2SLS model structures seem technically to satisfy all the technical requirements, namely non-linearity of daily number of motorized trips, excess zero problem in daily number of motorized trips, and endogeneity of daily motorized travel time. Firstly, these model structures are able to deal with the non-linear nature of number of daily motorized trips (motor_y). In this sense, GSEM is able to deal with the related nonlinearity of daily number of motorized trips by the indication of the NBRM structure into the GSEM design, while 2SLS is able to tackle with the same nonlinearity by applying natural logarithm (ln) transformation to this variable (motor_y). Secondly, both GSEM and 2SLS are able to cope with the potential problem of excess zero observations for the variable called number of daily motorized trips. Herein, it is able to be tackled with by the indication of the condition in the GSEM design asserting that the individual exhibits at least one motorized trips in the day (*mobility_dummy==1*). Hereby, it would be guaranteed that the potential zero counts in daily number of trips are automatically eliminated in such this GSEM design. On the other hand, the same excess zero observations in daily number of motorized trips have been able to be dealt with in 2SLS by adding constant ***1*** to each observation of this variable and then taking natural logarithm of this. As mentioned before, this mathematical operation (applied in 2SLS) makes zero observations of daily trips non-zero, which technically does not give any harm to the related data. Thirdly, both GSEM and 2SLS structures are able to tackle with the endogeneity of daily motorized travel time of an individual (motor_time) in modelling the number of daily motorized trips (motor_y). In GSEM, this has been satisfied via the indication of free motorized flow time (motorfft) as the theoretically justified instrumental variable in modelling daily motorized travel time (motor_time) seperately with the same socio-economic characteristics & the related dummy variables, as used in modelling number of daily motorized trips. In this sense, firstly the number of daily motorized trips (motor_y) has been modelled via the related socio-economic characteristics & the related dummy variables, then this derived estimated value of the number of daily motorized trips (estimated_motor_y) has been used in modelling the dependent variable called daily motorized travel time (motor_time). This has made the GSEM structure be able to cope with the endogeneity of daily motorized travel time (Table 3.4). On the other hand, in 2SLS, the same endogeneity of daily motorized travel time per passenger (motor_time) has explicitly been able to be tackled with the usage of instrumental variables for this endogeneous variable. As a reminder, the endogeneous variable, namely daily motorized travel time, has been modelled by the explicit exogeneous instruments called mobility_dummy and motor_distance.

Table 3.4: Comparison of the model structures [8].

Model	Copes with the non-linear nature of "motor_y"	Copes with the assumption of equidispersion or "motor_y"	Copes with the "excess zero problem" for "motor_y"	Copes with the "endogeneity" of "motor_s" in modelling "motor_y"
Poisson Regression Model (PRM)	✔	X	X	X
Negative Binomial Regression Model (NBRM)	✔	✔	X	X
Sample Selection Model (SSM)	✔	✔	✔	X

[8] In the Table 3.4, (✔): the asserted property is satisfied; (**X**): the asserted property is not satisfied.

Table 3.4 (continued) : Comparison of the model structures.

Model	**Copes with the non-linear nature of "motor_y"**	**Copes with the assumption of equidispersion or "motor_y"**	**Copes with the "excess zero problem" for "motor_y"**	**Copes with the "endogeneity" of "motor_s" in modelling "motor_y"**
Sample Selection Poisson Regression Model (SSPRM)	✔	X	✔	X
Instrumental Variable Poisson Regression Model (IVPRM)	✔	X	X	✔
Instrumental Variable Zero Truncated Poisson Regression Model (IVZTPRM)	✔	X	✔	✔
Generalized Simultaneous Equations Model (GSEM)	✔	✔	✔	✔
2SLS with *ln(y+1)* transformation for the dependent variable *y*	✔	✔	✔	✔

To summarize, both GSEM and 2SLS model structures would be the optional convenient models in modelling induced motorized passenger mobility with refers to the inter-relationship between daily number of motorized trips and daily motorized travel time of the urban passengers.

4. MODEL RESULTS

As stated earlier, to begin with, the count models are estimated for motorized trips in this thesis. The results of the count models for motorized daily trips are given in Table 4.1 below. First of all, a linear model is estimated as a benchmark model. Later, a PRM, a NBRM, a SSM and a SSPRM are estimated. As it can be seen from Table 4.1, likelihood ratio as specified in equation 3.7, are 50 and 5926. They are well above the critical chi-squared table value of 5.99 indicating overdispersion and favor to the NBRM for both models. Furthermore, the estimated distribution (scale) variable, θ, is significant in both models confirming an overdispersion. Concerning SSM and SSPRM, log likelihood ratios are well above 5.99 indicating all models outperforms their restricted forms.

Table 4.1 : Count models' estimations for motorized daily trips.

	LRM	t	PRM	z	NBRM	z	SSM	z	SSPRM	z
intercept	0,032	4,00	-1,462	-98,13	-1,579	-92,36	1,6503	103,7	0,534	34,1
motor_time	-0,01	-421,5	-0,005	-284,2	-0,008	-810,0	-0,003	-93,3	-0,0014	-45,4
male_d	0,013	3,20	0,093	13,24	0,08	10,08	0,0365	5,21	0,018	2,59
hh_head_d	0,090	19,52	0,14	18,55	0,135	15,89	0,1087	14,3	0,0513	6,85
hh_income	0,013	10,40	0,007	5,23	0,009	5,00	7,37e-06	4,57	2,69e-06	2,04
auto_n	0,148	53,69	0,167	40,61	0,185	38,54	0,1525	35,6	0,0695	16,9
hhsize	-0,01	-11,06	-0,017	-10,12	-0,019	-10,47	0,003	1,75	0,0016	0,92
age	-0,0003	-2,79	-0,002	-8,50	-0,002	-6,00	-0,002	-7,77	-0,001	-3,66
yr_schooling	0,014	29,61	0,031	44,57	0,027	30,44	0,0069	9,06	0,0036	4,85
worked	0,493	122,51	0,962	139,42	0,90	120,29	-0,0367	-5,44	-0,015	-2,25
student	0,361	49,45	0,854	79,04	0,795	64,14	-0,073	-6,58	-0,032	-2,91
Dispersion					0,177	51,94				
Lambda							-0,16			
Log L			-1201		-11710		-78580,5		-193614,7	
R^2	0,63									
Wild chi2(10)							10942,16		2591,8	
LR			**50**		**5926**					

Initial observation for the results of the parameters is that all variables have significant relationship with the total daily motorized trip frequencies, except only for household size. The results partially confirmed the initial expectations in LRM, PRM, and NBRM. However, when the model is designated towards SSPRM, the parameters converged to initial expectations. When we look at the result by SSPRM, it is seen that all variables affect total daily trip making positively. However, being student, being worked, daily motorized travel time, and age are the significant variables affecting daily number of motorized trips negatively. It should be noted that travel time is the most significant (negative as expected) variable among all. According to the results, all models significantly justify the fact that travel time refers to a kind of explicit disutility for individuals' propensities to mobilize in a day.

Beyond travel time, age, being a student, and being a worked show negative correlations with daily number of motorized trips, and this result is consistent since students' ages are relatively younger. In addition, being worked would refer purely to the obligatory trips called home based work trips, which would be able to minimize the other types of motorized passenger mobility such as recreation, shopping, and non-home based others in the case of İstanbul. The precursor factors -affecting daily motorized mobility of the individuals negatively- are being student, being worked, travel time, and age with coefficients 0.032, 0.015, 0.0014, and 0.001, respectively according to the results of SSPRM. On the other hand, the prominent factors inducing the urban motorized passenger mobility positively are number of automobiles owned by the household and the dummy variable asserting that whether the individual is household head or not.

Even though we modelled the characteristics of trip making, the core of our thesis refers to calculating the magnitudes of marginal effect of these variables on total number of daily motorized trip making as a proxy for the measurement of induced urban motorized passenger mobility, and these estimations are presented in Table 4.2.

Table 4.2 : Count models' elasticity estimations for motorized daily trips.

Variable	LRM	PRM	NBRM	SSM	SSPRM
motor_time	-0,0109	-0,0029	-0,0040	-0,0033057	-0,0026353
male_d	0,0134	0,0502	0,0398	0,0365087	0,0339419
hh_head_d	0,0904	0,0753	0,0675	0,1087435	0,0968455
hh_income	0,0131	0,0037	0,0042	7,37e-06	5,08e-06
auto_n	0,1482	0,0901	0,0924	0,1524823	0,131163
hh_size	-0,0101	-0,0093	-0,0099	0,0029819	0,0029253
age	-0,0003	-0,0009	-0,0009	-0,0019955	-0,0017546
yr_schooling	0,0137	0,0169	0,0137	0,0068702	0,0068009
worked	0,4925	0,5206	0,4507	-0,0366814	-0,028163
student	0,3605	0,4620	0,3973	-0,0729874	-0,06028

It should be noted that the key variable measuring induced urban passenger mobility is systematically over-estimated in LRM, PRM, and NBRM. When the bias -coming from discrete dependent variable- was eliminated by the count models, the elasticity of travel time lowers to around 3 per thousand. Elimination of endogeneity, further, decreases the core elasticity around to 3.3 per thousand in SSM and 2.6 in SSPRM for motorized trips.

If we should take SSPRM as reference, it is possible to say that one percent decrease in travel time makes the daily motorized trip makings increase by 0.27 percent with refers to the fact that people are more sensitive to the motorized travel time changes. Car ownership and dummy variable of being household head have the highest inducement effects on the daily amount of motorized trips with the marginal elasticity coefficients that are equal to 0.13, and 0.096, respectively. That is to say, one more automobile owned by the household induces 0.13 more motorized trips for each member of that household. In addition, household head has a propensity to exhibit 0.096 more motorized trips in a day when compared to others in a household. Another important finding is that sex also has significantly high inducement effects on trips with the marginal elasticity coefficient that is equal to 0.034. Herein, it is able to be stated that males exhibit 0.034 more daily motorized trips when compared to the females in the case of İstanbul.

Even though these findings confirm the existence of significant induced travel demand effect, as confirmed by the other studies in the literature of urban passenger mobility, the magnitude of it cannot be said overwhelming the magnitude of other trip determinants according to these model results. In any case, the magnitude of induced urban passenger mobility demand effect remains modest in comparison to the combined effects of individual socio-economical characteristics. For example, the combined effects of number of auto and household income have higher generative effect than the generative impact of travel time saving.

These travel time elasticities are higher than those of estimated by Barr (2000), which was -0,44 on average. A reasonable explanation for such a result could be that the Turkish cities are more compact than the western cities. Another possible reason could be that the transportation cost is higher since Turkey heavily taxes petroleum. In either case, people would be more sensitive to travel. Furthermore, there are several differences between Barr’s study and our approach. Barr used household level car trip data while our data include all trips and motorized trips at individual level. On the other hand, PRM and NBRM have only been able to deal with the non-linear nature of the daily number of motorized trips, while SSM and SSPRM have been able to deal with only non-linear nature of daily amount of motorized trips and with excess amount of zero observations in daily trips. In other words, none of these model structures asserted up to here, namely LRM, PRM, NBRM, SSM, and SSPRM has been able to cope with the endogeneous nature of daily travel time. In this sense, the concept of using instrumental variables in modelling endogeneous variable called daily travel time comes into considerations. Herein, Instrumental Variables Poisson Regression Model (IVPRM) and Instrumental Variable Zero Truncated Poisson Regression Model (IVZTPRM) are such kinds of model structures (Cameron & Trivedi, 2005), as indicated in the model discussions part of the thesis in detail. The results of these models have been exhibited in the Table 4.3.

Table 4.3 : IVPRM (with two[9] & three [10] instruments) & IVZTPRM [11] results.

Variable	IVPRM (with 2 instruments)		IVPRM (with 3 instruments)		IVZTPRM (with 3 instruments)	
	Coef.	z	Coef.	z	Coef.	z
motor_time	.005372	85.13	.0103077	96.30	.0016434	39.79
male_d	.1293632	14.90	.1209033	1.81	.0130574	4.41
hh_head_d	.1912472	18.71	.025207	0.33	.0391598	10.64
hh_income	.0000129	5.89	-7.24e-06	-0.31	2.79e-06	2.45
auto_number	.2359158	40.17	.0916965	2.34	.0676029	28.75
hhsize	-.0308132	-13.14	-.0347579	-1.94	.0008274	1.02
age	-.001912	-5.25	-.0019274	-0.77	-.000717	-5.68
tahsil_y	.0382773	35.32	.0445754	5.61	.003114	8.23
hbw_d	1.273507	123.14	1.395202	28.26	-.0096283	-2.65
hbs_d	1.074179	59.86	1.150497	13.86	-.0267135	-5.01
_cons	-1.686658	-72.17	-2.740163	-18.50	.5053046	61.34

IVPRM structure has been designated by two forms. In the first, two instruments,

[9] related STATA code is:
.ivpoisson gmm motor_y male_d hh_head_d hh_income auto_number hhsize age schooling_year hbw_d hbs_d (motor_time = motor_distance motorfft), vce(robust)
Instrumented: motor_time
Instruments: male_d hh_head_d hh_income auto_number hhsize age schooling_year hbw_d hbs_d motor_distance motorfft
[10] the related STATA code is:
.ivpoisson gmm motor_y male_d hh_head_d hh_income auto_number hhsize age schooling_year hbw_d hbs_d (motor_time = motor_distance motorfft mobility_dummy), vce(robust)
Instrumented: motor_time
Instruments: male_d hh_head_d hh_income auto_number hhsize age schooling_year hbw_d hbs_d motor_distance motorfft mobility_dummy
[11] related STATA code is:
.by mobility_dummy, sort: ivpoisson gmm motor_y male_d hh_head_d hh_income auto_number hhsize age schooling_year hbw_d hbs_d (motor_time = motor_distance motorfft mobility_dummy) if mobility_dummy==1, vce(robust)
Instrumented: motor_time
Instruments: male_d hh_head_d hh_income auto_number hhsize age schooling_year hbw_d hbs_d motor_distance motorfft mobility_dummy

namely motorized travel distance (motor_distance) and motorized free flow travel time (motorfft) have been asserted. In the second, the dummy variable indicating whether the individual has carried out at least one motorized trip(s) in the given day or not (mobility_dummy) has also been asserted in addition the previous two instruments, so this form is called IVPRM with three instruments.

According to the results, the coefficients of all variables in IVPRM with two instruments seem statistically significant, while there have been observed three insignificant coefficient estimations of IVPRM with three instruments with refers to the variables called hh_head_d, hh_income, and age according to their z values. In IVZTPRM, only *hhsize* seems statistically insignificant. The prominent positively affecting factors of trip making observations are auto_number, hh_head_d, and male_d according to the results of IVPRM (with both two and three instruments) and IVZTPRM. On the other hand, both IVPRM forms make *hbw_d* and *hbs_d* come into prominence as the leading factors enhancing the level of daily motorized trip makings, while IVZTRPM exhibits antipodal effect (negatively affecting prominent factor). This would have been caused by the elimination of zero counts from the daily amount of motorized trips in the IVZTPRM.

Lastly, as revealed in the Table 4.4, according to the estimations of marginal effect of daily motorized travel time (motor_time) on daily number of motorized trips (motor_y), the highest value comes from the IVZTPRM with refers to the coefficient 1.88. In other words, according to the result of IVZTPRM, one unit decrease in travel time induces approximately 2 more motorized trips per passenger, while the related coefficients are only 0.7, and 0.28 respectively for IVPRM with two and IVPRM with three instruments. These elasticities are much higher than the ones of all the previous model structures covered up to here.

Table 4.4 : Marginal elasticity calculations of IVPRM & IVZTPRM models for *motor_time*.

margins for "motor_time" (at their means)	∂y/∂x (response variable motor_time)			[95% Confidence Interval]	
	Margin	Std. Err.	z		
IVPRM WITH ENDOG 2 INST/ motor_y	**.6972516**	.0032443	214.92	.690893	.703610
IVPRM WITH ENDOG 3 INST/ motor_y	**.2771952**	.0023361	118.66	.272616	.281773
IVZTPRM /motor_y	**1.877319**	.0051726	362.94	186.718	1.887.45

On the other hand, neither the two forms of IVPRM nor IVZTPRM satisfies all technical requirements of our research question. On one side, even though IVPRM is able to deal with the non-linearity of daily amount of motorized trips, and endogeneity of daily travel time, it is not able to meet the requirement called coping with excess amount of zero observations in daily trip makings. On the other side, although IVZTPRM seems to be satisfying all the technical requirements, coping with non-linearity of daily number of trips, excess amount of zero observations, and endogeneity of daily travel time, it still does not belong to the most convenient model structure of this research framework, since it assumes that the mean and variance of the main dependent variable (motor_y) is equal to each other (assumption of equidispersion) due to the integration of poisson regression model structure in it (*see* equation 4.1).

In the light of all these views, it has been required further model structures, which are able to cope with non-linearity of dependent variable, excess amount of zero counts in the dependent variable, endogeneity of travel time, and eliminating the assumption of equidispersion with refers to the mean and variance of the dependent variable. In this sense, subsequent to the all model discussions exhibited up to here, GSEM and 2SLS have been asserted as the most convenient optional models for the empirical research of the thesis, as mentioned in the previous section. In the following parts, the related results of these two models with their primary indications have been exhibited.

4.1 Generalized Simultaneous Equations Model (GSEM) Results

According to the GSEM results, to begin with, the Negative Binomial Regression Model (NBRM) structure has overweighted Poisson Regression Model (PRM) with reference to modelling the dependent variable, namely daily number of motorized trips per passenger (motor_y). In this sense, according to the results of PRM & NBRM, the assumption of *equidispersion* of the mean and variance of the variable (motor_y) has been failed due to the related calculation of equation 4.1;

$$LR = 2 \times (\text{Ln } L_{NBRM} - \text{Ln } L_{PRM})^{12} \quad (4.1)$$

Here, "LR (likelihood ratio)" exhibits a type of "chi-square distribution" with two degrees of freedom (d.f.) and any calculated value, that is greater than the critiqual value for the related chi-square distribution, favors NBRM[13]. This asserts that the assumption of "equidispersion" of the variable "motor_y" has been failed (see Table 4.5).

Table 4.5 : PRM & NBRM results.

PRM	**Coef.**	**z**	**NBRM**	**Coef.**	**z**
male_d	**.**0817799	9.85	**male_d**	.029871	4.20
hh_head_d	.1405562	14.76	**hh_head_d**	.1075885	14.02
hh_income	6.79e-06	4.30	**hh_income**	7.80e-06	4.05
auto_number	.1617351	30.09	**auto_number**	.1146227	23.28
hhsize	-.0137819	-6.60	**hhsize**	-.0138107	-8.00
age	-.0020655	-6.50	**age**	-.0020207	-7.77
schooling_year	.0311067	30.95	**schooling_year**	.0240298	30.12
motor_time	.0054185	74.77	**motor_time**	.0085438	95.37
hbw_d	.9484354	96.50	**hbw_d**	.7390091	84.62
hbs_d	.8404261	56.93	**hbs_d**	.6927652	58.96
motorfft	.1482236	7.18	**motorfft**	1.516752	36.38
_cons	-1.473417	-73.06	**_cons**	-1.712916	-104.77
			/lnalpha =	-1.734055	
			alpha =	.1765671	
PRM			**NBRM**		
Number of obs		193,253	**Number of obs**		193,253
Wald chi2(10)		50232.86	**Wald chi2(10)**		112423.72
Prob > chi2		0.0000	**Prob > chi2**		0.0000
Pseudo R2		0.2717	**Pseudo R2**		0.2367
Log pseudolikelihood (Ln L_{PRM})= -184233.85			**Log pseudolikelihood (Ln L_{NBRM})**= -181271.32		

Moreover, all the related model coefficients are statistically significant according to the calculated z values, and these models (PRM & NBRM) have been asserted due to the justification for the selection of NBRM structure in modelling the dependent variable called daily number of motorized trips (motor_y) within the GSEM Path Analysis. On the other hand, these two single equation models (PRM & NBRM) are only able to cope with the non-linearity of the count variable, called "motor_y". The other respective conditions, namely excess zero problem in daily amount of motorized trips (motor_y) and the problem of endogeneity of daily motorized travel

---

[12] According to equation 4.1, *2*(-181271.32 - (-184233.85))=* ***5925.06*** (*LR test of alpha=0: chibar2(01) = 5925.05*), which is much greater than the critiqual chi-square value for 2 degree of freedom (d.f.) that is equal to **0.10** for 95 % confidence level (*Ln* L_{NBRM}*=-181271.32* and *Ln* L_{PRM} *= -184233.85* according to the related model results as indicated in *Table 4.5*).

[13] That is why for the final model specification asserted in GSEM path structure, the functional form of the variable "motor_y" has been indicated as in the functional form of "negative binomial mean distribution, with the link log".

time (motor_time) have not able to be tackled with via the PRM & NBRM models. In this context, the further model, namely GSEM Path Analysis, has been developed and exhibited in this thesis.

According to the results of GSEM (Table 4.6), all the calculated coefficients of the explanatory variables are statistically significant. To begin with, the coefficients of ***hhsize*** is negative in explaining the dependent variable of *mobility_dummy*. On the other hand, the coefficient of *age* is positive in explaining the same dependent variable *mobility_dummy* in GSEM. In other words, as the individual gets older, interestingly, his/her probability to travel in any given day increases according to the results of GSEM. Secondly, within the equation of *motor_y*, coefficients of both *hhsize* and *age* are negative. Thirdly, as it is expected, the coefficient of *motor_time* is negative within the model of the dependent variable *motorfft*. This result is not surprising, since the term *motorfft* refers to *motorized free flow travel time of an individual traveling unit* in the case that he/she is the only motorized traveler with his/her related free flow speed, which is, inversely related with the actual daily travel time of the individual (Table 4.6).

Table 4.6 : GSEM results.

Dependent Variable	Variable	Coef.	Std. Err.	z
mobility_dummy	**male_d**	.0876364	.008781	9.98
	hh_head_d	.1824174	.0098624	18.50
	hh_income	.0000381	5.01e-06	7.60
	auto_number	.2149051	.0060338	35.62
	hhsize	-.0340989	.0020449	-16.68
	age	.000905	.0002754	3.29
	schooling_year	.0537412	.0010146	52.97
	hbw_d	1.448814	.0078503	184.55
	hbs_d	1.230701	.0139649	88.13
	_cons	-1.498341	.0181383	-82.61
motor_y	**male_d**	.094116	.0078328	12.02
	hh_head_d	.1502352	.0086166	17.44
	hh_income	.0000143	2.67e-06	5.35
	auto_number	.1087255	.0053448	20.34
	hhsize	-.0162595	.0019109	-8.51
	age	-.0010601	.0002899	-3.66
	schooling_year	.0430605	.0008928	48.23
	hbw_d	1.052222	.0101024	104.16
	hbs_d	1.042079	.0122483	85.08
	motorfft	2.269169	.0490115	46.30
	_cons	-1.65532	.0177808	-93.10
motorfft	**motor_time**	-.0023983	.0000523	-45.89
	motor_distance	.0102411	.000213	48.07
	mobility_dummy	.2824551	.0030349	93.07
	_cons	-7.54e-18	1.28e-19	-58.70

Table 4.6 (continued) : GSEM results.

Dependent Variable	Variable	Coef.	Std. Err.	z
motor_time	**estimated_motor_y**	7.415484	.4303755	17.23
	_cons	30.81517	.3910488	78.80
motor_y	**lnalpha**	-.8130799	.030545	
	var(e.motorfft)	.0407052	.0113116	
	var(e.motor_time)	3392.502	41.8371	
		Log pseudolikelihood = -1314459.5		

Subsequent to the exhibition of GSEM esitmations, the estimation of the marginal elasticity coefficients of each factor has also been exhibited (Table 4.7). The negative marginal elasticity coefficients of ***motor_y*** with respect to ***motor_time*** vary from 0.00016 to 0.0025 [14]**.** In other words, one unit decrease in daily travel time of an individual, induces an average amount of 0.000953 more trips (see Table 4.7). To illustrate, specifically, if the individual is male (male_d=1), is household head (hh_head_d=1), is carrying out a type of home-based work trip (hbw_d=1) or home-based school trip (hbs_d=1), and selects to travel at least once in a given day (mobility_dummy=1); then one unit[15] decrease in travel time induces an amount of 0.0025 more trips for the individual per day. At that point, the *mobility_dummy* is fixed to be equal to one[16], asserting that, the number of daily trips (motor_y) is observed if the related individuals select to exhibit at least one trip in a day. The remaining dummy variables, namely *male_d, hh_head_d, hbw_d,* and *hbs_d* take the binary values of *0* or *1*, which makes *sixteen* [17] possible combinations for the marginal elasticity calculations for each factor. Among these combinations of these related dummy variables, the highest marginal elasticity coefficients for the related factors affecting *motor_y* are observed when all these dummy variables are equal to one. In other words, in the case of the passenger is male, household head, and the related trips are the *obligatory* ones [18], there exist highest amounts of marginal elasticity coefficients with reference to the factors affecting *daily number of the motorized trips of an individual* (motor_y). According to the results of marginal elasticity calculations, via the stated case of the dummy variables [19], the leading

[14] See Table 4.7.
[15] One unit refers to one minute for travel time.
[16] This has also been indicated so as to cope with the *excess zero problem* for the main dependent variable, called *motor_y*.
[17] 2*2*2*2 = 2^4 =16 different combinations. Here, the number *2* refers to the possible outcomes for each dummy variable, respectively, 0 or 1.
[18] such as home-based work (hbw) trips and/or home based school (hbs) trips as the obligatory trips.
[19] that all these dummy variables are equal to 1.

factors affecting *motor_y*, are *male_d, hh_head_d,* and *auto_number*, with the marginal elasticity coefficients, respectively, *0.1597, 0.256,* and *0.1884*. In other words, if the individual is being male and household head, then the number of daily trips of him/her increase respectively by 0.1597 and 0.256 more trips in a day. Similarly, as it is expected, a one unit increase in the number of automobiles owned by the household (*auto_number*) causes 0.1884 more trips per person in any given day in İstanbul. Furthermore, the negatively signed marginal elasticity coefficients, with refers to the factors, namely *hhsize* and *age*, are equal to *-0.028* and *-0.0017*. In this manner, one unit increase in the household size and the one unit increase in age of the individual make the daily number of trips decrease, respectively, by 0.028 trips and by 0.0017 trips. Lastly, the household disposable income (*hh_income*) is able to be stated as the weakest factor, according to the marginal elasticity coefficient, which is equal to *0.000025*, with relates to its effect on *motor_y*.

Table 4.7 : Marginal elasticity calculation results for GSEM.

(male_d h_head_d hbw_d hbs_d mot_y_s)	(dydx)(Δx/Δy) motor_time)	(dydx)(Δx/Δy) [20](male_d)	(dydx)(Δx/Δy) (hh_head_d)	(dydx)(Δx/Δy) (hh_income)	(dydx)(Δx/Δy) (auto_number)	(dydx)(Δx/Δy) (hhsize)	(dydx)(Δx/Δy) (age)	(dydx)(Δx/Δy) (schooling_year)	(dydx)(Δx/Δy) (hbw_d)	(dydx)(Δx/Δy) (hbs_d)
(0 0 0 0 1)	**-.0001616**	.0115671	.019187	2.18e-06	.0152942	-.0023205	-.0000981	.0055241	.1372912	.1324988
(0 0 0 1 1)	**-.0005874**	.0430542	.071847	8.36e-06	.0580784	-.0088293	-.0003461	.0206997	.5157669	.4958144
(0 0 1 0 1)	**-.0006166**	.0442775	.0735092	8.38e-06	.058715	-.0089111	-.0003728	.0211662	.5262388	.5075802
(0 0 1 1 1)	**-.0019541**	.1257064	.2025372	.0000202	.1502313	.022553	-.0013324	.0581155	1.426.084	1.403
(0 1 0 0 1)	**-.0001938**	.0141393	.0235682	2.73e-06	.0190015	-.0028876	-.0001149	.0067893	.1690849	.1626644
(0 1 0 1 1)	**-.0007051**	.0508195	.0844473	9.66e-06	.0675962	-.0102621	-.0004244	.0243182	.6048409	.5830472
(0 1 1 0 1)	**-.0007366**	.0518862	.0857164	9.57e-06	.0676694	-.010253	-.0004558	.0246671	.6119836	.5922027
(0 1 1 1 1)	**-.002283**	.1455893	.2339743	.0000231	.1723947	-.0258535	-.0015698	.0671157	164.505	1.622
(1 0 0 0 1)	**-.0001802**	.0130192	.0216488	2.48e-06	.0173562	-.0026355	-.0001081	.0062347	.1551127	.1494578
(1 0 0 1 1)	**-.0006558**	.0476989	.0794443	9.17e-06	.0639334	-.0097133	-.0003903	.0228835	.5697138	.5483626
(1 0 1 0 1)	**-.0006868**	.0488749	.0809531	9.14e-06	.0643072	-.0097522	-.0004199	.0233034	.578798	.5591276
(1 0 1 1 1)	**-.0021529**	.137881	.2218643	.000022	.16401	-.0246086	-.0014743	.0636514	1.561.018	1.537
(1 1 0 0 1)	**-.0002164**	.0159183	.0265857	3.10e-06	.0215316	-.0032742	-.000127	.0076603	.1909343	.1834501
(1 1 0 1 1)	**-.0007858**	.0561266	.0930538	.0000105	.0740876	-.0112391	-.0004782	.0267896	.6656621	.6426346
(1 1 1 0 1)	**-.0008188**	.0571366	.0941575	.0000104	.0738952	-.0111868	-.0005122	.0270886	.6713451	.650704
(1 1 1 1 1)	**-.0025132**	.1597193	.2564198	.0000251	.1884241	-.0282456	-.0017339	.0735453	1.802	1.777

[20] **(dydx)(Δx/Δy)** (response variable) gives marginal elasticity of the response variable (all the remaining covariates are at their mean values) on *motor_y* . Here, there have been calculated *sixteen* marginal elasticities for each, because there are four dummy variables, taking values "0" or "1" (only mobility_dummy =1 constant due to the nature of the zero truncation for the daily trip number(s)) in total that derives sixteen ($2^4 = 16$) different combinations.

4.2 Two Stages Least Squares (2SLS) Regression Model Results

According to 2SLS model results, it is explicitly seen that nearly all the explanatory variables are statistically significant at 99 % confidence interval according to their z values. In addition, the Wald test of two stages least squares (2SLS) model explicitly refers to the fact that the related estimations of all coefficients are different from zero and the coefficient of determination is 0.5228 (*see* Table 4.8).

Table 4.8 : 2SLS model results.

Number of observations	193,253
Wald chi2(10)	341429.18
Prob > chi2	0.0000
R-squared	0.5228
Root MSE	.38162

Variable	Coefficient	z	2SLS Marginal Effect Coefficients
motor_time	-.0086859	-453.42	-8.761673e-03 (-11.4 %)
male_d	-.0084113	-3.57	-8.340847e-03 (-1.1 %)
hh_head_d	.0193036	7.40	1.967985e-02 (2.5 %)
hh_income	2.27e-06	3.20	2.27001e-06 (-.00029 %)
auto_number	.050989	32.77	5.365630e-02 (6.7 %)
hhsize	-.0034356	-6.66	-3.423817e-03 (-.043 %)
age	-.0003358	-4.76	-3.356873e-04 (-.042 %)
schooling_year	.0008838	3.39	8.845814e-04 (.011%)
hbw_d	.1159134	49.31	1.301590e-01 (16.5 %)
hbs_d	.0671441	16.21	7.180723e-02 (9.1 %)

Following the designation of the 2SLS model structure with the related results, the calculation of the marginal effect of each explanatory variable come into considerations. In this sense, the related marginal effect of each variable on the number of daily motorized trips *at their means* (please *see* Cameron & Trivedi, 2005 for technical details) have been calculated and exhibited in this section of the thesis. Herein, the amount of induced number of daily trips of an individual -with respect to the one unit change in the daily travel time of this individual- has been calculated. Technically, this refers to the marginal effect calculation as revealed in equation 4.2;

$$\partial(\text{Ln}(\text{motor\_y}+1))) / \partial(\text{motor\_time}) = \beta_{1.1},$$

$$\text{Since Ln}((\text{motor\_y})+1) = \beta_{1.1} * (\text{motor\_time})\text{; then,}$$

$$e^{\beta 1.1 * (\text{motor\_time})} = (\text{motor\_y})+1;$$

$$e^{\beta 1.1 * (\text{motor\_time})} -1 = \text{motor\_y; then,}$$

$$\partial \text{motor\_y}/\partial \text{motor\_time} = \beta_{1.1} * (e^{\beta 1.1 * (\text{motor\_time})}) \quad (4.2)$$

According to the estimations, the marginal effects of the key leading figures have been estimated as *0.1302*, *0.0718*, *0.0536*, and *0.0197*, respectively for *hbw_d*, *hbs_d*, *auto_number*, and *hh_head_d* (*see* Table 4.8). Herein, an individual, who has practised at least one home based work trip (hbw_d) in the day, has a propensity to exhibit 0.1302 more trips in that day. This would reveal that one unit increase in the employment figure of İstanbul leads to 16.5 % increase in the average number of daily motorized trips per individual with refers to the explicit validity of the employment rate in affecting the amounts of daily trip makings in the case of İstanbul. Similarly, an individual, who exhibits at least one home based school trip (hbs_d) in the day, is able to carry out 0.0718 more trips in that day. That is to say, one more student participating in the education cycle in İstanbul leads to 9.1 % increase in the number of daily motorized trips. In addition, one more automobile owned by the related household leads to 0.0536 more trips with refers to 6.7 % increase in the amounts of daily motorized trip makings. Besides, any individual, who is household head (hh_head_d), carries out 0.0197 more motorized trips in a day when compared to other members of the household. Herein, the marginal change in the number of household heads makes the number of the daily motorized trips increase by 2.5 %. Lastly, the least inducing effects on the number of daily motorized trips (motor_y) come from the monthly disposable income of the individual (hh_income) and number of years of schooling of the individual (schooling_year). In this context, it is able to be stated that one hundred Turkish Lira increase in the monthly disposable income of the individual and one year of more education of this individual refer only to the 0.000227 (0.029 %) and 0.000885 (0.011 %) more daily trips for that individual, respectively.

On the other side, the estimated marginal effects of the leading negative factors of daily trip makings are equal to *-0.00876,-0.00834*, *-0.00342*, and *-0.000336*, respectively for *motor_time*, *male_d*, *hhsize*, and *age*. Herein, to begin with, one minute decrease in daily motorized travel time induces 0.00876 more daily motorized trips per passenger per day, which justifies the validity of induced travel demand theory once more. It is directly able to be stated that a ten minute decrease per capita in İstanbul Metro Area, which makes a 26 percent decrease in average motorized travel time, is able to make motorized trip makings increase by 11.4 % in

İstanbul. Secondly, in the case that the individual is male, there exist 0.00834 less motorized trips (1.1 % decrease) in a day. Thirdly, one unit increase in the household size precipitates 0.00342 less motorized trips (0.043 % decrease) in a day per passenger. Such a negative effect of household size on number of daily motorized trips would be mainly explained by the considerations of transportation cost per person in the related household under budget constraint. Fourthly, as an individual gets one year older, the number of daily motorized trips of that individual decreases by 0.000336 trips, which refers to 0.042 % decrease in daily motorized trips per passenger.

4.3 Comparison of GSEM and 2SLS Model Results

According to the findings of the thesis, one minute change in travel time induces 0.00095 and 0.0086859 more motorized trips per passenger in a day according to GSEM and 2SLS, respectively. In other words, any new transportation investment, that makes *Total System Travel Time (TSTT)* decrease by ten minutes, which makes a 26 percent decrease in average motorized travel time, would induce 261,250 (2.92 %) and 1,19 million (11.4 %) more motorized trips in a day according to GSEM and 2SLS, respectively in the case of İstanbul. That is to say, according to the results of GSEM, a ten minute decrease in average motorized travel time results in 174,167 more motorized vehicles in the daily traffic in addition to 14 million motorized vehicles in a day in İstanbul. On the other hand, according to the results of 2SLS, a ten minute decrease in average motorized travel time results in 793,333 more motorized vehicles in the daily traffic 14 million motorized vehicles in a day in İstanbul.

Such a difference between the related marginal elasticity estimations of these two optional models would have been caused by the difference in their model structures. In this sense, the usage of motorized free flow travel time (motorfft) as an intermediate variable (*see* Figure 3.1) between daily travel time (motor_time) and daily amount of motorized trips (motor_y) would have made the marginal elasticity estimation of GSEM decrease significantly when compared to the one of 2SLS. It is natural that the travel time sensitivities of passengers in the case of GSEM will dramatically decrease when the daily motorized travel time is intermediately explained by motorized free flow travel time (motorfft), which is a kind of purely disaggregated traffic parameter.

Furthermore, dummy variable standing for household heads (hh_head_d) and number of automobiles owned by the household (auto_number) are common leading factors generating daily motorized trips according to the results of both GSEM and 2SLS. On the other hand, the dummy variable standing for sex (male_d) comes into prominence according to the results of GSEM , while dummy variables with refers to the home based work (hbw_d) and home based school (hbs_d) come forward in making daily motorized trip makings explicitly increase according to the results of 2SLS. Even though the potential effect of these two models on policy implications would not differ so much, the selection among these two model structures (GSEM and 2SLS) would be *state of art* and more empirical researches from the cities of different countries will be required on this front.

5. DISCUSSION

In pursuit of the previous section of the thesis, exhibiting the two optional model results (GSEM and 2SLS), the discussions on these related results of these two models have been asserted in this section. In this sense, discussions on how to benefit from the related results of these models, namely Generalized Simultaneous Equations Model (GSEM) and Two Stages Least Squares (2SLS) Regression Model, have also been highlighted.

5.1 Discussions on GSEM and 2SLS Results

The findings of GSEM have confirmed the existence of induced demand in terms of traveling more. It is the first time in the literature that the measurement of induced urban passenger mobility demand is incorporated within a trip generation model via a series of regression models, each eliminating a different level of bias in the parameter estimates. This variation in the core parameter estimate proves that selection of correct model with correct speciation and estimator makes such biases dramatically decrease**.** In that sense, the path analysis, specified as the Generalized Simultaneous Equations Model (GSEM), seems one of the good candidates so as to measure the induced urban passenger mobility demand. Herein, the asserted measurement of the elasticities with GSEM, which refers to a complex structure, has also constituted another contribution of this study that may give a baseline for a future research direction in this area.

Non-linear structure of elasticity estimate of GSEM may further allow researchers to estimate the spatial variation of generative impact of induced urban passenger mobility demand, since it is possible to account for the individual characteristics in the estimation of elasticities as long as researchers have disaggregated spatial data. The proposed methodology is able to be used iteratively to estimate the generative impacts of different investment scenarios from the trip assignment phase of travel demand modelling.

The vital point for the investigation of the interrelationship between the *daily motorized travel time (motor_time)* and *number of daily motorized trips (motor_y)* is to perceive the backgrounding mechanism of the passengers' travel behavior via excluding any exogeneous effects as much as possible. In other words, the marginal willingness of an individual to travel, with reference to the motorized daily travel time of that individual, requires a kind of *disaggregated level intermediate variable*. Since, it is explicitly able to be asserted that the marginal willingness of an individual requires an intermediate level investigation between the related variables, called *motor_time* and *motor_y*. In this sense, even though the variables, namely motor_y and motor_time exhibit the types of *individual level disaggregated observations* (*see* Table 2.3 for definitions), an amount of daily aggregation have been still asserted [21]. Therefore, the addition of such a pure disaggragated intermediate variable ,namely *motorized free flow travel time (motorfft)* would be vital so as to minimize such an aggregation errors.

According to the results of GSEM, it would also be interesting to note that the coefficients of the variables, called *hhsize* exhibit negative signs for both dependent variables, namely *mobility_dummy* and *motor_y*. On the other hand, the coefficient of *age* for the dependent variable *mobility_dummy* is positive, while it is negative for the dependent variable *motor_y*. In other words, as the size of the household of the related passenger increases, then both the marginal propensity of this passenger to travel at least once in a day[22] and the number of daily motorized trips of this passenger tend to decrease. On the other hand, as *age* of the passenger increases, then the *number of daily trips of this passenger* (*motor_y*) decreases, while the *marginal propensity of this passenger to travel* (*mobility_dummy*) increases for the same case (as *age* increases) interestingly. Even though, at first glance, these two findindgs (signs of the coefficients of age) seem to be conflicting to each other, it would most probably be the case due to the fact that the *number of daily motorized trips of the passenger* as the main dependent variable involves *excess zero observations*. In other words, *the household size that the passenger belongs to (hhsize)* and *age of the individual (age)* induces an amount of increase in the probability of this passenger to

---

[21] ***motor_y***: sum of the numbers of daily trips & ***motor_time***: sum of the daily travel times of an individual. According to the related definitions (Table 2.3), an amount of low degree of daily aggregation is still exhibited due to summations.
[22] That refers to ***mobility_dummy =1***.

travel at least once in a day. Meanwhile, this makes *the number of the daily motorized trips* of this passenger decrease, which would be above zero.

Furthermore, with regards to the marginal elasticity estimations of GSEM, the leading factors, affecting *daily motorized number of trips*, are *hh_head_d*, *auto_number*, and *male_d* with the marginal elasticity coefficients of *0.256, 0.1884*, and *0.1597*, respectively. In other words, one unit increase in the number of *household head* and one unit increase in the number of *males* induce the amounts of *0.256* and *0.1597* more daily motorized trips in İstanbul, which would signal for the *patriarchal structure of the Turkish families* and *overweighting number of males in employment & in schools* in the daily trip flows with regards mostly to the obligatory trips, such as home-based work trips and home-based school trips. Lastly, one unit increase in the *number of private automobiles owned by the household (auto_number)* induces an amount of *0.1884* more trips in a day with refers to the automobile dependency in transportation flows, as it is the case in most of the developing countries.

Furthermore, the leading parameters enhancing the level of daily motorized trip makings are *hbw_d, auto_number, and hh_head_d* acording to the results of 2SLS. Besides, any new transportation investment, that makes total system travel time (TSTT) decrease by ten minutes, would be inducing 1.19 million more motorized trips in a day(793,333 more motorized vehicles in the daily traffic) in addition to the current daily 14 million motorized vehicles in İstanbul according to 2SLS. The leading factors are almost common according to the results of both GSEM and 2SLS.

5.2 Discussions on How to Benefit From These Findings

The findings of both GSEM and 2SLS model would strengthen the guidelines in setting any travel demand management policies. In this sense, the related policies would be channelized especially to the indicators; namely, *per cent of employment*, *number of automobiles owned by the related households* and *per cent of household head population among total within the related urban spaces* rather than to the classical indicators called household disposable income per month and urban population. In this manner, to illustrate, both 33 counties (with refers to 987 districts) of İstanbul and all the cities of Turkey would be systematically categorized according to their related indicators such as *employment level*, *number of students*, *average number of automobiles owned per household*, and *percent of household heads among population*. Afterwards, an amount of systematic prioritizations for some counties of İstanbul and for some cities of Turkey based on these parameters would be realized in implementation of the related travel demand management policies. Herein, the main objective would be minimizing the number of daily motorized trips in urban areas as much as possible and the changes in all those prominent parameters asserted above would be able to be defined as the performance indicators in using policy impact analyses of such travel demand management policies.

In addition, the results of both GSEM and 2SLS model would be able to be integrated into the classical *four stages travel demand models*, which are detailly explained by many textbooks involving the ones of Dickey (1983) and Ortuzar and Willumsen (2001). In this context, any new transportation investment, that makes *total system travel time (TSTT)* decrease by ten minutes, would induce 261,250 and 1.19 million more motorized trips in a day according to GSEM and 2SLS respectively, which should be re-integrated into the trip generation stage of the conventional four-step travel demand modelling process. That kind of reciprocal process is directly related to the inter-relationship between the *number of daily* ***motorized trips*** and ***daily motorized travel time*** of the passengers (*see* Figure 5.1).

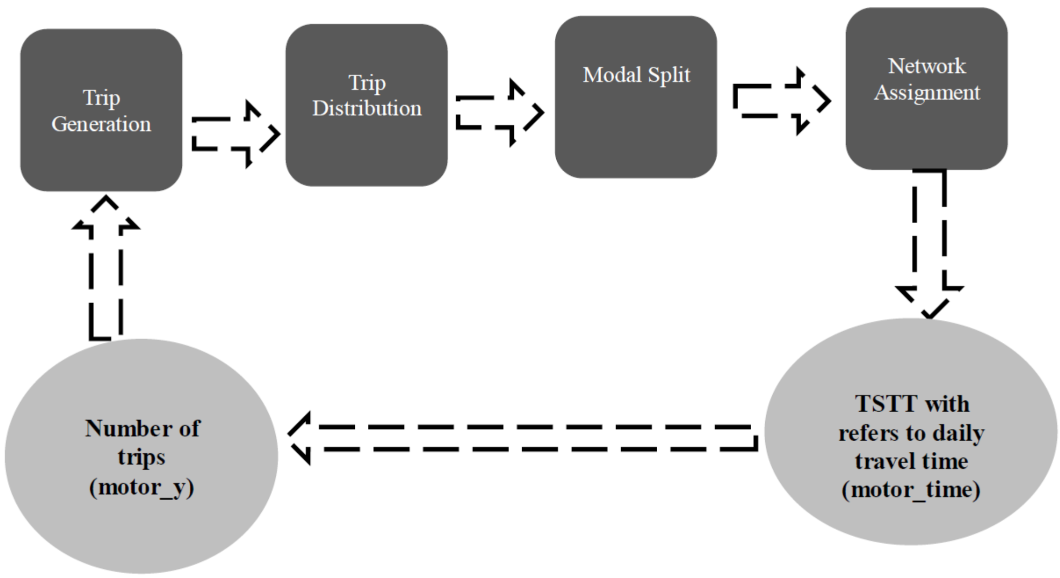

Figure 5.1 : Integration of the reciprocal relationship between *motor_y* & *motor_time* into classical four stages travel demand models.

In addition, such a kind of iterative process between the daily amounts of trips and daily travel time would also be integrated into the cost & benefit analysis of the related transportation investments in that the daily motorized travel time is a kind cost, which affects the daily number of trips reciprocally. Furthermore, the average monetary value of daily motorized travel time of each passenger would be calculated and then the potential changes in average daily travel time -as a result of new transportation investment- would be integrated into the cost & benefit analyses of such projects.

6. CONCLUSIONS AND RECOMMENDATIONS

The vitality of measuring motorized traffic with refers to the improvement of transportation infrastructure in urban spaces is mostly based on the concern of minimizing/managing motorized traffic within the considerations of social benefit. In this sense, stimulation of the usage of green modes of transport, which mostly refer to non-motorized travel modes (cycling, walking, etc.), so as to enhance the level of social benefits of urban societies. In addition, stimulation of he usage of such green modes of transport would stand also for the concern of minimizing the level of gas emissions in urban spaces.

In this sense, the thesis has proposed measurement of the motorized traffic so as to constitute a baseline in monitoring any travel demand management policy with especially regards to the motorized traffic flows in urban spaces. Furthermore, all these considerations asserted in the thesis would make policy makers see the urgency of discouraging enlargement of new urban roads for motorized traffic, since each new motorized route induces extra motorized travel demand. All these are directly related to understanding the mechanism of induced motorized traffic of passengers on the urban scale.

It is the first time in the literature that the measurement of induced urban passenger mobility demand is carried out so as to strenghten the baseline of a trip generation model via a series of models, each coping with a different level of bias in the parameter estimates. The proposed model structures, namely Generalized Simultaneous Equations System (GSEM) and Two Stages Least Squares (2SLS) Regression Model seem as the most convenient model structures so as to measure the induced urban passenger mobility demand. Furthermore, both model structures (GSEM and 2SLS) are able to be integrated into a trip generation model. Besides, these models are also able to be used iteratively so as to estimate the generative impacts of different investment scenarios from the trip assignment phase of travel demand modelling.

According to the results of the models, any new transportation investment ,that makes Total System Travel Time (TSTT) decrease by ten minutes, which makes a 26 percent decrease in average motorized travel time, would induce 261,250 (2.92 %) and 1,19 million (11.4 %) more motorized trips in a day according to respectively GSEM and 2SLS in addition to the current 21 million daily motorized trip counts for the case of İstanbul. Furthermore, these findings of GSEM and 2SLS refer to an increase in the number of motorized vehicles in the daily traffic by 174,167 and 793,333 more vehicles respectively in a day in addition to the current amount of 14 million daily motorized vehicles for the case İstanbul. These findings, in turn, should be re-integrated into the trip generation stage of the conventional four-step travel demand modelling process. That kind of iterative process is able to be run till the optimization between the numbers of daily motorized trips and daily travel time is reached. Such a kind of optimization process would explicitly constitute empirical inputs for the travel demand management policies and cost & benefit analysis of the related transportation investments.

In addition, the comparison of the potential changes in motorized travel times -as the result of developing new optional transportation projects- would be added to the cost & benefit analyses of these related projects. This would be carried out so as to select the optimum transportation project (or a bundle of projects) referring to optimum social benefit. Herein, the optimum social benefit would be formulated via the concern of minimizing daily amount of motorized trips (and so minimum gas emissions) and daily motorized travel time simultaneously as much as possible.

Another important finding of this thesis is that even though induced urban passenger mobility demand is well and alive, it is not the only and leading evil among other trip generating factors. Herein, high elasticity of being household head, being a male or owning more number of owned automobile might easily generate more trips than a new transportation network improvement according to the findings of both GSEM and 2SLS.

Furthermore, unlike traditional approaches in travel demand models, education levels and monthly income figures of the individuals exhibit the least positive effects on daily trip makings in İstanbul.

For the related future studies, the measurements of the generative impact of induced urban motorized mobility demand need to be tested in different urban settings with the asserted models via system-wide measurements in urban scales with disaggregated data in addition to the redistributive effects of induced demand measurements with the considerations of capacity improvements. Lastly, these measurements would refer to a kind of performance indicator for policy impact analysis of any transportation project about the policies for the considerations of travel demand management. Obviously, more measurings are required.

REFERENCES

Barr, L. (2000). Testing for the significance of induced highway travel demand in metropolitan areas. *Transportation Research Record: Transportation Planning, Public Participation, and Telecommuting*(1706), 1-8.

Breen, R. (1996). *Regression Models Censored Sample Selected or Truncated Data*. Thousan Oaks, California: Sage Publications.

Cameron, C., & Trivedi, P. (2005). *Microeconometrics:Methods and Applications.* New York: Cambridge University Press.

Cervero, R. (2002). Induced travel demand: Research design, empirical evidence, and normative policies. *Journal of Planning Literature, 17*(1), 3-20.

Cervero, R. (2003). Road expansion, urban growth, and induced travel: A path analysis. *Journal of the American Planning Association, 69*(2), 145-163.

Cervero, R., & Hansen, M. (2002). Induced travel demand an induced road investment: A simultaneous-equation analysis. *Journal of Transport Economics and Policy, 36*, 469-490.

Dalgleish, T., Williams, J., Golden, A., Perkins, N., Barrett, L., & Barnard, P. (2007). *Introductory Econometrics.* Wooldridge.

Dickey, J. (1983). *Metropolitan Transportation Planning* (Vol. 2). Taylor and Francis, Bristol P.A.

Downs, A. (1992). *Stuck in Traffic: Coping with Peak-Hour Traffic Congestion.* Brookings Institution Press.

Fulton, L., Meszler, D., Noland, R., & Thomas, J. (2000). A statistical analysis of induced travel effects in the U.S. mid- Atlantic region. *Journal of Transportation and Statistics, 3*(1), 1-14.

Genar. (2006). İstanbul Ulaşım Ana Planı Seyahat Başlangıç ve Bitiş Hanehalkı Anketleri Pilot Çalışması Raporu. İstanbul.

Green, W. (2003). *Econometric Analysis* (5th ed.). New Jersey: Pearson.

Green, W. (2007). Limdep Version 9.0. *Econometric Modelling Guide , v.2*. New York: Plainview.

Gujarati, D. (2003). *Basic Econometrics* (Vol. 4). New York: McGraw-Hill/Irwin.

Hansen, M., Gillen, D., Dobbins, A., Huang, Y., & Puvathingal, M. (1993). *The Air Quality İmpacts of Urban Highway Capacity Expansion: Traffic Generation and Land Use Change*. Berkley: University of California, Institute of Transportation Studies.

Hansen, M., & Huang, Y. (1997). Road supply and traffic in California urban areas. *Transportation Research Part A , 31*, 205-218.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica, 47*(1), 61-153.

Hill, R. (1998, July). What sample size is "Enough" in internet. *Interpersonal Computing and Technology, 6*(3-4).

Hogg, R., Tanis, E., & Zimmerman, D. (2014). *Probability and Statistical Inference* (Vol. 9). Pearson.

Holcombe, R., & Williams, D. (2010). Urban sprawl and transportation externalities. *The Review of Regional Studies, 3*(40), 257-273.

Hymel, K., Small, K., & Dender, K. (2010, February 5). Induced demand and rebound effects in road transport. *Transportation Research Part B Methodological, 44*(10),1220-1241.

İBB. (2007). İstanbul Ulaşım Ana Planı Hanehalkı Araştırması. İstanbul Büyükşehir Belediyesi, Ulaşım Daire Başkanlığı, Ulaşım Planlama Müdürlüğü, İstanbul.

Kennedy, P. (1998). *A Guide to Econometrics*. Cambridge, Massachusetts: The MIT Press.

Kroes, E., Daily, A., Gunn, H., & Van der Hoorn, T. (1996). The opening of the Amsterdam ring road: A case study of short term impacts on removing bottleneck. *Transportation, 23*, 71-82.

Lachin, J., McGee, P., Greenbaum, C., Palmer, J., Pescovitz, M., Gottlieb, P., & Skyler, J. (2011, November 10). Sample size requirements for studies of treatment effects on beta - Cell function in newly diagnosed type 1 diabetes. (Z. Su, G. Inc., & U. S. America, ed.) *Journal of Pone, 6*(11).

Lee, D., Guldmann, J., & Rabenau, B. (2018). Interactions between the built and socio-economic environment and driver demographics: Spatial econometric models of car crashes in the Columbus Metropolitan Area. *International Journal of Urban Sciences, 22*(1), 17-37.

Long, J. (1997). *Regression Models for Categorical and Limited Dependent Variables* . London, New Delhi, Thousand Oaks: Sage Publications.

Luk, J., & Chung, E. (1997). *Induced Demand and Road Investment: An Initial Appraisal*. ARR 299. Vermont South, Australia: ARRB Transportation Research .

Maddala, G. (1983). *Limited Dependent and Qualitative Variables in Econometrics*. New York: Cambridge University Press.

Melo, P., Graham, D., & Canavan, S. (2012). *The Effects of Road Investments on Economic Output and Induced Travel Demand: Evidence For Urbanized Areas in the US*. Transportation Research Board.

Mokhtarian, P., & Cao, X. (2008). Examining the impacts of residential self-selection on travel behavior: A focus on methodologies. Transportation Research Part B, 42, 204-228.

Mokhtarian, P., Samaniego, F., Shumway, R., Willits, N., & Azari, R. (2000). Analyzing Induced Traffic from Capacity Enhancements Used Matched Pairs: A California Study. *Davis: University of California, Institute of Transport Studies* .

Noland, R., & Cowart, W. (2000). Analysis of metropolitan highway capacity and the growth in vehicle miles of travel. *Transportation*, *27*(4), 363-390.

Ortuzar, J., & Willumsen, L. (2001). *Modelling Transport* (Vol. 3). New York, N.Y.: John Wiley & Sons.

Ozuysal, M., & Tanyel, S. (2008). Induced travel demand in developing countries: study on state highways in Turkey. *Journal of Urban Planning and Development -ASCE, 134*(2), 78-87.

Pells, S. (1989). User response to new road capacity: A review of published evidence. *University of Leeds, Institute of Transport Studies, Workin Paper 283*.

Shapiro, S., & Wilk, M. (1965). An analysis of variance test for normality (complete samples). *Biometrica*, *52*(3/4), 591-611.

Silva, A., & Costa, G. (2007, September). Urban sprawl and energy use for transportation in the largest Brazilian cities. *Energy for Sustainable Development* , *3*(6), 44-50.

Terza, J. (1998). Estimating count data models with endogeneous switching: sample selection and endogeneous treatment. *Journal of Econometrics*, *84*, 129-154.

Vos, J., & Witlox, F. (2013). Transportation policy as spatial planning tool; reducing urban sprawl by increasing travel costs and clustering infrastructure and public transportation. *Journal of Transport Geography, 33*, 117-125.

Wilson, C., Voorhis, V., & Morgan, B. (2007). Understanding power and rules of thumb for determining sample sizes. *Tutorials in Quantitative Methods for Psychology* , *3*(2), 48.

Winkelmann, R. (2008). *Econometric Analysis of Count Data* (Vol. 5). Verlag, Berlin, Heidelberg: Springer.

CURRICULUM VITAE

Name Surname : Enver Cenan İNCE

Place and Date of Birth : Konya/ Bozkır, 02.05.1987

E-Mail : envrcnn@gmail.com

EDUCATION :

- **B.Sc.** : 2012, Middle East Technical University, Faculty of Architecture, Department of City and Regional Planning
- **M.Sc.** : 2012, 2015, İstanbul Technical University, Faculty of Architecture, Department of Regional Planning

PROFESSIONAL EXPERIENCE AND REWARDS:

- 2013-2019: Research Assistant in İstanbul Technical University at the Faculty of Architecture, Department of City and Regional Planning.
- 2019 award of *early career professionals from academia* by Tower Transit in Thredbo 16 (in *16th International Conference on Competition and Ownership in Land Passenger Transport*, August 25-30, 2019, Singapore).

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

Ince, E.C., & Çelik, H.M. (2019). Measuring the Induced Travel Demand: Evidence from a Path Analysis of Generalized Simultaneous Equations Model for the Case of İstanbul. In M. Dalkılıç (Ed.), *Academic Researches in Science and Engineering* (pp. 15-42). New York, 387 Park Avebue South, 5th floor, 10016, USA: Gece Publishing.

Ince, E.C., & Çelik, H.M. (2019). Measuring Urban Motorized Passenger Mobility: Evidence from Sample Selection Poisson Regression Model in the Case of İstanbul. *Civil Engineering and Architecture, 7*(4), 215-227.

OTHER PUBLICATIONS, PRESENTATIONS AND PATENTS:

İnce, E.C., Türk, Ş.Ş. (2016, Aralık 1). Büyük Ölçekli Ulaşım Projelerine Yönelik Hukuki Süreçlerin İncelenmesi. *Süleyman Demirel Üniversitesi 16. Bölge Bilimi ve Bölge Planlama Kongresi*, s. 8.

İnce, E.C., Türk. Ş.Ş. (2017). Büyük Ölçekli Ulaşım Projelerine Yönelik Hukuki Süreçlerin İncelenmesi. Kadir Temurçin, Dulupçu M.A.(Ed.), *Türkiye'de Mekansal ve Bölgesel Dönüşümler* (ss. 315-334). Isparta: Süleyman Demirel Üniversitesi Yayınevi, ISBN: 978-605-9454-04-9.

İnce E.C., Akgün A. A. (2017). *Güneydoğu Anadolu Projesi (GAP) ve Bölgesel Gelişme (Turkish Edition).* LAP LAMBERT Academic Publishing, ISBN-13:978-659-90202-4, ISBN-10: 3659902020, EAN: 9783659902024 .

Cengiz, E.C., **Ince, E.C.,** & Çelik, H.M. (2019). Beyond Farebox: Financing Railway via Betterment Levies. In T. Transit (Ed.), *16th International Conference on Competition and Ownership in Land Passenger Transport (Thredbo 16).* Singapore.

Özdemir, M., Ayataç, H., & **Ince, E.C. (2019).** Şehirlerin Nüfus Kapasitelerinin Belirlenmesine Yönelik Bir Metodolojik Yaklaşım Önerisi. In I. o. Studies (Ed.), *IV. International Congress of Urban Studies.* Ankara: Kent Araştırmaları Enstitüsü, İdealKent (Kent Araştırmaları) Dergisi, TOBB, ETÜ, TEPAV.