MODELING THE EFFECTS OF LAND USE CHARACTERISTICS ON MODE CHOICE FOR HOME – BASED WORK TRIPS: THE CASE OF ISTANBUL

Uğur YANKAYA

İzmir Institute of Technology
December, 2010

MODELING THE EFFECTS OF LAND USE CHARACTERISTICS ON MODE CHOICE FOR HOME – BASED WORK TRIPS: THE CASE OF ISTANBUL

A Thesis Submitted to
the Graduate School of Engineering and Sciences of
İzmir Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in City Planning

by
Uğur YANKAYA

December 2010
İZMİR

We approve the thesis of **Uğur YANKAYA**

Assoc. Prof. Dr. Semahat ÖZDEMİR
Supervisor

Prof. Dr. Cemal ARKON
Co-Supervisor

Prof. Dr. Gökmen TAYFUR
Committee Member

Assoc. Prof. Dr. K. Mert ÇUBUKÇU
Committee Member

Assoc. Prof. Dr. Metin ŞENBİL
Committee Member

Assist. Prof. Dr. Yavuz DUVARCI
Committee Member

29 December 2010

Assoc. Prof. Dr. Semahat ÖZDEMİR
Head of the Department of
City and Regional Planning

Prof. Dr. Durmuş Ali DEMİR
Dean of the Graduate School of
Engineering and Sciences

ACKNOWLEDGMENTS

First of all, I would like to thank everyone who supported me during the preperation of this thesis. I offer my sincere gratitude to my advisor, Professor Cemal ARKON for his support, guidance, and patience. I would also like to special thanks to Assoc. Prof. Dr. Semahat ÖZDEMİR and Assoc. Prof. Dr. K. Mert ÇUBUKÇU for everything. This dissertation would not have been possible to accomplish without their support, interest, critical review, and help.

I would also like to thank my committe members, Prof. Dr. Gökmen TAYFUR, Assoc. Prof. Dr. Metin ŞENBİL, and Assist. Prof. Dr. Yavuz DUVARCI for their support and critical review. I would like to thank to Assoc. Prof. Dr. Aylin Alin and Assist. Prof. Dr. Emel Kuruoğlu in Department of Statistics from Dokuz Eylül University for their statistical support.

I very much appreciate the support and encouragement of my friends, Kader REYHAN, Nida Kamil ÖZBOLAT, Mert KOMPİL, Onur KINACI, Nilüfer DÜNYA, Emel GÜNAY, HamidrezaYAZDANİ, İlknur UĞUR, and Türker ATMACA.

I would like to say thank you to the Department of City and Regional Planning and Izmir Institute of Technology for providing the environment and opportunities throught my time at IYTE. I am very grateful to Istanbul Metropolitan Municipality and BIMTAŞ for data and their help.

Finally, I am very grateful to my parents, Taşdemir and Gülay YANKAYA, and my brother, Oğuz YANKAYA for their support, patience, encouragement, trust, and everything during this thesis. Without their support and encouragement, I would never finish this work.

ABSTRACT

MODELING THE EFFECTS OF LAND USE CHARACTERISTICS ON MODE CHOICE FOR HOME – BASED WORK TRIPS: THE CASE OF ISTANBUL

The cities in Turkey have been facing some of the same problems that European and North American cities have, including traffic congestion, traffic accidents, and air pollution. To overcome this situation, both local and central administrators who make urban policies and city planners have tended to optimize Land Use and Transportation Interaction (LU&T). In recent years, some new concepts concerning urban planning have suggested that shaping travel demands can be used as a tool to overcome these problems. The most common objectives of this concept are to reduce motorized trips and to promote public transit in metropolitan areas. To achieve these objectives, understanding the probable effects of land use on mode choice is crucial. However, the effects of land use on mode choice have never been answered fully, in Turkey. This dissertation empirically explores the relationship between travel mode choice and land use by employing different mode choice models for home - based work (HBW) trips in Istanbul at aggregate and disaggregate levels. The focus of this study is to understand how land use characteristics affect home - based work mode choice in the case of Istanbul. In this study, logit models and Bayesian Belief Networks (BBNs) are used to identify and quantify the effects of land use on travel mode choice at both levels. Empirical data were obtained from 2006 Household Travel Survey prepared for 2007 Istanbul Transportation Master Plan Study. The model results show that land use variables are statistically significant at both levels. The inclusion of land use variables increases models' explanatory level.

ÖZET

EV - UÇLU İŞ YOLCULUKLARI İÇİN ARAZİ KULLANIM KARAKTERİSTİKLERİNİN TÜR SEÇİMİ ÜZERİNE ETKİLERİNİN MODELLENMESİ: İSTANBUL ÖRNEĞİ

Ülkemiz kentleri, Avrupa ve Kuzey Amerika'daki kentlerde görülen trafik sıkışıklığı, trafik kazaları ve hava kirliliği gibi bazı problemlerin benzerleri ile karşı karşıya kalmaktadır. Bu durumun üstesinden gelebilmek için gerek kentsel politikalar üreten yerel ve merkezi yöneticiler gerekse kent plancıları arazi kullanım ile ulaşım etkileşimini (LU&T) eniyilemeye (optimize) yönelmişlerdir. Son yıllarda kent planlama ile ilgili bazı yeni anlayışlar, seyahat taleplerinin biçimlendirilmesinin bu problemlerin üstesinden gelmede bir araç olarak kullanılabileceğini önermektedirler. Bu anlayışın temel hedefleri, büyük kentsel alanlarda motorlu araçlarla yapılan seyahatleri azaltmak ve toplu taşımayı geliştirmektir. Bu objektiflere ulaşmak için, arazi kullanımın tür seçimi üzerindeki olası etkilerinin anlaşılması önemlidir. Ancak, tür seçimi üzerinde arazi kullanımın etkileri Türkiye'de tam olarak cevaplanmamıştır. Bu tez, ev - uçlu iş (HBW) yolculukları için İstanbul'da "seyahat tür seçimi" ve "arazi kullanım" arasındaki ilişkinin farklı tür seçim modelleri uygulayarak ampirik olarak toplu ve bireysel düzeyde incelenmesidir. Bu çalışmanın odağı, İstanbul örneğinde arazi kullanım karakteristiklerinin ev - uçlu iş yolculuk tür seçimini nasıl etkilediğini anlamaktır. Bu çalışmada, lojit modeller ve Bayesian Belief Networks (BBNs), seyahat tür seçimi üzerinde arazi kullanımın etkilerini her iki düzeyde tanımlamak ve ölçmek için kullanılmaktadır. Deneysel data, 2007 İstanbul Ulaşım Master Plan çalışması için hazırlanan 2006 Hanehalkı Anketinden temin edilmiştir. Model sonuçları, arazi kullanım değişkenlerinin her iki düzeyde istatistiksel olarak anlamlı olduklarını göstermektedir. Arazi kullanım değişkenlerinin ilave edilmesi, modellerin açıklama düzeyini arttırmıştır.

To my beloved parents, Gülay & Taşdemir YANKAYA,
who have trusted and supported me throughout my life

Hayatım boyunca bana inanan ve destekleyen,
sevgili annem ve babam, Gülay & Taşdemir YANKAYA'ya

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

HBW	Home - Based Work Trips
LOS	Level of Service
LU&T	Land Use and Transport
U.S.	United States
VMT	Vehicle Miles Travelled
IIA	Independence of Irrelevant Alternatives
MNL	Multinomial Logit Model
BBNs	Bayesian Belief Networks
OD	Origin Destination
BN	Belief Network
UTMS	Urban Transportation Modeling System
ASC	Alternative Specific Constants
GEV	Generalized Extreme Value
NL	Nested Logit
MLFFN	Multi Layer Feed Forward Network
DAG	Directed Acyclic Graph
CPTs	Conditional Probability Tables
NP	Non-Deterministic Polynomial Time
AIC	Akaike Information Criteria
MDL	Minimum Description Length
LL	Log Likelihood
CIDR	Conditional Independence Dependence Relations
EM	The Estimation - Maximization Algortihm

TAZ	Travel Analysis Zone
GIS	Geographical Information Systems
SRMSE	Standardized Root Mean Square Error
RMSE	Root Mean Square Error
MSE	Mean Square Error
ARV	Average Relative Variance
MAE	Mean Absolute Error
ANN	Artificial Neural Networks
ROC	Receiver Operating Characteristics Curve
MLE	Maximum Likelihood Estimator
HBS	Home - Based School Trips
HBO	Home - Based Other Trips
NHB	Non - Home - Based Trips
TL	Turkish Lira
LU	Land Use
TT	Travel Time
TC	Travel Cost
SE	Socioeconomic
LRT	Light Rail Transit
CBD	Central Business District
LRI	Likelihood Ratio Index
WTP	Willingness to Pay
VTTS	Value of Travel Time Savings
JHB	Jobs - Housing Balance Ratio

CHAPTER 1

INTRODUCTION

The common problems in the cities are traffic congestion, air pollution, automobile dependency, uncontrolled development, and decentralization of jobs that need to structure a link between land use and transportation (LU&T). Therefore, interest in land use and transportation relationship has gained popularity in recent years. In Europe, U.S., and also Asian cities, the most common land use and transportation policy is to reduce the number of motorized trips and encourage the use of public transport. In Western world, many policy makers and planners propose some land use strategies such as high density development, smart growth, new urbanism, transit oriented development, and mixed land use as a solution of those problems. It is intended that these land use policies are used to create changes in travel behavior. The logic behind these solutions is to create a land use pattern that provides improved accessibility to choices for housing, employment, retail, and other opportunities but with less demand for motorized trips. These are main agenda for urban planning over the last two decades. To achieve this objective, spatial configuration of land use in terms of planning and design should be integrated into the modeling process in urban transportation planning. However, since the relationship between LU&T has complex and multidimensional, the relationship is not well enough understood in the World and especially in Turkey. There is a lack of empirical support for the existence of an association in the case of the cities of Turkey. Also, the question of whether land use characteristics affect travel behavior has never been fully answered. Main reason for this lack of empirical study has been the unavailability of empirical data such as household travel surveys and land use data. This subject is important for developing land use policies aimed at reducing motorized trips in metropolitan areas in developing countries. In U.S. and Europe, land use planning and urban design concepts have been used as a powerful tool for shaping travel demand. On the other hand, there is ongoing argument whether or not this relationship is important for explaining travel behavior, even in developed countries.

Travel behavior consists of several aspects: mode choice, route choice, trip chaining, vehicle-miles traveled (VMT), commuting time, etc. (Cervero and Kockelman 1997, Zhang 2004, Coevering and Schwanen 2006). This study focuses only on mode choice for home - based work (HBW) trips as one aspect of travel behavior. Mode choice itself is a distinct area of traditional four - step transportation modeling. In existing literature, it is suggested that three classes of variables affect mode choice: socioeconomic characteristics of travelers, characteristics of the journey, and characteristics of the transport facility (Ortuzar and Willumsen 2006). Since the effects of land use attributes on travel behavior are complex, potential land use indicators in and around trip origin and destination remains unanswered. In empirical studies, land use variables are generally omitted from modeling process, except for density variable. However, density and other factors as mentioned above cannot capture the all effects of land use on travel behavior. Therefore, *a research question arises as whether land use attributes affect mode choice or not at aggregate level and disaggregate level?*

After 1990s, studies have focused on measuring the effect of land use characteristics on mode choice. In spite of growing interest and voluminous empirical literature, many issues needs to be explored. Firstly, there is no consensus about the factors affecting travel behavior. The empirical findings are mixed since complex and multi-dimensional relationship between land use and mode choice make analyzing difficult (reviewed by Crane 2000). For example, Stead (2001) found that both land use and socioeconomic characteristics influence travel pattern. However, socioeconomic characteristics explain more of the variation in travel patterns than land use. Cervero (2002) found that land use characteristics improved model’s predictability although not as significantly. In this study, the influence of urban design was more modest than land use. Zhang (2004) found that land use has an independent influence on mode choice like Cervero (2002) and Cervero and Kockelman (1997). In addition, most studies suggest that mixed land use, walk - friendly urban development, high density, and transit accessibility reduce motorized trips and travel distance. On the other hand, Ewing et al. (1996) found no significant relationship between total trip frequencies and land use. Crane and Crepeau (1998) and Rodriguez et al. (2006) did not found enough evidence on the relationship between neighborhood design and travel behavior change. Ewing and Cervero (2001) suggested that the association between the built environment characteristics and travel behavior is statistically significant, but the association has limited links. Despite the conflicting results, socioeconomic and land use factors affecting travel demand require further and comprehensive empirical studies.

In terms of urban planning, it has been suggested that land use factors (or built environment) have been thought to influence travel demand along three principal dimensions: density, diversity, and design (Kockelman 1997, Cervero and Kockelman 1997, Cervero 2002). The general hypotheses in previous studies are that higher densities and mixed land use are thought to decrease motorized trips. They are positively correlated with transit choice and non motorized trips. Density is a common measure in empirical studies. It assumes that people who lived in higher density areas use more transit (public transportation) and non - motorized modes because of parking problems, good access to transit service, and congestion problem. However, land use and transportation system attributes have been often treated as exogenous variables in travel demand analysis. The models may ignore the effects of these attributes that may play important role in residential location decisions. Although recent studies still have suggested that land use attributes may affect mode choice behavior, it cannot be said that there is a consensus on the degree of the impacts. Some empirical studies found a correlation between land use and mode choice. However, questions remain regarding strength and direction of the relationship. Another issue is that which land use characteristics influence travel behavior has not been adequately explained.

Several weaknesses of the existing studies still remain. One of them is that many empirical studies have been motivated by urban design approaches such as new urbanism and transit oriented development. These design concepts are assumed as a way of shaping aggregate and disaggregate travel demand in the USA and Europe. These design philosophies are new for the cities in developing countries. From the perspective of developing countries, cities mainly have been developed by lack of urban design concepts and planning decisions. In addition, land use (or urban form) data can not readily be available and measured. On the other hand, expensive public transit investments are more common in developed part of the world than those of developing countries. Rail transit networks are not widespread in comparison to developed counties. Therefore, rail modes cannot be an alternative mode for each zone in developing countries.

Mode choice models calibrated with disaggregate data are used to explain individuals' behaviors while the aggregate models analyze to predict the zonal shares of trips by different travel modes. In the existing literature, empirical analysis of mode choice is generally based on discrete choice models developed from consumer choice theory (Domencich and Mc Fadden 1975, Ben Akiva and Lerman 1985). Multinomial logit (and conditional logit) models are the most used and preferred probabilistic choice models up to now. Since probit models need computational effort, logit models have been used increasingly in mode choice studies, especially with disaggregate data. However, the assumption of independence of irrelevant alternatives (IIA) is an important restriction for the application of discrete choice models. In mode choice studies, alternative approaches that are more flexible than discrete choice models are needed to develop. For example, soft computing methods do not suffer some statistical assumptions. The application of soft computing methods for modeling and analyzing transport systems is new and unexplored in comparison with discrete choice models. Among soft computing methods (neural networks, fuzzy logic, neuro-fuzzy, and genetic algorithms), bayesian belief networks (BBNs) are rather new approach for dealing with decision problems under uncertainty. Traditional methods can not adequately explain the complex relationships. Therefore, new methods may provide more information under uncertainty and complex problem domains for city planners.

Even though the authors studying the mode choice have reached varying results in their findings, the urban environments they were analyzing shared certain similarities. In the existing literature, these urban settings mostly took place in the developed economies. North-American cities dominate the literature: Los Angeles Area (Cambridge Systematics 1994), Seattle Area (Frank and Pivo 1994, Frank, et al. 2007), San Francisco Bay Area (Cervero and Kockelman 1997, Kitamura, et al. 1997, Kockelman 1997, Cervero and Duncan 2002, Bhat and Guo 2007), Portland (Rajamani, et al. 2003), Maryland (Cervero 2002), and New York City (Chen and McKnight 2007). The Greater Dublin Area in Ireland (Commins and Nolan 2010), Hong Kong (Zhang 2004), and The Netherlands (Schwanen, et al. 2004, Limtanakool, et al. 2006) are the other urban environments analyzing the connection between land use and travel behavior in developed countries. There is noteworthy effort to analyze the relationship between land use and travel behavior in disaggregate analysis. For example, at disaggregate level, Zhang (2004) and Cervero (2002) found that land use variables improved disaggregate model's explanatory power. The significance of land use and urban form characteristics vary among the cases. However, there is no enough evidence at aggregate level. Little attention has been given to the analysis of zonal behaviors with different empirical models. Also, empirical studies have been made at either aggregate or disaggregate levels. To date, there have been few empirical studies analyzing and comparing the potential effects of land use on travel demand for both levels at the same time. Therefore, the aim of the study is to expand the understanding of the relationship between land use and mode choice by accounting for alternative approaches to choice models at aggregate and disaggregate levels in the case of Istanbul, Turkey, so as to achieve a better understanding of the effects of land use on mode choice. The study explores this research by offering a comparative empirical study on the performance of two different type models: Logit Models and Bayesian Belief Networks.

Under this framework, the objectives of the study are:

1. To examine the relationship between land use and travel mode choice with the application of the traditional (conventional) and alternative methods with respect to the usefulness of their information provided when estimating and forecasting travel behavior in terms of mode choice.

2. To explore how the effects of land use on mode choice may differ at both aggregate and disaggregate level.

In the content of the study, the measure of land use pattern is defined in terms of three core dimensions of spatial configuration in the city: ***density, diversity, and accessibility*** like Cervero and Kockelman (1997). Main hypothesis of the study is that land use characteristics affect mode choice decisions for home - based work trips in Istanbul at aggregate and disaggregate levels. In addition, this study aims to test following sub-hypotheses in Istanbul.

SH-1. Adding land use variables to the models at aggregate and disaggregate levels improves the model’s explanatory power.

SH-2. Alternative methods (BBNs) are superior to traditional (conventional) models (logit models) in mode choice modeling at both levels.

The following sub-hypotheses associated with land use variables are derived from the relationship between mode choice and land use in Istanbul are tested:

H.1. Population density is positively correlated with walking and transit mode choice.
H.2. Employment densities are positively correlated with motorized trips.
H.3. Diversity positively correlates with walk and transit mode choice.
H.4. Transit access increases the choice of transit mode.
H.5. Commuters whose trip origin and destination point is in the same zone are more likely to choose non-motorized alternatives.

In order to achieve the objective of the study, firstly, academic research focusing this relationship between mode choice and land use is reviewed. It is not paid enough attention to this subject, especially in developing countries and soft computing methods. There is a lack of empirical studies in Turkey, while the findings of empirical studies in the world are not generalized. Because of this reason, this study seeks to answer the following questions in Istanbul:

- Is there a statistically significant association between land use characteristics and travel pattern in terms of mode choice?
- Which land use attributes show statistically significant with mode choice and to what extent at aggregate and disaggregated levels?
- What are the similarities and differences for the relationship between LU&T in comparison with Western cities?

To have a comprehensive understanding to the influence of land use on mode choice, the study examines the relationship based on some dependent variables:

- The likelihood of using different modes (walk, car, service, and transit),
- The likelihood of traveling according to aggregate and disaggregate mode choice.

The contribution of the study to the existing literature is two-fold. First, new models, baseline category logit and BBNs, are introduced to explore the effects of land use attributes on mode choice at both levels. The methods are expected to provide more information under uncertainty and missing data in transportation applications. Such an alternative model can predict the choice probabilities as well as mode choice decisions that may be affected by land use policies. Second, from a methodological framework, this study presents a methodology for simultaneously and comparatively modeling the LU&T interaction for both levels. Data used in this study is based on 2006 Household Travel Survey conducted by the Transportation Department of the Metropolitan Municipality of Istanbul. This survey was prepared for 2007 Istanbul Transportation Master Plan. Response rate in this study is 263,768 people in 70,888 households. In the content of the study, the models are calibrated using aggregate and disaggregate data. The final data set for disaggregate models contain 116992 home - based work trips while zonal (aggregate) level sample includes 406 travel analysis zones. In Istanbul, 451 travel analysis zones are determined for 2007 Transportation Master Plan. In aggregate models, 45 zones are excluded from the models due to lack of land use data and few household survey studies for these zones. In order to test performance comparisons of the models, the data used in aggregate and disaggregate models is partitioned into two subsets, randomly: training and testing sets. Training sub-datasets are used to develop the models and testing sub-datasets that are not used in training process, are used to accuracy and performance comparisons of the models. Training data include 80% of total data while testing data include 20% of total data. The aggregate models are based on non OD (origin-destination) - based data whereas disaggregate models are based on OD data. In line with previous studies, logit models (MNL and Baseline Category Logit) are used as a traditional (conventional) model while BBNs are used as an alternative method to mode choice. SAS and Limdep - Nlogit programs for baseline category logit and multinomial logit models are used to estimate model parameters. Belief Network (BN) PowerConstructor and Hugin Researcher (Version 7.1) softwares are used to construct the network and estimate model parameters at both levels in Bayesian Belief Networks.

This study has six parts. Chapter 2 introduces a review of the literature. After that, modeling approach is discussed in Chapter 3. Chapter 4 presents description of data source and processing. This is followed by a presentation of the model results in Chapter 5. The study ends with conclusion in Chapter 6.

CHAPTER 2

LITERATURE REVIEW

In mode choice analysis, spatial configuration of land use may be an important factor for explaining individual behavior. However, land use characteristics can be still neglected in empirical studies. Since the relationships between land use and transportation are complex and multidimensional, the probable effects of land use vary considerably from case to case. Many empirical studies need to be explored in different urban settings. Therefore, the focus in this stage shifted to analyze the effects of land use (built environment or urban form) in recent years. In this chapter, firstly, traditional four - step modeling is discussed in Section 2.1. After that, land use and mode choice relationship are presented in Section 2.2. The section provides information about the pattern of travel demand in developing countries. The section also includes the different formulations used in measuring land use characteristics and empirical applications focusing this interaction. Alternative approaches to traditional methods are described in Section 2.3. Soft computing methods used in travel demand modeling are discussed in this section. The methods used for performance comparisons of different mode choice models and the algorithms used in the model estimation are described in Section 2.4 and Section 2.5, respectively.

2.1. Review of The Four - Step Models and Land Use - Transportation Models

In 1950s, city planners and civil engineers firstly developed urban transportation models. The four - step model as seen in Figure 2.1 (or known as the urban transportation modeling system) has been used increasingly in transportation modeling up to now. The classic four - step transportation modeling system is applied for a zoning and network system. The system needs detailed empirical data that are obtained from mainly travel surveys (e.g., household travel surveys, roadside surveys, modal surveys, etc.). Urban Transportation Modeling System (UTMS) consists of four major and sequent stages (Meyer and Miller 2001, Ortuzar and Willumsen 2006).

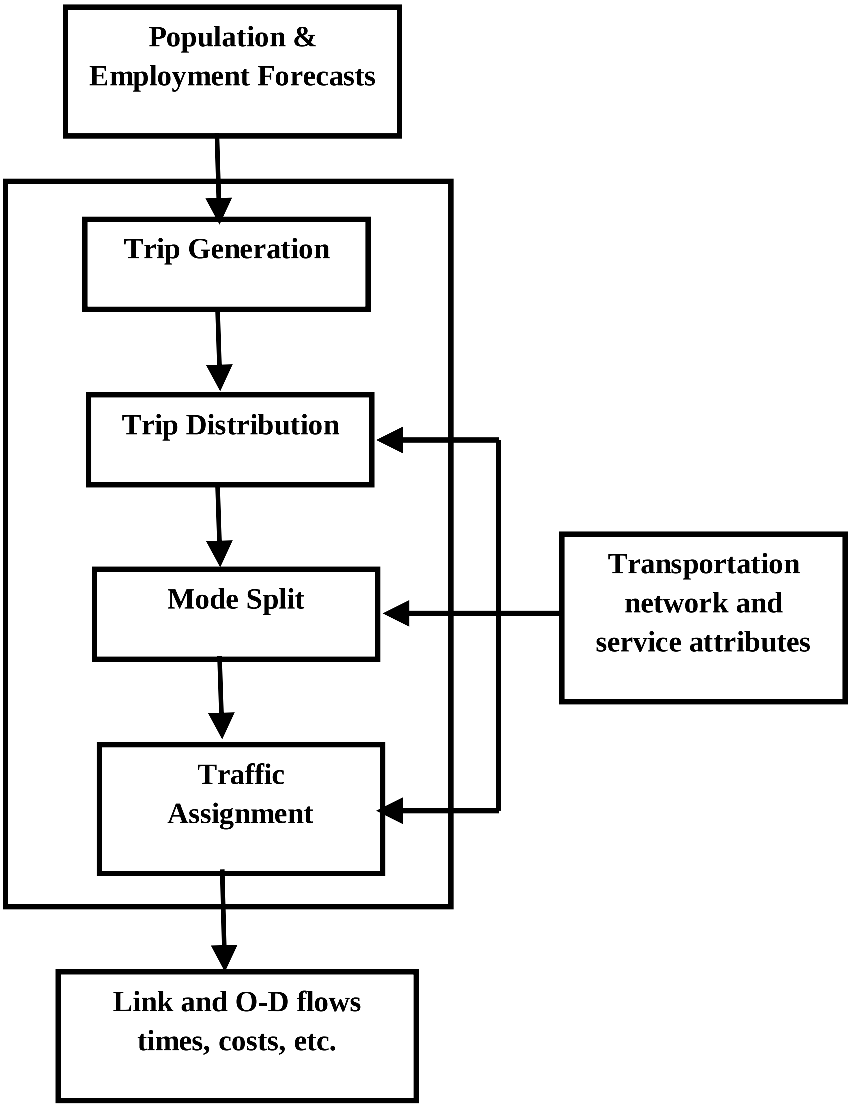


Figure 2.1. The urban transport modeling system
(Source: Meyer and Miller 2001)

1. Trip Generation Models are used to predict the number of trips produced by and attracted to each zone in a study area. This first step predicts total flows or total daily travel for each zone in a study area. Regression models, cross classification, and discrete choice models can be used in modeling trip generation.

2. Trip Distribution Models are used to predict spatial pattern of trip. The models can be called as destination choice model. In this stage, a trip table (origin-destination matrix) is used for showing number of trip ends and trips generated estimated by trip generation models between each zone in the study area. Gravity model and growth factor models are used mainly in this stage.

3. Mode Choice Models predict the percentage of travelers using each of the travel modes for particular types of trips. Another goal is to predict the share of number of trips according to the modes available to the given travelers. Discrete choice models are common methods in mode choice analysis.

4. Trip Assignment Models deal with the assignment of the predicted traffic flow on a network minimum path (all or nothing) assignment, stochastic methods, and congested assignment are commonly used traffic assignment techniques.

The four stage modeling is a sequential decision process. The usage of transportation models successfully encouraged the development of land use models. Lowry's study (1964) is one of the most known models. "*The principle use of a Lowry - type model is to allocate a fixed amount of population and employment to zones of a region, given known locations of some of that employment and the transportation characteristics of the region*" (Horowitz, et al. 2004, 167). The four stage modeling has been criticized recently. In the last twenty five years, integrated land use - transport models, microsimulation models of urban land use - transportation, and activity - based methods have been used and developed. Integrated models aim to analyze urban policies that might have impacts on land use and transportation. In other words, integrated models aim to predict of land use - travel patterns and their interactions (Timmermans 2003). Most of these models are aggregate models, except UrbanSim and urban areas are divided into the zones (Hunt, et al. 2005, Wegener 2004). There is a still considerable interest among planners in integrating land use and transport planning in order to assist in reducing car based travel and obtain sustainable development. In recent years, many land use and transportation models[1] regarding LU&T such as BOYCE, ITLUP, MEPLAN, TRANUS, UrbanSim, and POLIS have been developed. One of the alternative approaches in travel demand modeling is activity - based travel demand modeling. This approach is based on modeling the entire activity associated with trip making instead of the modeling for each trip purpose in the four stage model (Meyer and Miller 2001). Activity - based approach assumes that travel decisions are activity based. The model includes several subclasses of econometric model systems:

[1] Land Use and Transportation Modeling is discussed in detail in Hensher et al. (2004) and Pickrell (1999). Acitivity based approach is discussed in detail in Hensher et al. (2004) and Hensher and Button (2000).

trip - based systems, tour-based systems, and daily schedule system (Bowman and Ben - Akiva 1997).

2.2. Land Use and Mode Choice Relationship

2.2.1. The Pattern of Travel Demand in Developing Countries

For developed countries, travel behaviour and its relationship to land use has been the subject of the debate for urban transportation problems. Some land use policies such as compact development, TOD, and mixed land use are solution to the problems based on high level of private car usage. However, much remains to be learned about how land use characteristics affect travel behaviour for developing countries. Lack of coordination between land use and transportation cause serious transportation problems such as congestion and traffic accidents in developing countries.

Travel behaviour is generally measured in terms of trip time, mode choice, trip length, and route choice in empirical studies. In the content of this study, mode choice is focus of the study. Mode choice behaviour in developing countries are rarely investigated with respect to location (spatial configuration of the cities). There is a lack of empirical studies on this issue (criticized in Table 2.2 – 2.3). From the perspective of developing countries, urban transportation problems can be analyzed under four headings: congestion, deteriorating environment, safety and security, and declining public transportation for the poor people (Gwilliam 2003). These problems are highly based on rapid motorization process. Researchers have focused the effects of motorization process in the literature related to developing countries. For example, Dargay and Gately (1999) estimates the effect of income elasticities (the growth in per capita income) for national car and vehicle ownership for OECD countries and a number of developing economies including China, India, and Pakistan. The study found that car and vehicle ownership for the lower income countries (China, India, and Pakistan) will grow about twice as rapidly as per-capita income. Senbil et al. (2007) found that income has stronger effect on car ownership than motorcycle ownership. Income elasticity was estimated to be 1.75 for the Asian whole region. It means that one percent increase in income causes a 1.75 percent increase in passenger cars per thousand population. Therefore, it seems that income distribution is the most important determinant for explaining motorization.

Gakenheimer (1999) found that cars per 1000 population are positively correlated with the annual income of the top 20% of population of the low income developing countries such as Bangladesh, India, Pakistan, and Colombia. Baker et al. (2005) analyzed the factors affecting the demand for transport services by the poor who live in Mumbai, India. The study found that poor household made fewer trips than wealthier. The main mode is walking for poor households. 66% of commuters in poor households take walk or bicycle whereas over 30% of poor households take rail and bus for commuting. Poor households wanted to shorten travel distance due to high cost and travel time. The highest frequency for commute distance is 1-2 km whereas higher income workers are willingness to travel longer distance. The poor workers are highly commuting by walking while they take rail mode for commute distance with 5 km or more. For the highest income groups, the motorcycles and cars are the main commute mode. In Mumbai, public transit is important factor in mobility for the poor and the middle class. *"rail remains the main mode to work for 23% of commuters, while bus remains the main mode for 16% of commuters. The modal shares for bus are highest for the poor in zones 1-3 (21% of the poor in zone 2 take the bus to work) while rail shares are highest for the poor in the suburbs"* (Baker, et al. 2005, 46).

Hyodo et al. (2005) analyzed urban travel behavior characteristics of 13 cities[2] using by household interview survey data.

> Bicycle trips are biggest in Chengdu, the bicycle being a major mode in China. In Tokyo and Hiroshima, the bicycle is an important access mode to train stations and for short trips. The bicycle is not as important in the other cities due perhaps to the hot weather, culture, and others. About 30% - 40% of all trips is done by "walking" for Belem, Managua, Chengdu, Damascus, and Phnom Penh. The motorcycle is an important mode in KL, Phnom Penh, and Tripoli (Hyodo, et al. 2005, 34).

The World Bank (2002) suggested that most developing countries have fewer than 100 cars per 1000 people, compared with 400 or more per 1000 people in developed countries. The main mode for Hong Kong is public transportation. 48.3% of

[2] The cities are Tripoli (Lebanon), Phnom Penh (Cambodia), Damascus (Syria), Manila (Philippines), Chengdu (China), Managua (Nicaragua), Belem (Brazil), Bucharest (Romania), Cairo (Egypt), Jakarta (Indonesia), KL (Malaysia), Tokyo (Japan), Hiroshima (Japan).

total work trips is made by buses and minibuses while 25% of total work trips were made by Mass Transit Railway (MTR) in Hong Kong in 2001 (Lau and Chiu 2004).

Vasconcellos (2005) analyzed transportation conditions for the years between 1967 and 1997 in Sao Paolo, Brazil. The study found that work trips decreased from 50% in 1967 to 41% in 1997. Regarding the change in the use of motorized transport modes, from 1967 to 1997, the share of private modes (auto and taxi) have increased while public transportation modes (train, subway, and bus) have decreased. However, walking is the main mode for all trips. On the other hand, the number of auto trips per person firstly increased from 1967 to 1977 while the rate has stabilized since 1977. The number of public transportation trips per person increased from 1967 to 1977 while the rate has decreased since 1977. Liu (2006) analyzed travelers' choice behavior for work trip in Shanghai. The study found that income is important variable for work trip mode choice decisions. Individuals with higher income levels tend to commute by taxi more than bicycle and bus.

The rise in population and motorization is common problem for developing countries. For example, in Malaysia, the number of registered motor vehicles increased by 8,321,517 from 1990 to 2003 (Nurdeen, et al. 2007). Although vehicle ownership and usage is growing rapidly, private modes have a lower commuter mode share than public transport modes in developing countries. Public transit is the main mode for urban vehicular trips, approximately 75% of urban vehicular trips (Gakenheimer 1999). In spite of higher use of public transportation, the use of rail modes among public transportation is still lower-level. One of the problems related to urban transportation in developing countries is poor service quality of the public transit (Senbil, et al. 2005, Alpizar and Carlsson 2003). *"Although the vast majority of trips depend on public transportation in most cities services suffer from poor financial conditions, inadequate passenger capacity, low network integration, slow operating speeds, and deteriorating physical conditions"* (Gakenheimer and Zegras 2004, 162). The other one is that most urban public transit is hihghly road based (World Bank 2002). For example, public transportation in the city of Karachi and Pakistan that reached a population 14 million in 2004 is mainly based on road-based. The city is developing without a rail based mass transit system. In China, India, and Malaysia, the automobile sector is the biggest economic sector while in US and Europe policies aims to decrease the share of private modes and motorized travel distance. Nonmotorized (walking and cycling) modes play dominan role as a main mode for all trips in developing countries. For example, the share of walking is between 25 and 50 percent of trips in the major Indian cities and 50 percent of all trips in major African cities (World Bank 2002). On the other hand, in Hanoi (Vietnam) and Ouagadougou (Burkina Faso), motorcycles play a predominant role in 1990s (Vasconcellos 2001). Nowadays, motorization is dominated by motorcycles in Ho Chi Minh City, Vietnam. The share of motorcycles is 78% of journeys in the city. In fact, motorcycle usage have become the major mode due to low cost and effectiveness whereas public transportation is not highly used due to poor service labels and conditions. (Santoso and Tsunokawa 2005). In Asian cities, high levels of motorcycle ownership is common fact because buying a motorbike is cheaper than others (Senbil, et al. 2006). The major modes in the city of Addis Ababa (Ethiopia) are buses and taxis that are used for public transportation (Gebeyehu and Takano 2007). In Asia, the share of motorcycle mode is more than automobiles. In Taiwan, The motorcycle ownership per square kilometer is 302.8 whereas this rate is only 0.4 in America (Lai and Lu 2007).

Gakenheimer (1999) suggest that mobility and accessibility are declining in most of the large cities of developing countries, depending on the high level of congestion. An important issue in transportation in the world is environmental discussions. The most known solutions to this problem that have been hihghly discussed in sustainable transportation are to reduce automobile dependence, to increase the share of public transportation and non-motorized modes, and to develop land use policies such as mixed use, transit-oriented community, and higher density development. For example, Pucher and Renne (2003) examined the variations in travel behavior such as travel mode and mobility levels using by 2001 National Household Travel Survey (NHTS) in US. According to the results of the study, the share of private car for walk trips increased from 66.9% in 1960 to 87.9% in 2000 whereas the share of public transit for the same period decreased from 12.6% to 4.7%. For walk trips, this rate decreased from 10.3 in 1960 to 2.9 in 2000. In total, non-motorized modes (walking and bicycling) as a commuting mode was 3.3% in 2000. Also, auto's share for daily travel (all trip purposes) is high level. The share of auto for daily travel in the United States increased from 81.8 in 1969 to 86.4% in 2001. In the same period, the share of transit decreased from 3.2 to 1.6 while the share of walk mode is 8.6 in 2001. Another important issue is rise in work travel distance. Average travel distance to work in US increased from 9 miles in 1975 to 11.6 in 1995 (Hanson 2004). Average travel time to work in US increased from 21.7 minutes in 1980 to 22.4 minutes in 1990, and to 24.3 minutes in 2000 (Horner 2004). *"In 2001 the average journey to work covered twelve miles and took twenty – four minutes. By 2005 the mean travel time to work in the USA was twenty – five minutes"* (Pacione 2009, 265). Regarding developing countries, the average trip length in Delhi increased from 5.4 km in 1970 to 8.5 km in 1993. The average travel time in the city increased from 30 minutes in 1985 to 44.34 minutes in 1993. The average trip lengths as minutes are 12.40 for Mumbai, 7.30 for Chennai, and 6.70 for Bangalore in 1993. The average trip times (minutes / kilometre) are 33.37 for Mumbai, 21.62 for Chennai, and 17.60 for Bangalore in 1993. According to the statistics for both developed and developing countries, people are willingness to travel longer distances for home - based work trips.

In U.S., important finding for mode choice is that the share of public transit for all trip purposes have decreased. The share of transit mode decreased from 3.2% in 1969 to 1.6% in 2001. The share of walk mode decreased from 9.3% in 1969 to 8.6% in 2001 for all trip purposes. For work trips, the share of public transit in total work trips have declined from 12.6% in 1960 to 4.7% in 2000 in the United States (Pucher 2004). The detailed mode split for developing countries are presented in Table 2.1.

The development and planning in many cities of North America and Europe is integrated with rail transit system. For example, Stockholm is one of the best example for this integration between rail rail-transit systems and urban development. "*Half of the city's 750,000 inhabitants live in satellite communities linked to the urban core by a regional rail system*" (Pacione 2009, 276). In the city, 53 percent of workers live in satellite new towns commute by rail (Pacione 2009). *"In the USA the Bay Area Rapid Transit (BART) system in San Francisco CA carries more than half of all CBD – bound work journeys"* (Pacione 2009, 271).

Srinivasan and Rogers (2005) analyzed differences in travel behavior between two different locations where low-income residents lived in the city of Chennai (India). The one group of households lived close to the city center (Srinivasapuram) while the others lived close to the periphery (Kannagi Nagar). According to the survey, residents highly used non-motorized transport and transit for conducting both work and non-work activity. Also, work trips include 56% of trips made by persons in both locations. The share of work related activity and shopping trips are 19% and 23%, respectively.

Regarding mode choice, In Srinivasapuram, the major mode share was for walk including 69% of trips while in Kannagi Nagar, the major mode was bus including 50% of trips. In sum, location has a significant effect for explaining travel behavior, even for low-income residents of Chennai. *"Poor people typically make only one-third to one-half as many motorized trips per capita as the non-poor"*(Gwilliam 2003, 10).

Table 2.1. Mode split in selected developing country cities
(Source: adapted from the studies of Srinivasan et al. (2007), Srinivasan and Rogers (2005), VTPI (2010), Chang and Wu (2008), Vasconcellos (2005), Zhao (2010))

Cities	**Mode Share (%)**		
	Public Transportation	**Private Modes**	**Non-Motorized Modes**
Karachi (1987)	57	31	12
Karachi (2004)	52	48	0
Chennai (India)	39.53	54.09	6.32
Bangalore	45	29	27
Bishkek	80	12	8
Cebu	96	4	0
Colombo	77	17	6
Mandaluyong	45	39	16
Dhaka	33	6	61
Naga	58	38	4
Phnom Penh	25	70	5
Hohhot	2	6	92
Melbourne	43	56	2
Taipei	30.58	61.06	8.36
Sao Paulo	50.8	47.3	0.9
Yizhuang	23.7	40.6	35.7
Wangjing	20.4	49.7	29.9

More vehicles on the roads and longer commuter distance have created urban transportation problems. The most important debates to challenge transportation problems have been continued on urban commuting associated with land use policies, especially in developed countries. Modeling commuting flows has become important in urban policy and regional science (McArthur, et al. 2010, Ruwendal and Nijkamp 2004). Also, commuting pattern in the cities is one of the main causes of traffic congestion. It is assumed that commuting pattern is an indicator of urban spatial structure (Sohn 2005). Commuting behavior is related to three markets: labor, housing, and transportation. Therefore, it plays important role on urban economic models developed by Alonso (1964), Muth (1969), and Mills (1972) (Rouwendal and Nijkamp 2004).

In addition, some typologies related to the spatial pattern of commuting flows have been developed. Analyzing commuting pattern is described on these typologies. According to a known typology in the literature, commuting flows have five different ways as seen in Figure 2.2 (Plane 1981, Pacione 2009, 266):

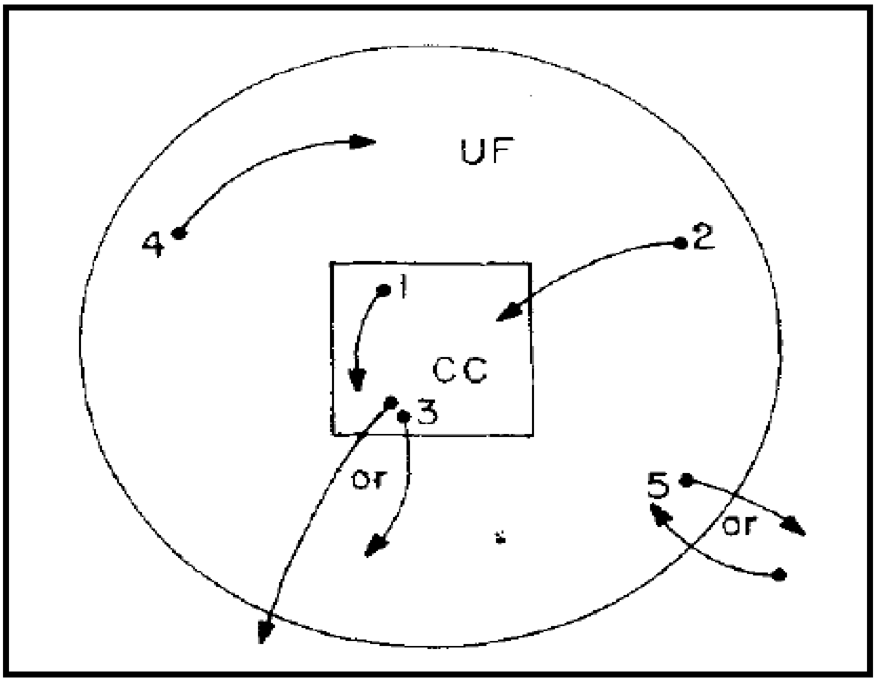


Figure 2.2. Typology of commuting flows
(Source: Plane 1981)

Type 1: within central city movements are trips made by workers who both live and work within the city's legal boundaries.
Type 2: inward commuting encompasses both the traditional commuters from suburbs and metropolitan villages to central cities, plus those workers living in one central city who commute to another.
Type 3: reverse commuting is composed of workers residing in the central city who work anywhere outside that city's boundaries.
Type 4: lateral commuting takes place within the commuter range of the city but both work place and residence locations are outside the central city.
Type 5. Cross-commuting flows are those entering or leaving the central city's commuter zone, meaning that only the workplace or residence is located inside the urban field.

Van der Lann (1998) and Schwanen et al. (2001) used a typology of daily urban systems. According to this typology, there are four types of functional daily urban systems: central, decentral, cross - commuting, and exchange - commuting. In addition, when considering the relationship between spatial structure and commuting behavior, the value of commuting travel time, mode choice, and commuting distance may be potential indicator of the relationship. In this study, mode choice behavior for home - based work trips are analyzed as an indicator of commuting behavior in Istanbul.

2.2.2. Measuring Land Use Characteristics

In existing literature, the factors influencing mode choice are divided into three groups (Wright and Ashford 1989, Ortuzar and Willumsen 2006):

1. The socioeconomic characteristics of traveler (income, car ownership, household structure, etc.),
2. The characteristics of travel mode (travel time, monetary cost, convenience, security, comfort, etc.),
3. The characteristics of journey (trip purpose and time of the day).

It is seen that land use attributes are omitted from this list. Also, the factors as summarized above cannot be expected to capture fully the effects of land use attributes. However, spatial configuration of land use (built environment) may be potential determinant of mode choice. Therefore, an empirical gap has occurred to test whether land use attributes are significant determinant of mode choice and also to what extent. This question brings two important tasks to researchers. One is that how land use attributes can be measured and entered into the models. The other one is to use alternative approaches that better explain the relationship between mode choice and land use may give better results than conventional models.

There is a growing interest in the relationship between built environment and travel behavior in recent years as a way of shaping travel demand. Therefore, some urban planning policies and urban design philosophies such as new urbanism and transit oriented development have gained popularity. These design concepts have common objectives as follows (Cervero and Kockelman 1997):

- Reduce the number of motorized trips (as called trip degeneration),
- Increasing the share of non motorized trips,
- Reducing travel distances and increasing vehicle occupancy levels or encouraging shorter trips or transit, para transit etc.

A key task is to define and analyze the effect of land use characteristics (or built environment) on travel decisions. Although there are several empirical studies investigating the relationship between land use and travel behavior, the results of the studies are mixed about the significance and the extent of the relationship. In this stage, one of first important contributions comes from Kockelman's study. Kockelman (1997) proposed some measures of land use pattern such as accessibility, land use balance, diversity, density, and dissimilarity index of land use mixture. Cervero and Kockelman (1997) suggested that the built environment influence travel demand along three dimensions (3Ds): ***density, diversity, and design***. In addition to this, accessibility measure can be an important indicator of land use pattern and urban form. It is suggested that these dimensions can be used for achieving design objectives. For example, "*underlying the New Urbanist movement is a belief that designing neighborhoods, communities, and regions to be more compact and walkable will result in increased pedestrian activity, increased transit use, and decreased reliance on the private auto*" (Reilly and Landis 2002, 2).

In measuring of land use characteristics (or built environment), population density, employment density, and job density are commonly used variables in the literature (Cervero and Kockelman 1997, Coevering and Schwanen 2006, Buchanan, et al. 2006, Limtanakool, et al. 2006, Newman and Kenworthy 1999, Zhang 2004). Density variable has been one of the most significant land use variables influencing travel behavior. Density is defined as the number of persons (or employment) per hectare. It is measured at metropolitan level in general. Empirical studies suggests that people living in high density areas makes less vehicle travels and they use public transport or walk mode (Maat, et al. 2005). Newman and Kenworthy (1989, 1999) examined this relationship in 46 cities worldwide. They studied auto - oriented land use for urban travel. They found a negative correlation between density and private auto use. They showed this relationship by a logarithmic curve. Coevering and Schwanen (2006) investigated the correlations between land use and travel for 31 cities in Europe, Canada, and the USA. They found that higher population density decrease the share of car trips and increase the share of walking / bicycling. Several critics have challenged Newman and Kenworthy's conclusions (Gomez – Ibanez 1991, Pickrell 1999). These critics suggest that income, household size, gasoline prices, automobile taxation, and transportation technology are related factors to automobile use. To determine the effects of urban density on automobile dependency, one should carefully specify the relationships among density, other important variables (e.g., income), and travel behavior. Automobile use is related not only density, but also income and otfer factors. In the literature, different density measures have been defined such as intersection density, bus stop density, and park and ride density (Frank, et al. 2007), but common density measures are population density and employment (or job) density at trip origins and destinations. In addition to job and population density, other density measures used in empirical studies are worker density (per sq mile number of workers), housing density (per sq mile housing units), road density (per sq mile road length), intersection density (per sq mile number of intersections), and overall density ((residents+jobs)/area) (Lin and Long 2008, Ewing, et al. 2004).

Diversity presents the degree of land use mixture. In other words, it represents spatial heterogeneity. Two indexes are highly used: land use mix (dissimilarity index) and land use balance (entropy index[3]). It assumes that more balance induce transit use and non drive alone travel. Entropy index provide a measure for the degree of balance across land use types (Kockelman 1997). Entropy measure has been used in different settings such as suburban employment centers (Cervero 1989), municipalities of Netharlands (Limtanakool, et al. 2006), Boston (Zhang 2004), Motgomery County, Maryland (Cervero 2002). Greenwald (2006) used housing balance and employment (economic) entropy indices for indicating the degrees to which a transportation analysis zone (TAZ) is in balance in terms of housing stock and diverse in economic activity. In some empirical studies, a different balance measure has been used. This balance is

[3] Entropy index as a land use balance measure is estimated as $\sum_j [P_j \times \ln(P_j)] / \ln(J)$. Where P_j presents the proportion of developed land in the jth use type. J is the number of land use categories. The mean entropy ranges from 0 (homogeneity) and 1 (heterogeneity). The details for this formulation is found in the studies of Kockelman (1997) and Cervero and Kockelman (1997).

called as jobs - housing balance (JHB). Jobs - housing balance represents the spatial relationship between the number of jobs and housing units within a given geographical area (Peng 1997). This balance is a planning tool that the local governments want to achieve a balance between the number of jobs and housing units. "*If planners designed communities with mixed uses, placing some jobs near residences, perhaps many more persons would be able to walk, use transit or carpool to work*" (Boarnet and Crane 2001, 10). The benefits of jobs - housing balance are (SCAG 2001):

- Reduced congestion and commute times,
- Air quality benefits,
- Economic and fiscal benefits,
- Quality of life benefits.

Jobs - housing imbalance (or spatial mismatch) causes to increase long distance work trips, higher automobile dependency, and more vehicle miles traveled (Cervero 1996). Several formulations of measuring the jobs - housing balance have been used in empirical studies. The most used formulation is the ratio of the number of employees to the number of households in a geographical area (Cervero 1989, 1991). Another formulation for jobs - housing balance is formulated as following (Cervero 1996):

$$ratioofjobstoemployedresidents = \frac{numberofwor\ker s\,\mathrm{in}\,thecity}{numberofresident\,s\,\mathrm{in}\,thecitywhoareemployed} \tag{2.1}$$

According to the findings of the empirical studies related to this variable, if jobs - housing balance occurs, people may want to live and work in the same area. It can be expected that long trips would be avoided (Cervero 1989, Sultana 2002, Wang and Chai 2009). For example, Sultana (2002) highlighted the fact that jobs - housing imbalance is an important determinant for longer commuting. The study found that job - rich areas tend to longer commuting times than areas of balanced JHB ratios. Also, employed residents living in housing - rich areas have longer commuting times than areas with balanced JHB ratios in Atlanta. Zhao et al. (2010) found that the jobs - housing balance has significant implications for commuting time in Beijing. Peng (1997) found that there is a non-linear relationship between the jobs - housing balance and commuting patterns in terms of vehicle miles travelled (VMT) per capita and trip length in Portland.

A simple formulation for JHB is presented as below:

$$\frac{|E_i - P_i|}{E_i + P_i} \tag{2.2}$$

Where *E* presents employment size and *P* is the population size at the relative zone. This value ranges from 0 to 1. 1 represents a pure nonresidential area or residential area while 0 indicates a balance between employment and population. Dissimilarity Index (Land Use Mix) as another type of diversity index presents proportion of dissimilar land uses within a tract. The index is based on distinct land use types. Different land use mix formulations can be used. One of the most known types of land use mix formulation is computed by the land use composition as seen in Equation 2.3. It varies between 0 and 1 (Rajamani, et al. 2003, Bhat and Guo 2007).

$$LandUseMixDiversity = 1 - \left\{ \frac{\left| \frac{r}{T} - \frac{1}{4} \right| + \left| \frac{c}{T} - \frac{1}{4} \right| + \left| \frac{i}{T} - \frac{1}{4} \right| + \left| \frac{o}{T} - \frac{1}{4} \right|}{\frac{3}{2}} \right\} \tag{2.3}$$

Where $T = r + c + i + o$, and *r* represent zonal hectares in residential use, *c* is zonal hectares in commercial use, *i* is zonal hectares in industrial use, and *o* is zonal hectares in other uses. A value of 0 means the land in metropolitan area has a single use and a value of 1 represents perfect mixing among land uses.

Design variables are associated with site, street, and block design in a neighborhood. For example, Cervero and Kockelman (1997) measured the variables of street design as predominant pattern of the street such as regular grid, proportion of intersections, number of blocks, number of dead ends and cul de sacs. On the other hand, site design variables were measured by proportion of commercial, retail, and service parcels with front and site lot parking. Also, under design category, one measure group is related to pedestrian and cycling provisions. They are proportion of blocks with sidewalks, street trees, bicycle lanes, and proportion of intersections with signalized controls, averages of block face length, sidewalk width, and bicycle lanes per developed acre. Ratio of sidewalks miles to road miles can be used for design variable (Cervero 2002). Since urban design philosophies aims to stimulate the use of public transportation, urban form characteristics may affect the choice of travel mode (Cervero and Gorham 1995, Frank and Pivo 1994). For example, Snellen et al. (2002) studied neighborhood characteristics including urban form typologies, transportation network types, and local - street network type for the cities in The Netherlands. The measuring of these design variables needs parcel level and Geographic Information System (GIS) data. It is very difficult to obtain design variables for the cities in developing countries.

Accessibility has long been identified as a key factor in urban theory. The previous studies by Alonso (1964), Muth (1969), and Mills (1972) have modeled a mono - centric city. Theory assumed that all the employment took place at the city center. Commuting time would be key determinant of the city rent curve. For example, savings in commuting time can be measured by monitoring the changes in a city rent curve. Many empirical studies have analyzed the effects of accessibility based on transportation investments in the city (detailed discussion Celik and Yankaya 2006, Yankaya 2004). In empirical studies, different accessibility measures have been used such as regional accessibility measure, recreation accessibility (Pinjari, et at. 2007), job and labor force accessibility (Cervero 2002, Cervero and Kockelman 1997) or proximity/distance variables to urban centre or transportation infrastructure such as a nearest transit station (Limtanakool, et al. 2006, Stead 2001, Zhang 2004). A common accessibility index in the studies is estimated as follow (Kockelman 1997):

$$Accessibility = \sum_j \frac{A_j}{f(t_{ij})} \tag{2.4}$$

Where A_j is attractiveness of zone j and t_{ij} is travel time from zone i to j. Another known form of the accessibility index is based on gravity type functional form.

$$A_{im} = \sum_{j}^{J} f(C_{ijm}) * R_j \tag{2.5}$$

In this equation, $f\,(C_{ijm})$ represents friction factor between zones i and j by mode m. R_j is employment in zone j, while J is the total number of travel zones (Levinson and Kumar 1995, Rajamani, et al. 2003).

2.2.3. Empirical Applications for Measuring The Influence of Land Use on Mode Choice

Interest in analyzing travel behavior has undergone considerable development in recent years. Table 2.2[4] represents several empirical studies focused on the relationship between and land use. In spite of growing interest and voluminous empirical literature, many issues needs to be explored. The main issues are summarized as follows:

1. There is no consensus about the findings for the relationship between land use and mode choice. For example, Cervero and Kockelman (1997), Cervero (2002), and Zhang (2004) found that land use has an independent influence on mode choice while Crane and Crepeau (1998) and Rodriguez et al. (2006) did not find enough evidence. Although some studies found a correlation between land use and mode choice, questions remain regarding strength and direction of the relationship. Another issue is that which land use characteristics influence travel behavior has not been adequately explained.
2. In existing literature as summarized in Table 2.2 and Table 2.3, many empirical studies have been conducted in North-American and European cities (Frank and Pivo 1994, Frank, et al. 2007, Cervero and Kockelman 1997, Kitamura, et al. 1997, Kockelman 1997, Cervero and Duncan 2002, Bhat and Guo 2007, Rajamani, et al. 2003, Cervero 2002, Chen and McKnight 2007, Zhang 2004, Crane and Crepeau 1998, Limtanakool, et al. 2006, Coevering and Shwanen 2006). In Hong Kong (Zhang 2004) and Asia (Lin and Yang 2009), there is also enough evidence to support the hypothesis. In line with previous studies for Turkey, the effects of land use on mode choice has been ignored. There is no evidence of any significant relationship to support the main hypothesis of the thesis.
3. In existing literature, evidence derived from empirical studies belongs to either aggregate or disaggregate analysis. Empirical studies combining and analyzing

[4] The more detailed discussion is included in TRB Special Report 282 (2005).

aggregate and disaggregate data together are rather limited. The effect of land use variables may change the scale of analysis.

4. In many studies, land use variables were not tested independently from other factors. Cervero (2002) and Zhang (2004) tested the marginal influence of land use and model's explanatory power by an expanded model.

5. Aggregate behavior is a result of individual choices in zones. The modeling aggregate choice behavior is highly related to individual choice. In the basis of individual choice theory, all decisions are probabilistic. According to the type of choice data, probability models may be applied to aggregate or disaggregate data. The models calibrated with disaggregate data is used to explain individuals' behaviors while the aggregate models analyze to predict the zonal shares of trips by different travel modes. Contrary to the disaggregate models, the aggregate models require characteristics of travel zones (average auto ownership, average income, etc.) and characteristics of o-d pair such as travel time. In existing literature, academic research is still heavily focused on disaggregate modeling for analyzing travel behavior (Zhang 2004, Cervero 2002, Pinjari 2007). These empirical literature stems from works of Domencich and McFadden (1975) and Ben Akiva and Lerman (1985). The disaggregate modeling is still widely preferred.

Empirical analysis of mode choice is mainly based on discrete choice model (Domencich and Mc Fadden 1975, Ben Akiva and Lerman 1985). The random component in MNL model is assumed to be independent and identically distributed with Gumbel distribution (McFadden 1974). However, the assumption of independence of irrelevant alternatives is an important restrictive for the application of discrete choice models (logit and probit models) into the modeling of choice behavior. Generalized extreme value models such as Nested Logit Model relax IIA assumption. Multinomial logit and conditional logit models are the most used and preferred probabilistic choice models up to now. Since probit models need computational effort, logit models have been used increasingly in mode choice studies, especially with disaggregate data. Despite some empirical evidence, traditional models cannot adequately exhibit this complex relationship. For example, Lin and Yang (2009) suggest that structural equation modeling is an appropriate technique for analyzing complex systems. On the other hand, discrete choice models suffer from some statistical assumptions such as independence of irrelevant alternatives (IIA). Soft computing methods do not suffer this assumption. Alternative approaches may be more flexible than discrete choice models. Also, in mode choice studies, alternative models are less used with aggregate data.

There has been little empirical attention paid to analyze the effect of land use characteristics on mode choice by using soft computing methods. Among the soft computing methods, artificial neural networks (ANN) have been widely used in mode choice studies while genetic algorithm (GA) and bayesian belief networks (BBNs) are less used methods. New algorithms in soft computing methods can be tested to increase model performance.

One of the important issues in empirical studies is residential self - selection factor. It assumes that some households may prefer to live a neighborhood with good transit service facilities. Cervero and Duncan (2002) analyzed self - selection factor by constructing a nested logit model in San Francisco Bay Area. The study found that residential location and commute choice are jointly related decisions among station-area residents.

Empirical studies in general analyzed the effects of land use on travel behavior in metropolitan areas. There are quite a few studies that have been done in small areas such as neighborhoods (Crane 2000, Pan, et al. 2009, Lin and Long 2008). Therefore, neighborhood characteristics may play an important role on travel behavior. However, neighborhood refers a spatial unit. In the literature, land use generally refers to built environment for various functions such as residential, commercial, industrial, natural areas while urban form includes design of the city.

To seperate out the influence of land use characteristics on mode choice, the effects of socioeconomic and travel characteristics should be analyzed independently of land use characteristics. Multivariate analyses may allow analyst to do so. In sum, despite the significant accumulation of empirical studies, many issues require further empirical studies including new models.

Table 2.2. Summary of the literature review related with mode choice and land use relationships

Case Study	Data Type	Land Use (or Built Environment) Variables Tested	Relationships significantly	Empirical Model
Cervero and Kockelman (1997), San Francisco Bay Area.	Disaggregate Data, OD based	Population density, employment density, accessibility to jobs, dissimilarity index.	- Mixed use and pedestrian friendly designs encourage non-motorized travel.	Binomial logit for work and non-work trips.
Cervero (2002), Maryland.	Disaggregate Data, OD based	Gross density, job accessibility land use diversity, ratio of sidewalk miles to road miles, labor force accessibility		Binomial logit and multinomial logit for all trip purposes.
Cervero and Duncan (2002), San Francisco Bay Area.	Disaggregate Data	Workplace distance to rail station, job accessibility index, neighborhood density		Nested logit for commute trips.
Zhang (2004), Boston and Hong Kong.	Disaggregate Data, OD based	Distance to nearest train station, population and job density, % non-culde sac, land use balance, public parking supply.	- Land use has an independent influence on mode choice. - Goodness of fit of the models improved after the inclusion of land use variables.	MNL model and Nested Logit Model for hbw trips
Limtanakool et al. (2006), Randstad, Holland.	Disaggregate Data, OD based	Population density, land use balance, local and national specialization index for services and urban center (core cities or suburban, type of municipality, availability of a train station.	- Population density and the provision of transport services have a statistically significant effect on mode choice. - Commuters are more likely to travel by train when traveling to a workplace with consumer services, urban facilities, and other activities nearby.	Binary logit model for commute, business, and leisure trips.

(cont. on next page)

Table 2.2. (cont.)

Buchanan et al. (2006), Christchurch New Zealand.	Aggregate Data, OD based.	Population density, Employment density, Distance from the CBD.	- Population density was not statistically significant variable as was distance from CBD.	Stepwise multiple regression
Pinjari et al. (2007), San Francisco Bay Area.	Disaggregate Data, OD based.	Household density, employment density, land use mix, recreation accessibility, street block density.	- Built environment attributes can indeed significantly impact commute mode choice behavior.	Joint flexible econometric model
Frank et al. (2007), Central Puget Sound (Seattle).	Disaggregate data, OD based.	Bike and transit intersection density, land use mix, retail area floor area.	- Land use mix, retail density and street connectivity measures proved significant for modes.	Logit model for home-based work and home-based other trips.
Lin and Long (2008), the cities in USA.	Aggregate data.	Neighborhood type (Urban elite, rural, suburban wealthy, etc.)	- Transit availability at place of residence tends to increase the transit mode. - Urban residents made higher percentages of transit, walk, and bicycle trips than the suburban and rural counterparts.	Descriptive analysis, ANOVA, hierarchical modeling.
Rajamani et al. (2003), Portland.	Disaggregate and Aggregate data, Non OD based.	Land use mix, park area, accessibility index, population density, percentage of culde sac.	- Mixed use planning promotes walking behavior. - Traditional neighborhood street design encourages walking mode.	MNL for non work
Çelikoğlu (2006), Istanbul.	Aggregate data, OD based.	-	-	Binary logit.
Bonnel (2003), France.	Aggregate Data OD based.	Density (population + jobs) of zone		Binary logit model

(cont. on next page)

Table 2.2. (cont.)

Schwanen and Mokhtarian (2005), San Francisco Bay Area.	Disaggregate data	Neighborhood indicators.	- Neighborhood type dissonance is statistically significantly associated with commute mode choice.	MNL model for commute mode choice.
Schwanen et al. (2001), The Netherlands.	Disaggregate data	Urbanization levels (core cities, suburbs, and growth centers)	- Deconcentration of urban land use to suburban locations and new towns almost certainly promotes the use of the private car for all purposes. It also leads to less use of public transport as well as of cycling and walking. - Decentralized and exchange commuting urban systems seem to promote public transport and biking.	MNL model for work, shopping, and leisure trips.
Coevering and Schwanen (2006), The major cities in Europe, Canada, and the USA.	Aggregate data	Population density, employment density, ratio of transit to road supply, parking places in CBD, population centrality.	- A higher population density is associated with a smaller share of the car and more walking/bicycling. - A good provision of public transport relative to road infrastructure and a lower number of public parking places in the CBD stimulate the share of public transport commutes.	Multiple regression models.
Lin and Yang (2009), Taipei, Taiwan.	Aggregate data	Building and emp. density, Housing-Job, Housing-Retail, road density, grid network, bus stop density, distance to metro station, transit, motorcycle, and car access.	- Density is negatively associated with car use (private modal split). - Mixed land use increases private modal split and a pedestrian – friendly built environment significantly reduces private modal split.	Structural equation modeling
Cervero and Gorham (1995), San Francisco and Los Angeles.	Aggregate data	Residential density and neighborhood type.	- Neighborhood type (1:Transit and 0:Auto) is a significant predictor. - Density has a significant effect on transit commuting in both transit and auto neighborhood.	Matched pair analysis (regression analysis) for commuting trips.

Table 2.3. Summary of the literature review related with mode choice and land use relationships for developing countries

Case Study	Data Type	Land Use (or Built Environment) Variables tested	Relationships significantly	Empirical Model
Hamed and Olaywah (2000) Amman, Jordan.	Disaggregate	Travel distance to work, Station distance, and work and home location	Bus, servis taxi, and private car commuters have different travel patterns.	MNL model for commuters' decisions.
Srinivasan and Rogers (2005) Chennai, India.	Disaggregate	Location variable (close to the city center or not)	- Differences in accessibility to employment and services have a strong effect on travel behavior. - Residents in the centrally located settlement were more likely to use non-motorized modes for walk and bicycle travel than the peripherally located residents.	Binary Logit Model for mode choice and trip frequency.
Wan et al. (2009) Huaibei, China.	Disaggregate O/D based.	Residential density, commercial use ratio.	In the higher residential density and commercial use ratio areas, the possibilities of commuters use public transport and motorcycle mode increase.	MNL model for commute mode choice.
Zhao et al. (2010) Beijing, China.	Disaggregate	Jobs-housing balance, population density, and transportation infrastructure-based accessibility.	- Jobs-housing balance has a statistically significant and negative relationship with commuting time. - High and middle population density have a negative effect on commuting time, but not significantly.	Multiple Linear Regression for workers' commuting time.
Zegras (2010), Santiago de Chile.	Disaggregate O/D based	Live in apartment, Dwelling unit density, Diversity index, 4-way intersections per km, Distance to CBD.	- Diversity index has a negative effect on household motor vehicle choice. - Households living further from the CBD have a higher likelihood of owning motor vehicles. - Dwelling unit density, diversity index, and 4-way intersections per km did not play significant role on automobile use. - Distance to CBD and metrostations have a strong association with vehicle use.	MNL model for household motor vehicle choice and Regression model for detecting the effects of built environment on automobile use.
Senbil et al. (2007) Jabotek (Indonesia), Kuala Lumpur (Malaysia), and Manila (Philippines).	Disaggregate	Distance to city center, land use diversity, ratio of commercial / residential /undeveloped land use, residential density, job density, length of all roads.	Density variables have not significant effect on motorcycle ownership levels, but the ratio of commercial land use have. Road supply has a significant effect on car ownership.	Bivariate ordered probit model of household motocycle and car ownership.

Boarnet and Crane (2001) proposed a typology (see Table 2.4) for organizing the empirical studies that have focused on travel and land use. In empirical studies, different types of travel outcome measures can be used for dependent variable: trip frequencies (number of trips), total miles traveled, mode choice, commute length, cumulative person miles traveled, trip lengths (distance and time). Empirical studies as reviewed above have some limitations in analyzing land use and mode choice. Many studies included some land use measures, but they did not allow separating out the effect of land use on travel from socioeconomic and travel variables. A few studies have recognized the magnitude of land use effects and tested the variations of land use effects on mode choice for different travel purposes (Cervero 2002, Cervero and Kockelman 1997, Zhang 2004). Zhang (2004), Cervero (2002), and Cervero and Kockelman (1997) found that land use has an independent influence from travel time and monetary costs on mode choice.

Table 2.4. A typology for the relationship between urban form and travel (Source: adapted from Boarnet and Crane 2001)

Travel Outcome Measures	**Urban Form and Land Use Measures**	**Method of Analysis**	**Other Distinctions and Issues**
1. Total miles traveled (e.g., vehicle miles traveled)	1. Density	1. Simulation	1. Land use and urban design
2. Number of trips	2. Land use mix	2. Description of observed travel behavior in different settings (e.g., commute length by city size)	2. Composition of trip chains and tours
2. Car ownership	3. Traffic calming		3. Use of aggregate versus subject-specific data
3. Mode	4. Street and circulation pattern	3. Multivariate statistical analysis of observed behavior	
4. Congestion			
5. Commute length	5. Jobs – Housing Balance and / or land use balance		
6. Other commute measures (e.g., speed, time)	6. Pedestrian features		
7. Difference by purpose (e.g., work vs. non-work)			

Method of analysis is categorized under three groups: simulation, descriptive, and multivariate techniques. Among these, simulation methods cannot provide guidance about the effects of land use on travel behavior. Multivariate analysis includes regression and logistic regression analysis. As seen in Table 2.2, logit models (or logistic regression) are widely used in mode choice studies. Binary Logit and Multinomial Logit models have been the most preferred methods (Pan, et al. 2009, Srinivasan 2002, Vega and Reynolds-Feighan 2008). Empirical analysis of home - based work (or commuting) trips can be performed by means of a Nested Logit and Probit models (De Palma and Rochat 2000, Cervero and Duncan 2002, Zhang 2004, Vega and Reynolds-Feighan 2006, Jou, et al. 2010). In addition to these approaches, structural equation modeling, soft computing methods, and activity based models are alternative methods for analyzing the relationship between travel demand and land use. Model calibrations are generally based on OD based data. Studies analyzing mode choice empirically collect data from several sources. Empirical analysis may focus on different trip purposes: work and nonwork trips (e.g., home - based school and home - based other). The analysis of commuting trips is dominant in existing literature (Limtanakool, et al. 2006, Zhang 2004, Cervero 2002).

Main data source are generally trip records drawn from household travel survey data, census, regional inventories, and field surveys. For example, Cervero and Kockelman (1997) used a digital database which belongs to the Association of Bay Area Governments on dominant land uses for hectare grid cells in San Francisco region. Travel surveys provide much information about variables. Travel data (e.g., mode, trip length), personal data (e.g., age, gender, education), and household data (e.g., income) can be obtained from travel surveys. Also, travel data may include information about geographical location of origin and destination of all trips. In some studies, quasi - experimental design data can be used (Snellen, et al. 2002). The origin and destination locations derived from travel survey data can be matched and integrated using with GIS based land use database. Land use attributes are measured in defined buffer zone such as one kilometer area. Design data obtained from field surveys (e.g. block length), and regional maps (e.g. proportion of intersections). Household travel surveys are generally cross-sectional data that presents the information about household’s characteristics at the same point of time.

The mode split studies are made by trip purposes: commute, business, leisure etc. Under the category of trip purposes, the probability of an individual choosing an alternative (travel modes) is used as dependent variable of the models. In this situation, common research question is how the effect of land use may vary for different trip purposes? Some studies focused only the effects of urban form on travel behavior for work and nonwork trips (Pan, et al. 2009, Rajamani, et al. 2003).

Zhang (2004) found that the inclusion of land use variables into the mode choice models improved the goodness of fit of the models. In Boston for work trips, higher population densities at trip origin and destination is positively correlated with commuting by transit or non - motorized trips while for non work trips, population density is not significant factor. Increasing in employment density is positively relationship taking non driving modes. However, this variable is not significant for people's decisions for work trips in Boston. In Hong Kong, higher population and job densities at origin and destination increase the share of transit and nonmotorized modes for commuting trips. Job density is significant in Hong Kong while population density is not. In Boston, entropy of land use balance had no influence on mode choice for work trips.

Chen et al. (2008) examined the effects of density in mode choice decisions in home - based work trips, using the data collected in the New York Metropolitan Region. The study used two - equation system. The study found that employment density at work is more important role than population density.

Lin and Long (2008) used five travel measures: number of trips per household, mode share, average travel distance and time per trip, and vehicle miles of travel (VMT) to compare 10 different neighnorhood types on household travel and vehicle use. They found that transit availability increase transit mode shares regardless of household automobile ownership and income level, job - housing tradeoffs. Urban residents choose transit, walk, and bicycle trips more than suburban and rural counterparts.

Buchanan et al. (2006) found that as the city has expanded, the effects of urban structure upon model choice have become important factor. In this study, distance from the CBD play significant role for predicting modal split. Population and work density was not strong variable as did distance from the CBD. Limtanakol et al. (2006) found that travelers living in high density areas tend to use the car less frequently in The Netherlands. Population density and availability of railway stations at origin and destination have a statistically significant effect on mode choice for work trips. At the destination point, land use balance and density is positively correlated with train use. Ewing et al. (2004) found that density and job mix were not significant in choosing travel mode to school. Rajamani et al. (2003) found that mixed uses and higher residential densities encourage walking and transit mode for nonwork travel.

Lin and Yang (2009) studied urban form impacts on travel demand using structural equation modeling in Taipei at aggregate level. They found that density is negatively correlated with private mode split. Mixed land use increases private mode split whereas a pedestrian friendly built environment significantly reduces private mode split. Jou et al. (2010) used multinomial probit modeling for analyzing commuters' mode-switching behavior from private transport to public transport in Taipei. The study found that private commuters were more likely to switch to mass rapid transit than to bus and that auto commuters are generally more likely to switch to public modes than are motorbike commuters. If commuter homes are far away from workplaces, commuters are not likely to switch to public modes due to higher commuting time.

Pan et al. (2009) studied the influence of urban form on travel behavior in four neighborhoods of Shangai using logistic regression for work and non-work trips. They found that urban form affects travelers' choice after the effects of socioeconomic characteristics are controlled. For example, pedestrian / cyclist friendly urban form increase the choice of non-motorized trips. Srinivasan (2002) examined the effects of neighborhood characteristics on mode choice for work and non-work tour using multinomial logit model in Boston. Commercial residential mix and balance are statistically significant and positive for non-auto trips in the work tour.

Vega and Reynolds-Feighan (2008) examined that how the spatial distribution of employment affects travel behavior in Dublin region across the sub-centers using binary logit model at aggregate level. Employment density is negatively correlated with car use and significant. Demand for car and public transport depends on the spatial distribution of employment. Travel attributes (time and cost) have an important effect on the choice of travel mode. An interesting development is that increase in sub-employment centers tend to switch from public transport to car use due to low transport costs (Vega and Reynolds-Feighan 2008).

Cervero and Wu (1997) studied the influence of land use environments on commuting choices in U.S. metropolitan areas using the 1985 American Housing

Survey. They found that neighborhood densities have a stronger influence than mixed land uses except for walking and bicycling.

Abane (2010) examined travel behaviour of the commuters in four metropolitan areas in Ghana at disaggregate level using by Multinomial Logit Model. According to the empirical data, in all the metropolitan areas, the most frequently used modes are trotro[5] (71.4%) and taxis (15.9%). Commuters are more likely to choose trotros and taxis due to perceived good behaviour of drivers and the availability of these modes.

Zegras (2010) aimed to answer the question: "What role might Santiago's built environment play in household automobile ownership and use" using by Multinomial Logit Model. The study found that income play important role on the household vehicle ownership decision. Regarding built environment characteristics, household in the zones with a higher diversity index have a lower probability of owing vehicles. A more gridded street has a negative effect on owning motor vehicles. For household automobile use, distance to the metro stations significantly affect household auto use. Living within 500 metres of a metro reduces car ownership. Dwelling unit density, diversity index, and four-way intersections per km have not significant effect on automobile use.

Kutzbach (2009) examined the motorization process (car and bus) in developing countries. The results of the study suggested that income inequality may increase motorization at low income scales, and reduce motorization at higher income scales. According to the study, this result in abrupt variations for motorization. Population growth and commute distance increase car use and rapid motorization.

Srinivasan et al. (2007) investigated mode choice decisons among commuters in the Chennai city in India. The study found that individuals with vehicles are much more sensitive to travel times of public transportation modes. For short work trips including travel distance lower than 8 km, the sensitivity to public transportation costs is largest among all modes. If work distance increases beyond 8 km, the sensitivity to two-wheeler cost declines by more than two-fold. It means that a unit change in cost variable has a smaller influence on mode choice.

[5] Trotro is an inexpensive public transportation (public minibuses) in Ghana for short and long journeys.

Alpizar and Carlsson (2003) studied the determinants of mode choice decisions for work trips in Costa Rica. The study found that travel time (bus and car) and travel cost for car alternative are the most important determinants of mode choice.

Gebeyehu and Takano (2007) found that increase in income tends to decrease the choice of bus mode whereas increase in household size tend to increase the probability of choosing buses. Higher waiting time for bus increase the probability for choosing a taxi. Bus frequency is the most important determinant of public transportation mode choice.

Hamed and Olaywah (2000) analyzed the factors that influence the commuters' travel related decisons (the morning departure time to the workplace and type of after work activities). The results suggest that travel distance to the work has a significant influence on commuters' departure time decisons. Increase in the distance to the work place affect bus and servis taxi commuters to depart early. Home and work locations have differential impacts on commuters' morning departure time decisons and type of after work activity.

Wan et al. (2009) analyzed the impact of land use variables on commute trip mode choice in China. After the inclusion of land use variables, model performance improved. Increase in residential density in origin encouraged the commuters choose public transport and motorcycle mode whereas increase in commercial use ratio at the origin increase the share of the same modes.

Zhao et al. (2010) analyzed the impact of the jobs - housing balance on urban commuting in Beijing using by multiple linear regression. The study found that jobs-housing balance has a statistically significant and negative effect on individual worker's commuting time. In other words, increase in JHB reduce reduce commuting time. The effect of this variable on commuting time is stronger than population density.

Wang and Chai (2009) analyzed the differences in commuting behavior between the commuters living in houses provided by Danwei and those living in houses in Beijing, China. The commuters for Danwei are more likely to be working and living in the same district. They rely on non-motorized modes. The study suggests that more balanced jobs - housing balance cause shorter commuting trips and icrease in the usage of non-motorized modes.

Alpkokin et al. (2005) analyzed the impacts of polycentric employment growth on urban commuting pattern in Istanbul using by travel surveys during the years 1985 and 1997. According to the results of the study, commuting times and average morning peak hour trip time declined over this period due to opening the second Bosporus Bridge and the multicentric growth of the city. The car usage in 1997 for the employment centers in Istanbul ranged from 38% in Eminonu to 45% in Sarıyer.

Kaldo (2005) examined the relationship between urban density and car usage for commuting trips in the cities that is densely built-up areas, in Japan. The main mode for commuting trips is car including 45.4% of residents. 33.2% of residents used motorized modes (bus, train and other types) whereas 9.5% of residents is walking mode. The study found that there was a strong correlation between driving to work and population density. In other words, people who live in the cities with lower population densities were more likely to take car journeys to work.

Senbil et al. (2006) examined the effect of land use characteristics on motorcycle ownership and its use in Jabotabek metropolitan area in Indonesia at disaggregate data using the tobit model and the ordered probit model. The study found that the ratio of commercial land use and land use diversity decrease motorcycle use while socioeconomic and demographic characteristics promote motorcycle ownership and its use. Also, it is found that the supply of public transport decrease motorcycle use. Regarding transportation system characteristics, accessibility to rail station and road supply increase motorcycles ownership. Distance from the city center has negative effect on motorcycle ownership and its use.

Senbil et al. (2009) studied the relationship between residential location, vehicle ownership, and mobility in two metropolitan areas of Asia, Kei-Han-Shin area of Japan and Kuala Lumpur area of Malaysia using structural equation modeling. The study found that land use mix decreases auto ownership in Kei-Han-Shin. For Kuala Lumpur, public transit access increases auto ownership. Households with more autos in Kei-Han-Shin are located away from the city center. Bicycles generally are used for shopping and to access public transit.

The empirical studies discussed in this section have some limitations. Firstly, several studies have used typical logit formulation: multinomial and binary logit models. The analyzing of mode choice at aggregate level with land use characteristics has not paid enough attention. Second, the magnitude of land use effects still remains unexplored at zonal level although a few exceptions at disaggregate level exist (e.g. Cervero and Kockelman 1997, Zhang 2004). Furthermore, empirical studies have rarely focused on mode choice problems in the cities of developing countries. Making progress in handling these limitations, an alternative approach is proposed to classical logit models. The relationship between land use and travel mode choice is investigated with baseline category logit models at aggregate level.

2.3. Alternative Approaches to Discrete Choice Models: Soft Computing Methods

The use of soft computing methods in the field of transportation is rather new and unexplored in comparison with discrete choice models. Most of the soft computing applications have been based on fuzzy logic and neural networks. The share of the empirical studies that are based on fuzzy logic and neural networks in traffic and transportation studies is around 72% in 2004 (Avineri 2005). Among soft computing methods, Bayesian belief networks are rarely used in transportation modeling. In this section, soft computing methods and Bayesian belief networks in mode choice modeling are discussed.

2.3.1. Soft Computing Methods in Travel Demand Modeling

The presented study intends to compare performance of mode choice models. Discrete choice models, especially logit models have been the workhorse for empirical analysis. However, soft computing methods have emerged as an alternative approach to conventional models in travel demand modeling and transport economics, over the last 15 years. Relative literature suggests that soft computing methods may need less information about problem domain. However, they may give more information and better model performance than conventional approaches. For this reason, soft computing methods can be more suitable and robust models than conventional models. In this part of the literature review, soft computing literature in mode choice modeling has been discussed over empirical studies. These studies represented in Table 2.5 have been pioneer of soft computing approaches to conventional models in mode choice modeling. The important point is that the research question, *how land use attributes affect mode choice*, generally has been ignored. In other words, the potential effects of land use characteristics generally are still ignored.

Nijkamp et al. (1996, 1997) analyzed the impact of high speed train in Italy using logit and neural network model at aggregate level. Nijkamp et al. (2004) studied interregional European freight transport flows by comparing discrete choice models (logit and probit) and the neural network at aggregate level. Abdelwahab and Sayed (1999) introduced neural networks to behavioral choice modeling to analyze U.S. freight transport market at disaggregate level. Hensher and Ton (2000) compared neural networks and nested logit models for commuter mode choice at disaggregate level in the Australian cities. They did not find enough evidence to recommend that ANN is better than Nested Logit models. Cantarella and Luca (2005) analyzed mode choice for commuting trips within the Italian region of Veneto using Multi Layer Feed Forward Network (MLFFN) and random utility models (multinominal and nested logit models) at disaggregate level. Vythoulkas and Koutsopoulos (2003) studied modeling discrete choice behavior using fuzzy set theory, approximate reasoning, and neural networks in The Netherlands at disaggregate level. Celikoglu (2006) studied radial basis function neural network and generalized regression neural network in Istanbul using only time and cost input variables at aggregate level for home-based work (HBW) trips. Xie et al. (2003) compared the capability and performance of data mining methods (decision trees and neural networks) and multinomial logit (MNL) models for work trips in San Francisco Bay Area at disaggregate level. Demir and Gercek (2006) studied mode choice behavior in urban passenger transportation using with soft computing methods (fuzzy logic, neural networks, and neuro-fuzzy logic) and binary logit in Eskişehir. Torres and Huber (2003) performed BBNs to trip generation as a function of socioeconomic variables for home - based work trips at disaggregates level using with 1996 Dallas Household Travel Survey. The study used found that accessibility variables have causal links with the trip generation variables. Janssens et al. (2006) examined the predictive capabilities of decision tree and Bayesian networks for modeling individual choice in The Netherlands. Scuderi and Clifton (2005) investigated the relationship between mode choice and land use using with BBNs in Baltimore metropolitan area at disaggregate level. The study found that the strongest relationships for mode choice are the availability of a private car, the driver status, age, and how empty the land-space looks around the point of origin. Household size, income, and number of commercial spaces are the least influential variables associated with mode choice. The performance of BBNs in the study was not measured.

Empirical studies mentioned above found that soft computing methods outperform conventional models. On the other hand, Hensher and Ton (2000) did not found enough evidence about which approach is better. Xie et al. (2003) found that data mining methods (decision tree and neural networks) are slightly better performance than MNL. Nijkamp et al. (2004) found that the predictive performance of ANN is higher than that of logit model. Cantarella and Luca (2005) found that ANN outperformed random utility models. Celikoglu (2006) found that the performance of neural networks is higher than multivariate linear regression. In the majority of these studies using alternative approaches, land use variables were omitted from the input variables and travel characteristics (time and cost) only entered into the models. Also, neural networks, fuzzy logic, and hybrid approaches are common models in travel demand modeling. Different algorithms and hybrid approaches can be tested in future studies. Therefore, better performance and low error term can be obtained. Bayesian belief networks are one of the alternative methods that rarely used in mode choice modeling.

Table 2.5. Literature review of empirical studies employing soft computing methods used in mode choice

Case Study	Data Type	Empirical Models Compared	Variables	Land Use Characteristics
Hensher and Ton (2000), in six Australian cities.	Commute Mode Choice, Disaggregate Level.	Artificial Neural Networks (ANN), Nested Logit Models.	Travel Characteristics (Cost and Time), Socieconomic and level of service (LOS) attributes, and ASC.	Not included.
Vythoulkas and Koutsopolos (2003), in The Netherlands.	Analyzing choice behavior between rail and car, Disaggregate Level.	Fuzzy Logic, Neuro-Fuzzy, and Binary Logit Models.	Cost, Time, and Rail Access Time.	Not included.
Nijkamp, Reggiani, and Tritapepe (2004).	European Freight Flows, Aggregate Level.	ANN, Probit, and Logit Models.	Distance and Cost	Not included.
Cantarella and Luca (2005), two cases in Italy.	Commuter trips. Disaggregatye Level.	ANN and MNL Models.	Travel Characteristics (Cost and Time), socieconomic and level of service (LOS) attributes, ASC	Whether destination zone is inside the urban center or not (only used in logit models).
Celikoglu (2006), in Istanbul.	Home-based work trips. Aggregate Level	Neural Networks, Linear Regression, and Binary Logit Models.	Time and Cost.	Not included.
Demir and Gercek (2006), in Eskisehir.	Mode choice for different income group, Disaggregate Level	ANN, Fuzzy Logic,Neuro-Fuzzy, and MNL Models.	Time, Cost, and Socioeconomic Attributes.	Not included.
Scuderi and Clifton (2005), in Baltimore metropolitan region.	Disaggregate Level	Only Bayesian Belief Networks.	Socioeconomic Characteristics.	Population density, road density index, commercial, industrial, vacant land rates.

2.4. The Methods for Performance Comparison of Mode Choice Models

The measure of the ability of a statistical model how well it fits observed data is goodness of fit statistics that are quantitative indicators for the difference between observations and predictions. Goodness of fit statistics provides a useful comparison of the accuracy with two or more models (Fotheringham and Knudsen 1987). Each statistical model may include different goodness of fit statistics. In the content of the study, different statistical models are estimated at aggregate and disaggregate levels. Also, one of the hypotheses is that soft computing methods are superior to logit models in mode choice modeling at both levels. In order to make performance comparisons of selected models correctly, there are many methods that are used for comparing the predictive ability (performance) of the soft computing methods.

One of the most useful methods for performance comparison is based on error estimations. Error estimations are derived from the difference between values predicted by a statistical model ($\hat{y}_i$) and actual values (y_i). Standardized root mean square error (SRMSE), root mean square error (RMSE), and mean squared error (MSE) are represented as follows (Nijkamp, et al. 1996, Fotheringham and Knudsen 1987):

$$SRMSE = \frac{1}{\bar{y}}\left[\sum (y_i - \hat{y}_i)^2 / n\right]^{1/2} \tag{2.6}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2.7}$$

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{N} \tag{2.8}$$

According to the equations above, there is a relationship between different formulations as seen below:

$$RMSE = \sqrt{MSE} \tag{2.9}$$

and $$SRMSE = \frac{RMSE}{\bar{y}} \tag{2.10}$$

Where $\hat{y}_i$ is the probability of selecting mode i, y_i is the actual choice of mode i and n represents the number of alternatives in the choice set. Nijkamp et al. (1996, 2004) compared the performance of logit and neural network models in terms of models' applicability in the European freight flows. The average relative variance (ARV) is used as a statistical indicator of performance. Average relative variance is defined as (Nijkamp 2004, Fischer and Gopal 1994):

$$R^2 = \frac{\sum (y - y^*)^2}{\sum (y - \bar{y})^2} \tag{2.11}$$

Where y is the observed probability, y^* presents probability, predicted by the adapted model. Canterarella and Luca (2003, 2005) tested performance of multinomial logit and neural network models by means of mean square error function (MSE). MSE represents error between observed and simulated mode choice probabilities. Celikoglu (2006) used the root mean square error (RMSE) and the coefficient determination (R^2) for performance comparisons of logit, neural network, and linear regression models. Hensher and Ton (2000) used a prediction success table which is a format for comparing the prediction capability of nested logit and ANN models. This evaluation measure presents predicted share less observed share for every mode of travel and the weighted percent correct and weighted success index. The performance test of discrete choice models (logit and probit) and soft computing methods can be analyzed by using success rates (% correctly predicted) of the models (Abdelwahab and Sayed 1999, Sayed and Razavi 2000, Vythoulkas and Koutsopoulos 2003). Succes rate of the models is obtained from a contingency table (crosstab or confusion matrix). Contingency table represents predicted choice outcomes for a test sample set versus the actual choice outcomes. Contingency table also provides information about overall error rate. Andrade et al. (2006) used the root mean square error (RMSE) and the mean absolute error (MAE) measure to compare model performance between multinominal logit and neurofuzzy models. Tortum et al. (2008) compared the performance of logit model, multiple regression model, neural networks, and neuro-fuzzy inference systems using root mean square error and correlation coefficient. Demir and Gercek (2006) studied to identify best performance measures to compare binary logit and soft computing methods (artificial neural networks, fuzzy logic, and neuro-fuzzy logic) for modeling mode choice in urban passenger transportation in Eskişehir. The performance measures used in the study are R^2, % correctly predicted, kappa statistics, and ROC curve.

Another potential technique can be Receiver Operating Characteristics Curves (ROC). The curve method has been used in different areas such as predicting multilateral credit risk (Tang and Chi 2005) and in biomedical and psychophysical applications (Türe, et al. 2005, Dirican 2001, Jaimes, et al. 2005, Phibanchon, et al. 2007). The area under the ROC curve (AUC) is an important index of a general measure of features of the underlying distribution of forecasts.

In the content of the study, performance analysis between logit and Bayesian belief networks is made. However, there is a lack of empirical studies associated with Bayesian belief networks in travel demand modeling in existing literature. Therefore, the methods for performance comparisons need to be determined. Error estimations and crosstab may be used as an indicator for performance comparisons.

2.5. Model Estimation Algorithms

The dependent variable in mode choice studies, P_i, can take an infinite number of values. In other words, dependent variable is a probabilistic. Therefore, ordinary least square method is not suitable for discrete choice models. In the calibration or estimation of discrete choice models, maximum likelihood estimation (MLE) is the most preferred statistical method to estimate model parameters ($\theta_1,\theta_2,\ldots,\theta_k$). It is defined that "*a maximum likelihood estimator is the value of the parameters for which the observed sample is most likely to have occurred*" (Ben Akiva and Lerman 1985, 20). The logic behind the estimation is a searching for the maximum value of a likelihood function or parameter values that maximize the likelihood function. The maximum likelihood procedure selects those estimates that maximize the probability of the observed sample (Ramanathan 1998). The maximum likelihood function is written as follows:

$$L(x/\theta) = \prod_{i=1}^{n} f(x_i, \theta) = f(x_1, \theta) \cdot f(x_2, \theta) ...... f(x_n, \theta) \quad (2.12)$$

Where x is a random variable, θ is the parameter (or coefficients). $f(x/\theta)$ is the probability density. The value for L is the highest will be chosen.

Maximum likelihood estimators are consistent, asymptotically efficient, and asymptotically normal, and asymptotically unbiased (Kennedy 1981, Ramanathan 1998). The method can deal with complex data due to its robustness. MLE is especially used for small sample properties, but some econometric assumptions such as normal distribution for disturbance term limit the use of MLE. Also, its computational difficult is an another limitation. However, many types of software include this estimation. MLE is an iterative procedure. In this estimation process, Newton-Raphson's method can be used.

CHAPTER 3

MODELING METHODOLOGY

Mode choice model is the third step of traditional four - step transportation modeling. In this stage, discrete choice models have been extensively used. Discrete choice models are derived under the assumption of utility – maximizing behavior. Theoretical contributions of the models comes from psychology (e.g., Marschak 1960) and econometry (e.g., Ben Akiva and Lerman 1985, Domencich and McFadden 1975, Manski 1973, Luce 1959). Different assumptions for the error terms give rise to different discrete choice models such as logit and probit models.

The goal of this study is to explore the effects of land use characteristics on mode choice behavior and make the performance comparison of mode choice models (Logit and BBNs) in Istanbul. Both aggregate and disaggregate models are estimated in the content of the study. Individual choice theory and existing literature provide domain knowledge for selecting explanatory variables. Some of these variables have been used in the study, including socioeconomic characteristics, travel characteristics, and population density. However, for this study, this guidance is not enough, since the number of empirical studies about this subject in the case of developing countries is very few. Several land use variables are entered into the models instead of using standard variable set used in mode choice studies. Many of these variables have never been used in mode choice modeling studies in the case of developing countries. Also, baseline category logit and Bayesian Belief Networks in mode choice studies have been rarely used in mode choice studies. The remainder of this chapter is organized as follows. Firstly, theoretical background of the models is discussed briefly in Section 3.1. The section introduces theories of individual choice behavior that are used in the formulation of traditional choice models. Discrete choice models are presented in Section 3.2. Section 3.3 introduces Bayesian Belief Networks (BBNs). After that, empirical mode choice models including research design - methodology and the model structure of the models are presented in Section 3.4.

3.1. Theoretical Background

The empirical analysis of mode choice in this study applies discrete choice model (MNL), baseline category logit model, and bayesian belief networks (BBNs). They are all models that are currently being used in probabilistic choice. Discrete choice modeling has been highly used in transportation modeling for the last fourty years. The probabilistic choice models such as discrete choice models are based on economic consumer choice theory (Ben Akiva and Lerman 1985, Domencich and McFadden 1975). In general, mode choice in transportation modeling is evaluated in consumer choice theory. The neoclassic economic theory suggests that a decision maker is able to compare two alternatives in the choice set. In consumer theory, utility plays an important role in the determining the behavior of individuals. Random utility theory is more suitable with consumer theory. Next section introduces random utility theory and individual choice behavior.

3.1.1. Individual Choice Behavior and Random Utility Theory

Choice is an important factor of the modeling of individual behavior. Choise itself is a complex process. A choice is conceptualized as an outcome of a sequential decision making process that include following steps (Ben Akiva and Lerman 1985, 31):

1. Definition of the choice problem,
2. Generation of alternatives,
3. Evaluation of attributes of the alternatives,
4. Choice,
5. Implementation.

On the other hand, choice theory includes following elements (Ben Akiva and Lerman 1985, 32):

1. Decision maker,
2. Alternatives,
3. Attributes of alternatives,
4. Decision rule.

A decision maker may be an individual or a household. An individual, called as consumer in micro economy, is defined as traveler (commuter) in mode choice analysis. Decision makers choose among a set of alternatives like consumer. Decision makers may face different choice situations (alternatives). The set of alternatives is called as choice set in theories of individual choice behavior. Choice set is defined as following characteristics (Train 2003, 15):

1. Alternatives must be mutually exclusive from the decision maker's perspective. Choosing one alternative necessarily implies not choosing any of the other alternatives. The decision maker chooses only one alternative from the choice set.
2. The choice set must be exhaustive, in that all possible alternatives are included. The decision maker necessarily chooses one of the alternatives.
3. The number of alternatives must be finite. The researcher can count the alternatives and eventually be finished counting.

The attractiveness of the alternatives in a choice set is evaluated by a set of attribute values that are measured as ordinal or cardinal (Ben Akiva and Lerman 1985). Attributes of alternatives might be generic (e.g., travel time and travel cost) or alternative specific (modal preference) attributes. The set of alternatives (choice set) may influence choice probabilities. From the selection of an alternative, individuals may have different tastes or different level of satisfaction. Assume that commuters choose between driving a car and using public transit, choice between car and public transit is determined by a comparison of the attributes of the alternatives and individuals. Since commuters may have different income levels (or budget constraints) and live different residential locations, the preferences of commuters may vary substantially. In choice analysis, an analyst must to decide on how to measure the factors that affect a decision maker's preference for car over public transit or vice versa. The choice between two different alternatives should be determined by a comparison of the attributes of the alternatives. Therefore, it must be found a way of measuring a decision maker's preferences (Hensher, et al. 2005). Preferences of commuters are evaluated by assigning a numerical score to each combination of the attributes. Numerical scores[6] are used to quantify the preferences of decision makers. The selection from the choice set is the alternative preferred by a decision maker. This selection requires a decision rule or behavioral rule. In existing literature, the most common rule (or numerical measure) is

[6] Numerical score is referred to as "level of satisfaction", in psychology while it is called "level of utility" in economics (Hensher, et al. 2005).

utility (Ben Akiva and Lerman 1985). Attractiveness of alternatives is formulized as a utility function. Therefore, the utility of each alternative represents a measure of preferences for that alternative. This means that decision maker assigns a utility value to each of the alternatives in the choice set. For the commuter mode choice example, it is expected that commuter will select the alternative with the highest utility (Ben Akiva and Lerman 1985). This preference is based on the combination of the attributes of alternatives that provides the highest utility to decision maker. Decision maker tries to maximize the level of satisfaction. Thus, this behavioral rule is called as *"utility maximizing behavior"* (Hensher, et al. 2005). In random utility theory, a decision maker is always assumed to select utility – maximizing alternative.

In theory, measurement of choice is based on different assumptions: deterministic choice and stochastic choice. Deterministic choice is a linear choice function, *V(i)*, of the demand and supply variables. The deterministic choice function is written as follows (Kanafani 1983):

$$V(i) = A_i X_i, \forall i \in I \tag{3.1}$$

Where X_i presents a vector of demand and supply variables influencing choice and A_i is a vector of parameters representing the effect of each variable. The decision rule for a deterministic choice model is as follow (Kanafani 1983):

$$V(j) = \max[V(i)] \tag{3.2}$$

According to this utility function, decision maker chooses the alternative with highest utility level. The theory assumes that individual facing same alternatives will choose the same choice over time. It means that decision makers having similar socioeconomic characteristics make the same choices when faced with the same alternatives. However, deterministic choice is accepted as unrealistic for real life situations due to three primary reasons as follows:

> Three primary reasons suggest that a stochastic model of choice may be preferable. One is that the behavior of individuals may not always follow the rational rules of choice exactly and that the idiosyncrasies of traveler behavior cannot be anticipated in a deterministic model. The second is that it is usually not possible to include in the choice function V(.) all the variables that

> can possibly influence choice. If such a function were possible, it would no doubt be so complicated as to render it impractical. The third reason is that the typical potential traveler is not likely to have perfect information about the transportation system and the alternatives it offers (Kanafani 1983, 122).

It is more realistic that a choice function is accepted as a random function that produces probabilities for given variables in a choice set. It means that the attributes of the alternatives are perceived differently by decision makers. Stochastic choice models have been widely used in travel demand modeling.

Choice probabilities are affected by the attributes of the alternatives. In other words, the attractiveness of the alternatives is represented in terms of a vector of attribute values (Ben Akiva and Lerman 1985). However, some attributes of alternatives cannot be known or measured such as comfort, security and convenience. These unobserved attributes may be an important part of choice analysis.

> The behavioral basis of individual choice theory presumes that all decisions are probabilistic, and that they are derived from a comparative evaluation of utilities. The probability or likelihood a specific alternative will be chosen by an individual is based on the utility associated with that alternative. The utility of the alternative is composed of its attributes. In making a choice among the available alternatives, an individual is assumed to assesses the attributes of each alternative. Based on this assessment, a utility value is assigned to each of the alternatives (Taaffe, et al. 1996, 342).

In this stage, an important contribution to discrete choice models comes from random utility theory that is more suitable with consumer theory (Domencich and McFadden 1975, Ben Akiva and Lerman 1985). The theory assumes that (Ortuzar and Willumsen 2006, 223):

> 1. Individuals belong to a given homogeneous population,… act rationally and possess perfect information, i.e. they always select that option which maximizes their net personal utility….
> 2. There is a certain set $A=\{A_1,\ldots,A_j,\ldots A_N\}$ of available alternatives and a set X of vectors of measured attributes of the individuals and their alternatives.
> 3. Each option $A_j \in A$ has associated a net utility U_{jq} for individual q. The modeller, who is an observer of the system, does not possess complete information about all the elements considered by the individual making a choice; therefore, the modeler assumes that U_{jq} can be represented by two components:
>   - a measurable, systematic or representative part V_{jq} which is a function of the measured attributes x; and
>   - a random part ε_{jq} which reflects the idiosyncrasies and particular tastes of each individual, together with any measurement or observational errors made by the modeller.

In random utility theory, individuals are accepted as a rational decision maker. Rational decision maker maximize utility relative to his/her choices. The choice set may be different for individuals since all the alternatives in the choice set may not be available to all individuals. For example, car alternative cannot be suitable for a decision maker without driving license. In other words, the alternative that is selected provides the highest utility in comparison to other alternatives. However, a choice analyst may not measure directly attributes. Therefore, the utilities are treated as random variables due to four distinct sources of randomness that were identified by Manski (1973) (Ben Akiva and Lerman 1985, 56):

1. Unobserved attributes,
2. Unobserved taste variations,
3. Measurement errors and imperfect information,
4. Instrumental (or proxy) variables.

The utility assigned to each alternative depends on the characteristics (or attributes) of alternatives and also individuals. It must be recognized that the utilities derived from the attributes of alternatives are not known to the analyst with certainty. Because of this, in this theoretical framework, random variables are taken into account in the utility function by an analyst. An important contributions to random utility theory belonged to Marschak (1960) who provided a derivation from utility maximizing and McFadden (1974) who developed the utility function as a function of a vector of attributes, socioeconomic characteristics, and unobserved vector containing all the attributes of the alternatives and characteristics of the individual which analyst are unable to measure. Decision makers are assumed to have perfect discriminating capability, but sources of randomness limit information about an individual's utility. Therefore, the choice analyst has less information than decision maker. In this framework, the uncertainty is taken into account with a random variable. After defining a set of observed and unobserved influences of the attributes on individual choice behavior, the utility function of an alternative that an individual associates with alternative *i* in the choice set is expressed as (Ben Akiva and Bierlaire 1999, McFadden 1974, Train 2003).

$$U_i = V_i + \varepsilon_i \tag{3.3}$$

Where V_i represents the deterministic (or systematic) part of the utility, and ε is the random term which is usually used to refer to the unobserved influences as error. Choice probabilities are based on the assumptions about the random term, sometimes known as error term. It captures the uncertainty. U_i is the overall utility of an alternative. Therefore, the random utility of an alternative is represented by a sum of systematic (or representative) component and random component. As mentioned above, this utility is known to the decision maker, but not known by the analyst. This utility function represents the measure of the level of satisfaction that individuals derive from their choices. Attributes of alternatives and socioeconomic status of individuals affect the magnitude of utility functions (Papacostas and Prevedovros 1993). Utility of a travel mode for a given trip should be measured by the total bundle of attributes (Oppenheim 1995). Utility function is expressed as follows:

$$U_i = U_i(X_i, S_t, \varepsilon_i) \tag{3.4}$$

Where U_i represents the utility of the ith alternative, X_t is a vector of observed attributes of ith alternative. S_t is a vector of observed socioeconomic characteristics of individuals while ε_i is random component of utility. The assumption for random utility theory is that individuals are assumed to choose the utility - maximizing alternatives. From the perspective of decision makers, a decision maker compares all possible alternatives in the choice set $(U_1, U_2, \ldots, U_i, \ldots, U_j)$ and one alternative with highest utility will be chosen such as U_i. Therefore, the probability of an alternative i which is chosen by decision maker from a choice set is greater than or equal to the choice probability of alternative j. The probability that decision maker n chooses alternative i is written as follows (Ben Akiva and Lerman 1985, Train 2003):

$$P_{ni} = \Pr ob(U_{ni} > U_{nj} \forall j \neq i) \tag{3.5}$$

$$= \Pr ob(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj} \forall j \neq i) \tag{3.6}$$

$$= \Pr ob(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj} \forall j \neq i) \tag{3.7}$$

The last equation means that the probability of individual *n* choosing alternative *i* is equal to the probability that the difference in the random utility of alternative *j* and alternative *i* is less than or equal to difference in the representative utilities of the alternatives. Different discrete choice models are derived from different assumptions about the distribution of the random component as explained below (Section 3.2).

3.2. Discrete Choice Models

Discrete choice models have been a popular method in travel demand modeling. The models are commonly used to analyze decision makers' choices among two or more discrete alternatives. In other words, discrete choice models are used to estimate the probability that a decision maker chooses a particular alternative in a choice set relate to the attributes of alternatives and decision makers.

Individual choices among a finite set of alternatives may indicate a huge amount of variability. *"This variability, often referred to as heterogeneity, is in the main not observed by the analyst. The challenge is to find ways of observing and hence measuring this variability, maximizing the amount of measured variability (or observed heterogeneity) and minimizing the amount of unmeasured variability (or unobserved heterogeneity)"* (Hensher, et al. 2005, 62). Therefore, a theoretical framework is obtained from the theories of individual choice behavior including probabilistic choice theory and random utility theory. Different logit models are derived from different assumptions about random component of the utility function. Multinomial logit, binary logit, conditional logit, binary probit, multinomial probit, and mixed logit are the types of discrete choice models. Multinomial and binary logit models have been used widely due to estimation easiness up till now.

3.2.1. Derivation of a Choice Model: Logit and Probit Models

After definition of the utility function, how the functions of representative utility (V_i) and random utility (ε_i) are to be represented is important. From the perspective of an analyst, he or she does not observe decision maker's utility. Random term (ε_i) capture the factors affecting utility, but are not included in V_i. Since the representative utility

(V_i) includes observed (or measured) factors, a functional form can be derived. In general, representative component of utility is expressed as a linear function in which each attribute is a linearly weighted by a coefficient to account for relative attribute's marginal utility input that is (Hensher, et al. 2005):

$$V_i = \beta_{0i} + \beta_{1i} f(x_{1i}) + \beta_{2i} f(x_{2i}) + \beta_{3i} f(x_{3i}) + ... + \beta_{Ki} f(x_{Ki}) \qquad (3.8)$$

Where β_{1i} is the weight (or parameter) associated with attribute X_1 and alternative i, β_{0i} is a parameter not associated with any of the observed and measured attributes, called the alternative - specific constant, which represents an average the role of all the unobserved sources of utility (Hensher, et al. 2005). In addition to a linear function, a logarithmic form or quadratic form can be used. An analyst does not know anything about random component of the utility. It means that any numerical value cannot be assigned to random component. "*The best place to start is to recognize that each individual will have some utility associated with an alternative that is captured by the unobserved component. Across the sample of individuals, each person will have such data. That is, there will exist a distribution of such unobserved sources of utility across the sampled population*" (Hensher, et al. 2005, 76). In this stage, in order to derive an operational random utility model, an analyst needs to make some assumptions about the joint probability distribution of the full set of disturbances. Randomness in utility function is highly associated with a way of capturing information in random component. Since choice analysts do not have any idea about numerical value to assigned to it, some specific distributions of the random component are applied under assumptions. It might be thought as some structure applications on ε_{nj}. "*Once a particular distribution of the random component has been selected, the analyst is well on their way to having all the necessary data to derive a choice model*" (Hensher, et al. 2005, 83). Also, different assumptions about the distribution of the random component (unobserved portion) of the utility function lead to derive different choice models such as logit and probit models.

Logit models are the most widely used discrete choice models. If the dependent variable is dichtomous or represented by a dummy variable (e.g., 1 for taking public transport to work and 0 for drive to work), classic estimation methods such as least square methods must not be used. In other words, regression models breaks down.

Because the value of dependent variable represents a probability measure for which the realized value is 0 or 1. It is expected that the predicted value of the dependent variable is interpreted as the probability that an individual makes a travel decision on mode choice. These models whose dependent variable takes a binary form are called as linear probability model (or binary choice models) (Ramanathan 1998). In this situation, probabilistic distribution is needed to lie inside 0-1 interval.

Logit models are convenient model for studying the determination of categorical variables. Logistic model is used to find the probability of an event occurring. Its functional form is expressed as (Ramanathan 1998):

$$Ln\left[\frac{P}{1-P}\right] = \alpha + \beta x + \varepsilon \tag{3.9}$$

Where P represents the probability of an event (between 0 and 1), α and β are parameters (or coefficients). X is an independent variable. ε is an unobserved random variable (the error term). If applying first exponentiating both sides, probability of an event (P) which represents the predicted probability that an event occurs is rewritten as (Ramanathan 1998):

$$P = \frac{1}{1+e^{(-\alpha+\beta x+\varepsilon)}} \tag{3.10}$$

When there are many independent variables, the logistic model can be written as follow (Gujarati 1995):

$$P = \frac{1}{1+e^{(-Z)}} \tag{3.11}$$

This equation is called as the cumulative logistic distribution function and provide information about the choice of a travel mode. Z represents linear combination of parameters. In other words, Z represents the relative attractiveness of a travel mode. If P gives the probability of an event, the probability of an event not occurs is represented by (*1-P*). It is written as (Gujarati 1995):

$$1 - P = \frac{1}{1 + e^{Z}} \tag{3.12}$$

or this equation equals to that (Gujarati 1995):

$$\frac{P}{1 - P} = \frac{1 + e^{Z}}{1 + e^{-Z}} = e^{Z} \tag{3.13}$$

When taking the natural log of the equation, the following formula is obtained as seen in the following equation. This model can be estimated by ordinary least squares.

$$L = \ln\left(\frac{P}{1 - P}\right) = Z \tag{3.14}$$

$$= \beta_1 + \beta_2 x + ... + \beta_n X_n + \varepsilon \tag{3.15}$$

Where L is called as the logit (or logit model). In this model, the coefficient β_2 measures the change in L for a unit change in X. In other words, β represents the relative importance of each of the explanatory variables (X). Logit model assumes that the log of odds ratio is linearly related to X (Gujarati 1995). On the other hand, in order to predict mode choice of an individual, the utility function is transformed into a probability using the logit model.

In discrete choice analysis, multinomial logit and binary logit models have been used widely in travel demand modeling. In general, if there are only two alternatives, binary logit model is used. If there are more than three alternatives, multinomial logit models are used. Multinomial logit model is derived from the assumption that random residuals (or error term), ε_{ni}, is identically and independently distributed extreme value. This distribution is known as Gumbel and type I extreme value. The density for random component of the utility is written as (Train 2003):

$$f(\varepsilon_{nj}) = e^{-\varepsilon_{nj}} e^{-e^{-\varepsilon_{nj}}} \tag{3.16}$$

Under this assumption that random term (ε_i) is logistically distributed, the choice probabilities for alternative *i* is given by (Domencich and McFadden 1975, Ben Akiva and Lerman 1985, Train 2003):

$$P_n(i) = \frac{e^{V_{in}}}{\sum_{j \in C_n} e^{V_{jn}}} \qquad (3.17)$$

This equation is known as multinomial logit model[7]. It represents logit probabilities. The model is derived from the assumption that $\varepsilon_n = \varepsilon_{jn} - \varepsilon_{in}$. The choice probabilities for all alternatives in the choice set must sum to one. It means that decision maker can select only one alternative. The logit probability is sigmoid and S shaped as seen in the Figure 3.2. On the other hand, probit model is derived from the assumption that error terms (or unobserved components of the utility) are distributed jointly normal (Train 2003, Ben Akiva and Lerman 1985). The logistic and normal density functions are seen in the Figure 3.1 and Figure 3.2. Logit analysis (or logit regression) is different from classic regression models, but there may be some similarities.

> Unlike regression, the logit model permits of a specific economic interpretation in terms of utility maximization in situations of discrete choice. Among economists this confers a higher status on the model than that of a convenient empirical device. And there is a subtle distinction in that the ordinary regression model requires a disturbance term which is stuck on to the systematic part as a necessary nuisance, while in the logit model the random character of the outcome is an integral part of the initial specification. Together with the probit model, the logit model belongs to the class of probability models that determine discrete probabilities over a limited number of possible outcomes (Cramer 2003, 1).

On the other hand, both models present causal relationships between dependent variable and independent variables, and also permits of all sorts of extensions and of quite sophisticated variants.

[7] In the multinomial logit model, explanatory variables contain only characteristics of individuals while the conditional logit model is used when choice – specific data is available. In other words, alternative – specific variables are entered into the conditional logit model.

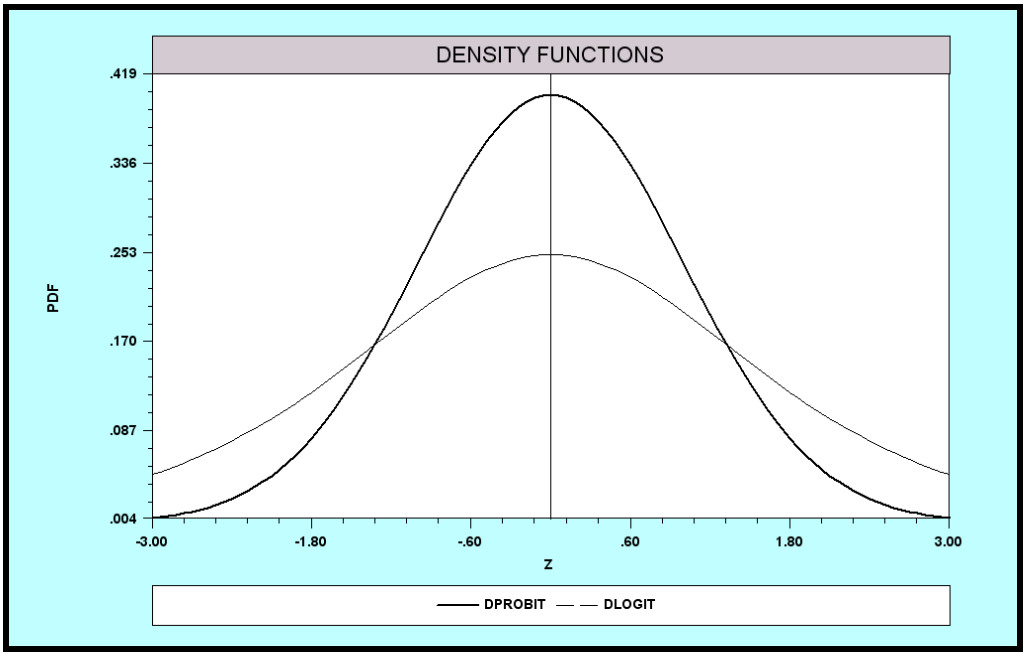


Figure 3.1. Density functions for probit and logit models

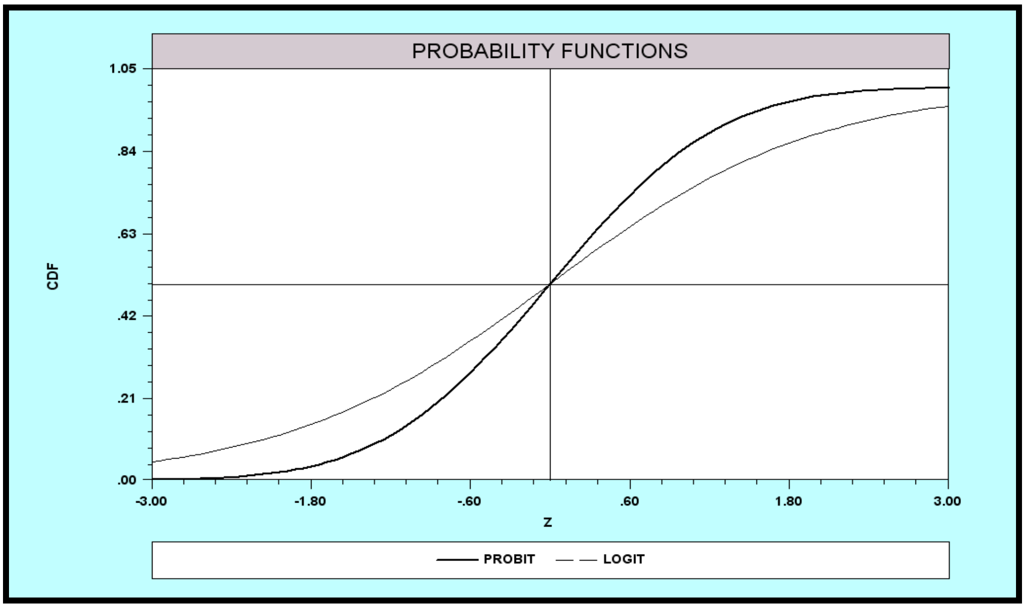


Figure 3.2. Cumulative distribution function for probit and logit models

One of the most important properties that restrict the use of multinomial logit model is the independence from irrelevant alternatives (IIA). As mentioned above, an analyst does not know and observe all the attributes of the alternatives in a choice set. Because of this reason, an analyst treats to utility as random, but a decision maker knows systematic and unobservable components of the utility derived from choosing an alternative. It can be suspected that alternatives in the choice set may share common unobserved attributes such as comfort and convenience. This situation increases correlation pattern. Shared unobserved attributes among alternatives cause to correlation in a choice set. Logit models cannot take into account these patterns. IIA property assumes that the ratio of the choice probabilities of any two alternatives is unaffected by the systematic utilities of other alternatives (Ben Akiva and Bierlaire, 1999). After introduction of a new mode to the choice set, the ratio of market share must not be affected by the new mode. This situation appears when unobserved attributes of the alternatives in the choice set are identical. The ratio of logit probabilities for any two choices (i and j) is as follows (Train 2003):

$$\frac{P_n i}{P_n k} = \frac{e^{V_{ni}} / \sum_j e^{V_{nj}}}{e^{V_{nk}} / \sum_j e^{V_{nj}}} = \frac{e^{V_{ni}}}{e^{V_{nk}}} = e^{V_{ni}-V_{nk}} \tag{3.18}$$

IIA requires that, all else being equal, an individual's choice between two alternatives is unaffected by other choices[8]. In existing literature, there are several tests[9] that are used for checking the assumption of IIA such as Hausman and McFadden test proposed by Hausman and McFadden (1984) and Small and Hsiao test (Cheng and Long 2007). Generalized Extreme Value (GEV) models are developed to capture the correlations among alternatives, when all correlations are not zero. GEV models are an extension of multinominal logit models. One of the most known GEV model is nested logit model. The nested logit model firstly was proposed by Ben Akiva (1973) to capture the correlation pattern in the choice set. The Nested Logit (NL) model assumes that there may be a probability that alternatives may share information about unobserved attributes. In other words, information for random component (ε_i) is possibly expected to be correlated or similar for some alternatives. For example, some unobserved attributes such as comfort and convenience may be the same for bus, train, and subway alternatives as seen in the Figure 3.3. This cannot be observed by an analyst. In the NL model, the alternatives sharing unobserved attributes are partitioned

---

[8] The most known example for IIA property is blue-bus-red bus paradox that is discussed in Ben Akiva and Lerman (1985) and Train (2003).

[9] The tests for IIA property are discussed in Greene (2003) and Train (2003).

into subsets. The subsets are called as nests for removing the IIA property. In addition to GEV models, probit model does not suffer from IIA property. The main assumptions for discrete choice models are summarized in Table 3.1.

Table 3.1. Discrete choice models
(Source: adapted from SAS User Guide: The MDC Procedure 2008)

Model Type	**Assumption for Random (Error) Term**
Multinomial Logit	Type I Extreme Value Independent and identical
Multiomial Probit	Multivariate Normal Distribution Correlated and non-identical
Nested Logit	Generalized Extreme Value Correlated and identical
Mixed Logit	Type I Extreme Value Independent and identical
HEV Models	Heteroscedastic Extreme Value Independent and non-identical

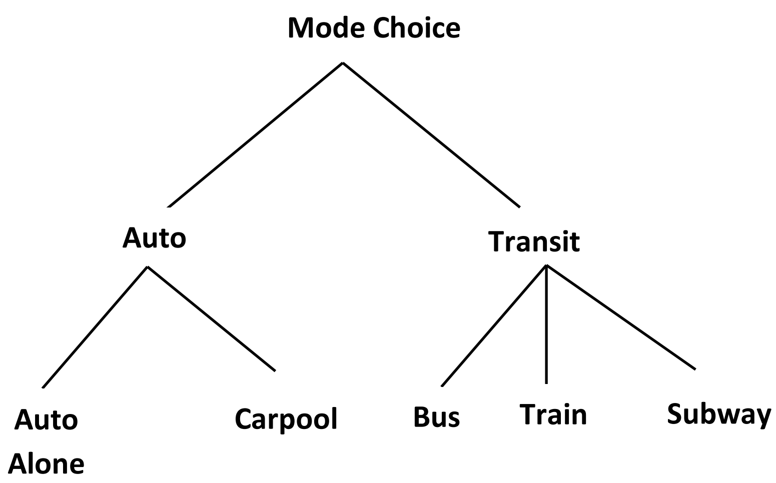


Figure 3.3. Two - level nested logit model

Random utility theory presents theoretical framework for discrete choice models. Discrete choice models are used when individuals have to select an option from a finite set of mutually exclusive and exhaustive alternatives. Discrete choice model assumes that "*the probability of individuals choosing a given option is a function of their socioeconomic characteristics and the relative attractiveness of the option*"

(Ortuzar and Willumsen 2006, 220). The probability of an alternative that is chosen from a choice set is defined as the probability that it has the highest utility among a set of possible alternatives (McFadden 1974).

3.3. Bayesian Belief Networks (BBNs)

Bayesian belief networks (also called as causal probabilistic networks, causal nets, and probabilistic graphical networks) provide a statistical tool for dealing with uncertain and complex domains. The development of BBNs was started during the 1990s in parallel with the development of softwares such as Netica and Hugin. BBNs that were developed in the fields of artificial intelligence and machine learning is a graphical representation of probabilistic relationships among a large number of variables in a problem domain (Pearl 1988, Jensen 2001). BBNs is a probabilistic model. The networks are based on probability theory developed by Thomas Bayes. Bayes networks allow researchers to do probabilistic inference. They have been applied to many problems, ranging from environmental modeling and management to pattern recognition and classification, medical diagnoses (Bromley, et al. 2005, Lee, et al. 2003, Kahn, et al. 1997, Aktaş, et al. 2007), operational risk management in banks (Cornalba and Giudici 2004) to resource planning and management. Therefore, bayesian networks have become a popular method in recent years for handling uncertainty in complex domains. However, the application of BBNs into transportation modeling is rather limited.

3.3.1. General Terminology in Bayesian Belief Networks

BBNs provide a graphical model (DAG) representing dependencies and independencies among the variables in terms of conditional probability distributions (CPTs) (Alpaydin 2004). BBNs consist of two components: a directed acyclic graph (DAG, qualitative part) and conditional probabilities (CPT, quantitative part) for each variable in a problem domain (Pearl 1988, Torres and Huber 2003). It is considered that conditional probabilities are model parameters. The degree of the relationship is expressed quantitatively by probabilistic terms. The networks are used to assess cause and effect relationships among the variables. BBNs consist of following properties (Jensen 2001, 19):

1. A set of variables and a set of directed edges between variables.
2. Each variable has a finite set of mutually exclusive states.
3. The variables together with the directed edges form a directed acyclic graph (DAG).
4. To each variable A with parents $B_1,\ldots\ldots,B_n$, there is attached the potential table $P(A \setminus B_1,\ldots\ldots,B_n)$.

This network has its own terminology. According to this terminology, a directed acyclic graph (DAG) as a structural part of the network is denoted by N (G, P). Where the graph *G* represents vertices *V* (a set of nodes) and edges (or arrows) between nodes. *P* represents a set of conditional probability distributions. In other words, bayesian belief networks are directed acyclic graphs in which each variable is represented by a ***node*** (or variable), and causal relationships are denoted by an ***edge***. The values of the nodes in the network are represented by ***states***. There are arrows (or edge) between nodes. An arrow represents a causal relationship between two nodes. The direction of an arrow indicates the direction of causality. The meaning of an edge drawn from node B to node C is that node B has a direct influence on node C as seen in Figure 3.4. It means that child nodes are conditionally dependent upon their parent nodes. Conditional probability tables (node C) show how one node influences another. When two nodes are joined by an edge, the causal node is called the parent of the other node. Therefore, changes in the states of any variable may cause changes in the states of other variables. This change is highly related with the strengths of dependencies between variables. The dependencies (or the strength of the influences) among variables are represented by conditional probability tables (CPT). Each node has a conditional probability table (CPT). Conditional probabilities may represent likelihoods based on prior information or past experience.

In sum, a Bayesian network $N(G, P)$ represents a joint probability distribution. Joint probability distribution is the probabilities of each of the combinations of states of the nodes in a bayes network. For a probability distribution, $P(X)$, over a set of variables $X=\{X_1,\ldots..,X_n\}$, the joint probability distribution $P(X) = P(X_1,\ldots.X_n)$ is the product of all potentials specified in bayes network. The **chain rule** of probability for bayes network is as follow (Jensen 2001).

$$P(X) = P(X_1, \ldots\ldots, X_n) = \prod_{i=1}^{n} P(X_i \,/\, pa(X_i)), \tag{3.19}$$

In this equation, $pa(X_i)$ is the parent set of X_i. The chain rule for Bayesian networks are extracted from conditional independence property. BBNs represent joint probability distributions by means of DAGs which represents the dependencies and independencies among variables in a domain as well as the conditional probability distributions of each variable, given its parents in the graph (Neapolitan 1990). Each conditional probability distribution, $P(X_i \setminus pa(X_i))$ includes a set of rules (Kjaerulff and Madsen 2008). For Bayesian networks, the chain rule property is explained by the Markov assumption (or Markov Condition). Chain rule yields a joint probability table for modeling purpose.

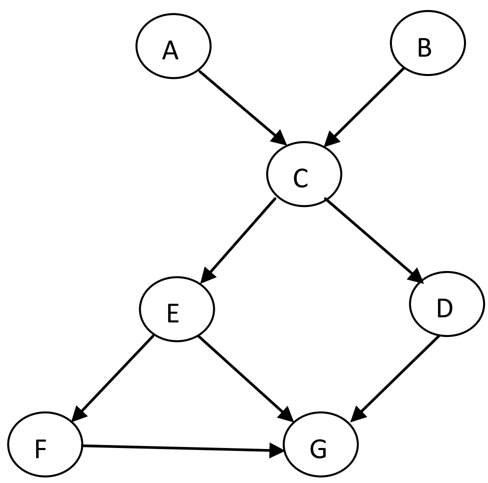
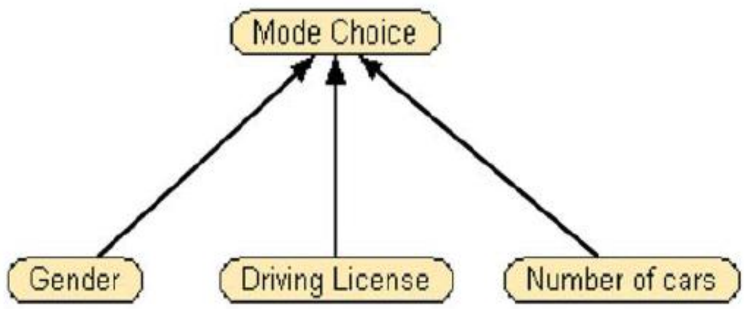


Figure 3.4. A directed acylic graph (DAG).

According to the Figure 3.4, prior probabilities $P(A)$ and $P(B)$ and also conditional probabilities $P(C \setminus A, B)$, $P(E \setminus C)$, $P(D \setminus C)$, $P(F \setminus E)$, and $P(G \setminus D, E, F)$ must be determined. The relationship between nodes and their parents are represented by conditional probability tables. CPT represents prior distributions (prior probability). These distributions (or prior information) are called as *beliefs*. Prior probability, ($P(A)$ and $P(B)$ for Figure 3.4) can be used when no other information is available (Lee and Abbott 2003). However, new information can be obtained for the states of the nodes. An advantage of Bayesian Belief Networks is to compute posterior probability distributions when new information is available about a current situation. New information about a current situation of the nodes is called as *evidence*. BBNs allow an analyst to enter a probability for evidence information of a node. When evidence is entered into the network, it will change the states of other variables. Examples of Bayesian belief network with DAG and CPT are shown in Figure 3.5 and Figure 3.6 for mode choice and marketing research.

Gender	Driving License	Number of cars	*P* (mode choice = bike)	*P* (mode hoice = car)
Male	Yes	1	0.2	0.8
Male	Yes	>1	0.6	0.4
Male	No	1	0.7	0.3
Male	No	>1	0.4	0.6
Female	Yes	1	0.4	0.6
Female	Yes	>1	0.8	0.2
Female	No	1	0.1	0.9
Female	No	>1	0.3	0.7

CPT Mode Choice



P(Gender = male)	*P* (Gender = female)
0.75	0.25

CPT Gender

P (Driving License=Yes)	*P* (Driving License=No)
0.6	0.4

CPT Driving License

P (Number of cars=1)	*P* (Number of cars>1)
0.2	0.8

CPT Number of Cars

Figure 3.5. A bayesian network with CPT for a mode choice problem
(Source: Janssens, et al. 2006)

Figure 3.5 represents that the nodes, "Number of Cars, "Gender", and "Driving License" are parents of the node, "Mode Choice". It means that mode choice is the child of the nodes, "Number of Cars", "Gender", and "Driving License" (Janssens, et al. 2006). The network indicates that gender, Driving License, and number of cars directy influence mode choice. In order to get the prior probabilities for the node, "Mode Choice", bayes' rule is used and written as follows:

$$\begin{aligned} &P(Choice, Gender, NumberofCars, DrivingLicense) \\ &= P(Choice / Gender, NumberofCars, DrivingLicense) \\ &\times P(Gender, NumberofCars, DrivingLicense) \end{aligned} \tag{3.20}$$

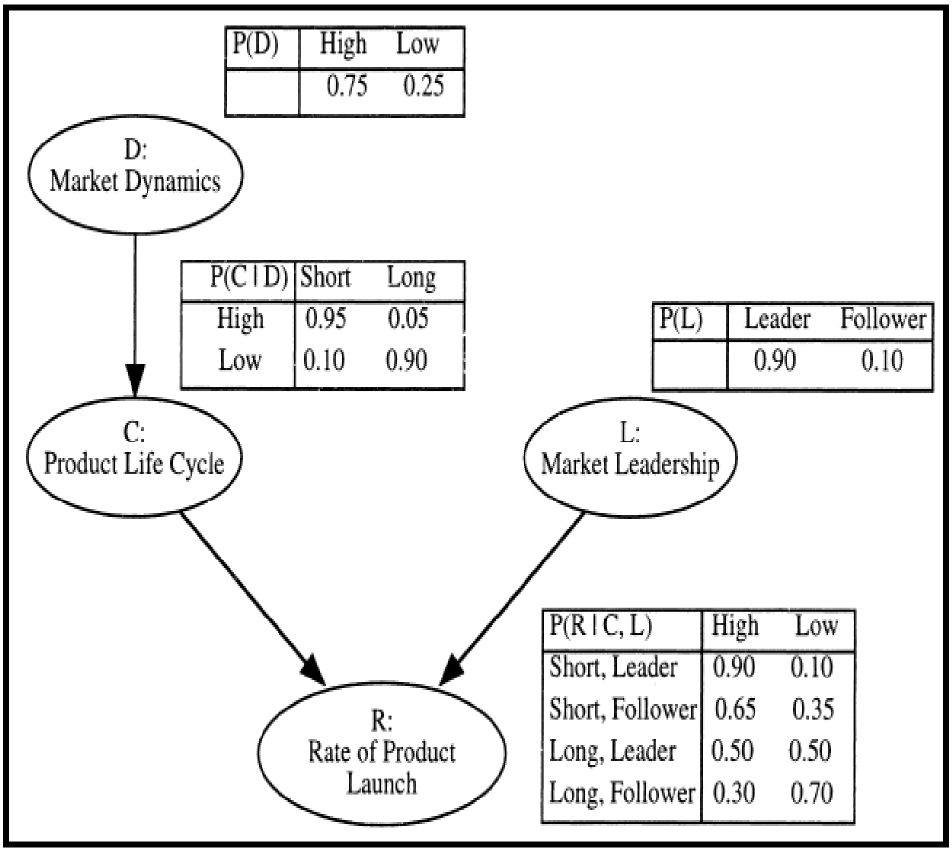


Figure 3.6. A bayesian network with CPT for a marketing research
(Source: Nadkarni and Shenoy 2001)

According to Figure 3.6, all of the links in the network are causal. Product life cycle (C) and market leadership (L) directly affect the rate of product. In this sample, the nodes, C and L are the parents of the node, "Rate of Product" (R). The node, "Market Dynamics" is the parent node of "Product Life Cycle".

If many nodes are dependent in the network, computations may become difficult. In this situation, this can be done by means of probabilistic inference algorithms that are included in some softwares such as Hugin and Netica softwares.

3.3.2 Conditional Probability and Bayes Theorem

Theoretical background of Bayesian belief networks comes from Bayes' Theorem. Bayes theorem relates conditional and marginal probabilities of random events. The theorem is used for calculating posterior probabilities given data (or observations). In this stage, firstly, conditional and marginal probability as a basic concept of bayesian analysis is explained. After that, bayes' theorem is explained. Conditional Probability means that if the events *A* and *B* are dependent, it is gained information about *P(A)* if the information that event *B* has occurred is known. The statement for conditional probability is that *"Given the event B, the probability of the event A is x"*. This statement is written as $P\ (A\backslash\ B) = x$ (Jensen 2001, Lynch 2007). The theorem states simply:

$$P(A \backslash B) = \frac{P(B \backslash A)P(A)}{P(B)} \tag{3.21}$$

This equation can be rewritten as:

$$P(A, B) = P(A / B)P(B), \tag{3.22}$$

$$P(B, A) = P(B / A)P(A). \tag{3.23}$$

P(A, B) = P(B, A) gives the equation 3.21. This equation yields the Bayes' rule. Bayesian Belief Networks are based on the work of the mathematician and theologian Rev. Thomas Bayes, who worked with conditional probability theory in the late 1700s to discover a basic law of probability, which was then called Bayes' rule. Marginal probability of event *A*, *P(A),* is computed as the sum of the conditional probability of *A* under all possible events B_i (Lynch 2007):

$$P(A) = \sum_{B_i \in S_B} p(A \backslash B_i)\, p(B_i) \tag{3.24}$$

Where *P(A)* is the prior probability or marginal probability of event *A*. It is 'prior' in the sense that it does not take into account any information about *B*. *P(A | B)* which means *"the probability of event A, given event B"*, is the conditional probability of *A*, given *B*. It is also called the posterior probability because it is derived from or depends upon the specified value of *B*. *P(B | A)* is the conditional probability of *B*, given *A*. *P(B)* is the prior or marginal probability (sometimes called as unconditional probability) of event *B*. The probability of an event *A* occurring is expressed with prior or unconditional probability. If there is evidence about relevant that event, this probability becomes posterior (or conditional) probability. According to the theorem, a conditional probability for event *A* given event *B* is equal to the conditional probability of event *B* given event *A*, multiplied by the marginal probability for event A and divided by the marginal probability for event *B* (Lynch 2007). With this formulation, the theory provides an opportunity for calculating the probability of interest. In Bayesian terminology, marginal probabilities such as *P(B)* or *P(A)* represents prior information of events in the domain. This information may come from previous researches and expert knowledge. This information is used in estimating posterior probabilities. This probability is called as posterior probability. It can be repeated in the next step as prior probability to estimate a new posterior probability. In other words, once new information is available, the conditional probability of *P(A\B)* that means the probability of *A*, given *B* will changed in the network.

3.3.3. Inference in Bayesian Networks

Inference in a statistical analysis, sometimes called as probabilistic inference, is important for making predictions and decision making. Bayesian networks as one of recent advances in artificial intelligence provide a powerful tool for making inferences in decision making process. *"Probabilistic inference refers to the process of computing the posterior marginal probability distributions of a set of variables of interest after obtaining some observations of other variables in the model"* (Nadkarni and Shenoy 2001, 484). Inference in a Bayesian network is based on the evidence propagation. Bayes networks in general are used to find posterior distribution of variables given evidence. It is called as probabilistic inference in bayes networks. As mentioned before, bayesian networks is a specification of a joint probability distribution of several variables in terms of conditional distributions for each variable in the network (Pearl

1988). Efficiency of the inference in bayesian networks is highly related with the structure of the DAG.

Bayesian analysis provides an efficient tool for reasoning under uncertainty. Reasoning under uncertainty needs the task of computing the updated beliefs in (unobserved) events given observations on other events such as evidence (Kjaerulff and Madsen 2008). Bayes theorem provides a practical application for statistical inference. Bayes networks performs bayes' theorem to problems. Inference in bayesian networks means computing posterior beliefs given evidence. However, in real life applications, inference is accepted as an NP - hard task. There are some inference methods used in bayesian inference: markov chain monte carlo, query - based inference, arc reversal, and message points in junction trees. All of these methods work with inference algorithms, but these algorithms are worst case nonpolynominal time and the problem of approximate inference is NP - hard. (Neapolitan and Jiang 2007). If the structure of bayes network is simply, inference can be simple. Also, the result of inference algorithm is based on the structure of a bayes network. Therefore, some softwares such as Netica and Hugin have been developed for this purpose. The softwares provide users to automate the process of inference.

3.3.4. Learning Bayesian Networks From Data

Bayesian networks are defined as graphical representation (or graphical structure) for the probabilistic relationships among random variables. It allows doing probabilistic inference with the variables in a problem domain. In this stage, bayes' theorem is used for probabilistic inference and to compute the conditional probability distribution among the variables. However, conditional probabilities in a large amount of variables cannot be computed easily by applying of a standard bayes' theorem. For this reason, bayesian networks were developed to do inference correctly, and to compute conditional probabilities in an acceptable amount of time.

After deciding which variables and their states that are used in the model, a researcher meets two tasks in data mining process using bayesian networks. The first stage is to construct of a bayesian network. In other words, bayesian network structure (DAG) must be defined. The resulting DAG represents a set of dependence and independences causal relationships among the variables. The second stage is to assessment the prior and conditional probabilities. It can be defined as parameters. These tasks in the existing literature are called as ***learning bayesian networks from data***. It means to define the optimal structure and local probability distributions given data (Heckerman 1996). Learning means that the task of finding a generic model of empirical data (Pearl 1988). *"Up to until the early 1990s the DAG in a Bayesian network was ordinarily hand-constructed by a domain expert then the conditional probabilities were assessed by the expert, learned from data, or obtained using a combination of both techniques"* (Neapolitan and Jiang 2007, 111). Therefore, since the construction bayesian networks from domain experts can be considered as a labor-intensive task, many algorithms for learning bayesian networks from data have been developed. In general, learning in bayesian networks from data traditionally is comprised of two subtasks: **Structural Learning and Parameter Learning**. It must be noted that in bayesian analysis, DAG means the structure of a network and the conditional probability distributions are defined as model parameters. In this section, learning methods and algorithms as a data mining tool are discussed.

3.3.4.1. Structure Learning

Structure Learning includes the task of inducing the structure (DAG) of a bayesian network from data. "*Structure learning determines the dependence and independence of variables and suggests a direction of causation, in other words, the placement of the links in the network*" (Janssess, et al. 2006, 24). As mentioned above, bayesian network can be typically constructed from expert knowledge. This method builds the structure of a bayesian network manually. Learning bayesian network with these methods can be difficult for complex domain. However, correct structure must be defined to estimate model parameters. In the structure learning phase, there are mainly two different approaches (Steck and Tresp 1999): **Constrained based and Search-and-Score Algorithms**. These two methods differ from each other. The task for searching for a good network structure needs efficient learning algorithms which can find close to optimum solutions in a reasonable amount of time because the number of possible networks is super - exponential in the number of nodes. Therefore, it is not easy to test all search space entirely without algorithms. The size of possible DAGs is a function of the number of nodes G(n) and it grows super-exponentially with the number of nodes in the graph (Kjaerulff and Madsen 2008, Scuderi and Clifton 2005). Firstly, Robinson (1977) suggested a formulation for calculating the number of DAGs, *f(n)*, including *n* variables. Table 3.2 shows the relationship between number of nodes and the number of possible DAGs.

$$f(n) = \sum_{i=1}^{n} (-1)^{i+1} \frac{n!}{(n-i)!\,i!} 2^{i(n-1)} f(n-1) \tag{3.25}$$

Where *n* is the number of nodes in BBNs. According to this formulation, $f(0)=1$ and $f(10)=4{,}2.10^{18}$. Therefore, searching all possible structures for the optimal network becomes very difficult. Table 3.2 represents the relationship between the number of nodes and possible DAGs.

Table 3.2. Number of directed acyclic graphs as a function of the number of nodes (G) (Source: Scuderi and Clifton 2005)

G(n)	DAGs
1	1
2	3
3	25
4	543
5	29,281
6	3,781,503
7	$1.1\text{x}10^{9}$
8	$7.8\text{x}10^{11}$
9	$1.2\text{x}10^{15}$
10	$4.2\text{x}10^{18}$

The finding an optimal solution for DAG is defined as an NP-hard problem. Due to the complexity of this estimation, learning algorithms (constrained based and search and score algorithms) have been developed. In search and score approach, learning Bayesian network structure can be considered as an optimization problem. The algorithms search for the best model structure from empirical data using a scoring metric. The main idea is to search in the space of all possible bayesian networks (or

DAGs) trying to find the network with optimal score (Abellan, et al. 2006). Different scoring criteria can be used for evaluating the structure. The approach aims at maximizing a scoring function by means of heuristic search algorithms which can determine a bayesian network close to optimum. These algorithms in search and score method can be divided into two groups (Steck 2001, Scuderi and Clifton 2005): a local search algorithm (e.g. Hill Climbing) a global search algorithm (e.g., Markov Chain Monte Carlo). One of the most popular approaches in search and score strategy is K2 algorithm developed in 1992 by Cooper and Herskovitz. K2 algorithm tries to optimize scoring function. K2 algorithm searches over a data set for a bayesian network structure that maximizes the probability of the structure given the data. "*It starts by assuming that a node lack parents after which in every step it adds incrementally that parent whose addition most increases the probability of the resulting structure. K2 stops adding parents to the nodes when the addition of a single parent can not increase the probability*" (Larranaga, et al. 1996, 913).

Hill Climbing Algorithm, Genetic Algorithm, Markov Chain Monter Carlo Algorithm, Tabu Search, Naive Bayes, and Tree Augmented Naive Bayes (TAN) are learning algorithms that based on search and score method. The search algorithms as mentioned above are implemented using by local score metrics (Witten and Fank 2005, Bouckaert 2008). These algorithms evaluate the structure of a bayesian network as a representation of a set of data. Quality measure of a given network is based on some criteria (measures). Two popular measures for evaluating the quality of a network are the Akaike Information Criteria (AIC) and the Minimum Description Length (MDL) criteria. These measures provide score metrics that is used within search algorithms (Witten and Fank 2005).

$$AICscore = -LL + K \tag{3.26}$$

$$MDLscore = -LL + \frac{K}{2}\log N \tag{3.27}$$

Where K is the number of parameters, LL is log-likelihood and N represents the number of instances (or records) in the data (Witten and Frank 2005). In the constrained based approach, the graph (DAG) of a bayesian network is considered as an encoding of a set of conditional dependence and independence relations (CIDRs). Then, structure learning is the task of identifying a DAG structure that best encodes a set of CIDRs from a set of CIDRs derived from the data by statistical tests (Madsen, et al. 2003, Kjaerulff and Madsen 2008). In the constrained based structure learning, validity of independence relationships needs to be tested by statistical hypothesis tests. In this stage, χ^2 test or likelihood G^2 test statistic under the null hypothesis can be used to test and decide independence given subsets of other variables. For example, in the case of marginal independence testing[10] for X and Y variables, the hypothesis to be tested is as follow (Kjaerulff and Madsen 2008):

$$\text{The null hypothesis, } H_0\text{: } P(X, Y) = P(X)\, P(Y), \qquad \text{i.e., } X \perp\!\!\!\perp_p Y \tag{3.28}$$

$$\text{The alternative hypothesis, } H_1\text{: } P(X, Y) \neq P(X)\, P(Y). \tag{3.29}$$

In comparison with search and score approaches, constrained based approaches have some advantages. *"Constrained based approach does not suffer from getting stuck at local optima unlike the search strategies aimed at optimizing a scoring function. For the same reason, equivalent DAG are not a particular problem for constrained – based algorithms"* (Steck 2001, 38). Also, they do not require any prior knowledge. They are computationally easy. There are mainly two different algorithms for structure learning in constrained-based approach: The PC algorithm and the NPC (Necessary Path Condition) algorithm. PC algorithm was developed by Peter Spirtes and Clark Glymour (1991). The main idea behind this is to derive a set of conditional independence and dependence statements (CIDRs) by statistical tests. PC algorithm in learning of the structure of bayesian network performs four main steps (Madsen, et al. 2003):

1. Statistical tests for conditional independence between each pair of variables.
2. Identifying the skeleton of the graph induced by the derived CIDRs.
3. Identifying colliders.
4. Identifying the derived directions or directions of all edges.

[10] The statement $X \perp\!\!\!\perp Y$ is often referred to as marginal independence between X and Y (Kjaerulff and Madsen 2008).

PC algorithm has been proven under the assumptions of infinite data sets. If data sets are finite, PC algorithm can not find the best DAG which represents all CIDRs because of deriving too many conditional independence statements (Madsen, et al. 2003). In this situation, NPC algorithm which is an extension of the PC algorithm should be preferred. NPC algorithm brings a criterion of a Necessary Path Condition as a solution. It is suggested for solving of the problems of constrained based learning algorithms (i.e. PC algorithm). It is developed by researchers at Siemens in Munich (Steck 2001, Steck and Tresp 1999). NPC algorithm like PC algorithm tries to generate a skeleton derived through statistical tests for conditional independence.

The NPC algorithm is based on a criterion of the necessary path condition. There may be inconsistencies among the set of CIDS. Uncertain links result in appear the ambiguous regions. The NPC algorithm allows the researcher to specify uncertain links that need to be directed by user (Kjaerulff and Madsen 2008, Hugin GUI Help 2010). On the other hand, the user or an expert is offered to resolve ambiguous regions. Users can decide the direction of undirected links.

3.3.4.2. Parameter Learning

After a satisfactory dependence is constructed by structure learning algorithms (i.e. PC or NPC algorithms), the parameters of the model that encodes the strengths of the dependences among variables are estimated. A Bayesian network is determined by a graph, *G*, and a set of parameters. Graph represents qualitative component, while parameters represent how the states of a given a node depend on the states of the parents of this node. Structure learning algorithm provided a graph representing the nodes and their dependencies. After that, parameter learning is to learn prior conditional probability distributions given graphical structure and data. In other words, parameter learning is to estimate the values of the parameters (probabilities) from data corresponding to a given DAG structure. In a Bayesian network (or graph), a CPT $P(A/B_1,....,B_n)$ has to attached to each variable A with parents $B_1,.........,B_n$. If A has no parents, unconditional probabilities $P(A)$ must be specified. Like structure learning, parameters of the model can be determined by expert knowledge. The other and efficient method is to use a learning algorithm.

One of the most used parameter algorithms is the Estimation-Maximization algorithm (EM) for estimating the conditional probability distributions in database. EM algorithm developed by Lauritzen (1995) includes two steps: the expectation E step and the maximization M step. The algorithm performs iteratively. "*The EM algorithm is well-suited for calculating maximum likelihood (ML) and maximum a posterior (MAP) estimates in the case of missing data"* (Kjaerulff and Madsen 2008, 206). E and M steps are iterated until convergence or a limit on the number of iterations (threshold) is reached. When the difference between the log-likelihoods of two consecutive iterations is less than or equal to the log-likelihood threshold (δ) times the log-likelihood. The value of δ can be chosen by researcher (i.e. δ=0, 0001) (Kjaerulff and Madsen 2008).

Conditional probabilities are estimated from database by an Expectation - Maximization (EM) algorithm. As mentioned above, the algorithm performs iteratively calculates maximum likelihood estimates for the parameters of the model given the data and the Bayesian network structure of the model. The advantage of this algorithm is to provide an opportunity to handle missing observations (Spiegelhalter, et al. 1993, Laurizen 1995, Heckerman 1996). In sum, searching for a Bayesian network that represent (best) dependence relationships in a data set is difficult because of the large number of possible DAG structure. The task of searching for a good network structure can be found if the right metric is used for scoring (Witten and Frank 2005). Shaughnessy and Livingston (2005) suggested that when using a search and score algorithm over the space of possible graphs to produce a causal network, the choice of scoring function (i.e. Bayesian metric) is much more important than the choice of search method in determining the resulting DAG.

3.4. Empirical Mode Choice Models

3.4.1. Research Design and Methodology

As explained in Section 3.1, the theories of individual choice behavior provides detailed information about decision making process. A major improvement in travel demand modeling is the development of disaggregate travel demand models based on discrete choice models (Ben Akiva and Lerman 1985). Discrete choice models derived from Random Utility Models can be made at aggregate and disaggregate levels according to data source. Especially, multinominal logit model is the most dominant model for travel behavior analysis during the last 25 years, but in recent years, soft computing methods have also been used in travel demand analysis. In existing literature, academic research have heavily focused on disaggregate modeling for analyzing travel behavior (Zhang 2004, Cervero 2002, Pinjari 2007). In the case of the cities in Turkey, there is an empirical gap at both aggregate level and disaggregate level in empirical mode choice studies. Therefore, there is not enough evidence about the factors affecting mode choice decisions in Turkey, especially in terms of land use characteristics.

It is possible to classify mode choice models into two categories: disaggregate and aggregate models. This study includes the modeling of individual behavior (disaggregate) and zonal (aggregate) behavior. At the disaggregate level, the study aims to explain individuals' behaviors for selection of a particular travel mode while the aggregate models used in the study analyze to predict the zonal shares of trips by different travel modes and examine how zonal attributes affect travel mode choice in Istanbul. Contrary to the disaggregate models, the aggregate models require characteristics of travel zones in terms of zonal averages (e.g., average household income, the number of cars per 1000 people, and average household size).

The presented study includes two different goals. One is to test whether the land use characteristics affect mode choice decisions or not at both levels. The other one is to compare the performance of traditional (logit) models and alternative model (BBNs) in mode choice analysis at both levels. To achieve the aim of the thesis, the research framework of the study is shown in Figure 3.7. The research design used in this study summarized as follows:

1. Problem definition within the context of the travel demand modeling,
2. Comprehensive literature review.
3. Discussing the research methods, empirical results, data and data sources, variables, and empirical models used in the empirical studies.
4. Defining a study area: The boundaries of Istanbul Metropolitan Municipality are chosen as a case study.

5. Data gathering: 2006 Household Travel Survey and zonal land use data including zonal averages of socioeconomic characteristics and the size of different land use types.

6. Data manipulation: Preparing the appropriate data structures for logit models and BBNs.

7. Testing different model formulations. Choosing the best model and its specification best fit to the data.

8. Running the models.

9. Performance comparison and evaluating of the model results.

10. Deriving out the general conclusions for existing situation and future studies.

In order to compare the performance of different models (traditional and alternative method) at both levels, the database is divided into two sub sets: training data set and testing data set. Training data set is used for building a model (or developing the model) while testing data set that was not used in the training process is used for comparing the predictive ability of the models. If statistical performance of multinomial logit and probit models are similar, logit model is generally used because of its computational easiness. In this case, multinomial logit model (MNL) is selected as a discrete choice model due to similar performance with probit model in disaggregate level.

Limdep Nlogit Version 4.0 and SAS softwares are used to estimate logit models whereas Belief Network (BN) PowerConstructor and Hugin Researcher 7.1 are used to estimate and compile bayesian belief networks. Hugin and BN PowerConstructor are software programs including learning algorithms.

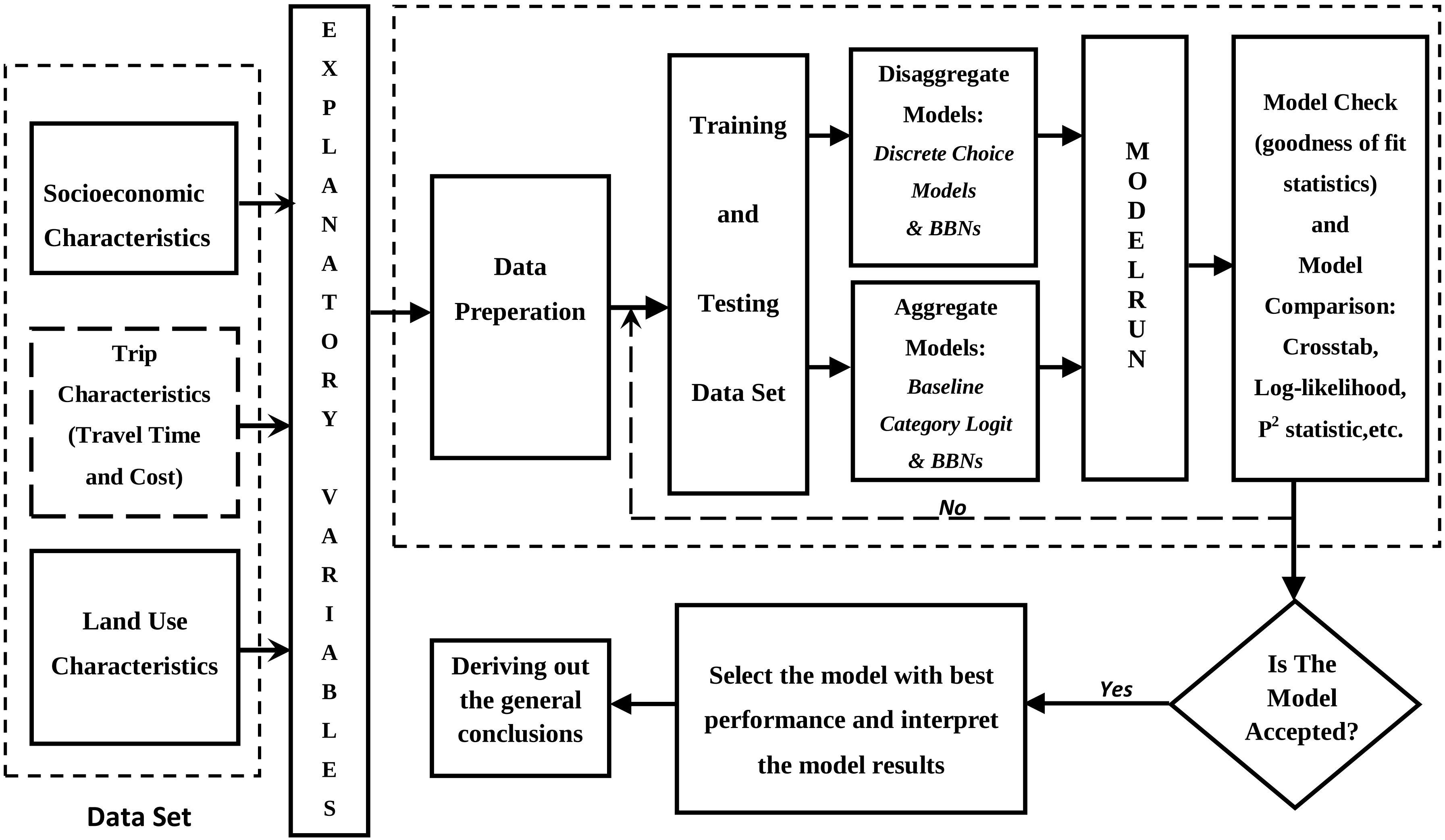


Figure 3.7. Research Framework

In transportation planning, urban areas (or city region) are divided into analysis zones. They are called as "*travel analysis zones*" (TAZ). These zones are expected to exhibit homogeneous land use and population structure. Also, it is expected that the zones generate equivalent daily traffic. A journey between travel zones is called **trip**. In 2006 Household Survey Data in Istanbul, a trip was defined as all types of motorized mobility and walk mobility that do not come back to origin within 15 minutes. Trips are divided into four categories by trip purpose[11]: home - based work (HBW), home - based school (HBS), home - based other (HBO), and non home - based trips (NHB). If origin or destination of trips is at home and other point is at work, this trip is called as home based work trip. Non home - based trip means that a trip does not start or end at home while home - based school trip means that a trip starts at home and end at school. Home based other trip means that origin or destination of a trip at home, the other point of the trip is not at work and school. In the content of the study, *mode choice models are calibrated by only home - based work trips (HBW) at both aggregate and disaggregate level*.

In the calibration of transportation demand modeling prepared for 2007 Istanbul Transportation Master Plan, four main modes were determined. In the case, mode choice models at both levels are estimated by the four main modes: walk, auto, service, and transit as seen in the Figure 3.8. Mode related variables describing the alternatives to the travelers (e.g., travel time and travel cost) were only estimated by these four main modes in TRANSCAD. In this case, choice set includes mainly these modes due to the data availability. According to this aggregation, walk mode includes walk and bicycle. Car mode includes auto drive alone, auto shared ride, taxi and motorcycle. Public transportation (sometimes known as public transit or mass transit) aims to serve only public in opposition to private modes. In Istanbul, there are several public transportation modes. Public transit modes include dolmuş, minibus, public bus, private bus, metro, light metro, tram, funicular, ferry, sea bus, sea motor, suburb train, and other vehicles. Service mode includes only personnel service vehicles.

[11] Trip purpose sometimes can be categorized as work trips, shopping trips, social-recreation trips, and business trips. Detailed discussion is included in Meyer and Miller (2001).

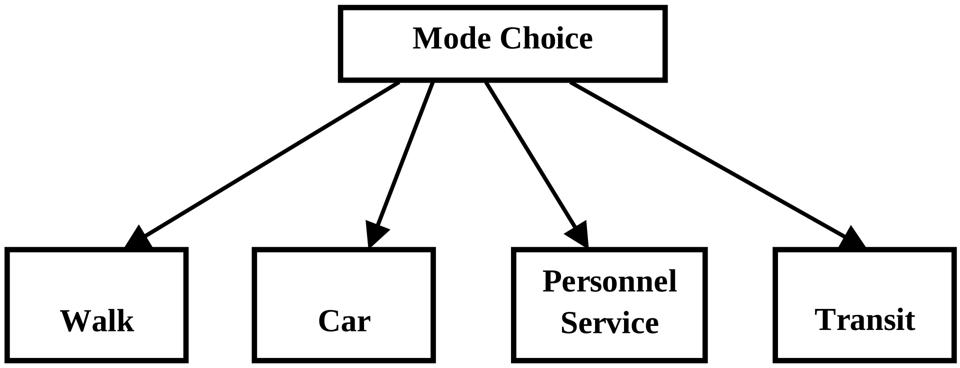


Figure 3.8. Travel modes in Istanbul

The empirical analysis of mode choice in Istanbul includes baseline category logit model and bayesian belief networks at aggregate level while multinomial logit model (discrete choice model) as a traditional (conventional model) and bayesian belief networks as an alternative approach are used at disaggregate level as seen in Figure 3.7. Following models are developed in the content of the study at both levels:

1. Conventional (logit) models using all input variables with different model specifications.
2. The models using selected input variables with statistically significant are determined. In empirical studies, these are called as the most efficient or the best conventional models.
3. Bayesian Belief Networks (BBNs) as an alternative method using most efficient input variables and different algorithms is constructed by train data set.
4. Performance analysis provides information about how well the predictions of the models match the observations using with test data set.

For conventional models, the models are estimated firstly for all possible input variables, and then in order to find the smallest possible number of input variables, models are re-run. For each run, some variables whose have low explanation levels (low t statistics) are excluded from the data set. After several model runs, most efficient models are found. The study pay enough attention to that these attempts will not result in reducing performance of the models. Model calibration process for soft computing method (BBN) is performed in Hugin and BN PowerConstructor while the process for conventional models (logit models) is performed in Limdep (Nlogit Version) and SAS. ArcGis software is used for measuring spatial data. The origin and destination locations derived from 2006 Household Travel Survey for Istanbul are matched and integrated into a geographic information system (GIS) based land use database.

The models are estimated using two sets of data separately and mainly for base model and expanded model. While the base model include variables typically considered in the analysis of mode choice (socioeconomic and travel characteristics), expanded model includes land use variables with other independent variables in the base model. It is expected that this separation provide information about the marginal impacts of land use on mode choice exactly. This approach provides information to make comparison of previous studies in the literature.

3.4.2. Variables

Firstly, the factors influencing mode choice in Istanbul are classified into three groups: socioeconomic characteristics, travel characteristics (time and monetary cost), and spatial configuration of land use at disaggregate level. At aggregate level, since a non origin and destination based mode choice model is estimated, trip characteristics (time and cost) are not entered into the models. At disaggregate level, disaggregate models based origin – destinations (OD) are calibrated. Therefore, the variables entered into the models differ from each other. Before empirical mode choice models are presented, it is worthwhile to explain how and which land use characteristics enter into the models. In this stage, existing literature provide guidance about the relationship. For example, Kockelman (1997) suggested four measures of land use: *entropy index of land use balance, dissimilarity index, accessibility, and density*. Cervero and Kockelman (1997) suggested that built environment is defined in terms of three core dimensions (3Ds): ***density, diversity, and design***. Among these dimensions, design variables needs urban form data at parcel level such as pedestrian and cycling provision and site design variables. Urban form and design variables are omitted from the models due to lack of empirical data in Istanbul. As mentioned before, there is no evidence for the relationship between travel demand and land use in Istanbul. The selection of model variables began by collecting represents from previous empirical studies. Several land use variables are determined by existing literature. For example, density and land use mix are common land use measures. However, different formulations can be applied to obtain these variables. Moreover, some of land use variables and their formulations highly depended on the data availability. In sum, land use variables in the content of the study are classified into three categories: density, diversity, and accessibility for aggregate and disaggregate models. Main reason for this classification is data availability and the other one is to make comparisons of the results with the previous studies.

On the other hand, one of the common problems in logit and regression analysis is multicollinearity which occurs when there are strong relationships (or dependency) among the explanatory variables. In the presence of multicollinearity, standard errors may have large values. Also, correct effects of the explanatory variables cannot be detected. For diagnosing multicollinearity, some diagnostic measures are used: the correlation coefficients for all pairs of explanatory variables, tolerance and variance inflation of explanatory variables. Following Kennedy (1981), the explanatory variables whose correlation coefficients smaller than 0.7 among explanatory variables, are entered into the models. The model variables are selected using stepwise method at aggregate level. Many model specifications are tested to find the best model. From a set of explanatory variables, only 14 variables (8 for land use characteristics and 6 for socioeconomic characteristics) are entered into the aggregate models. At the disaggregate level, there are 12 variables in total. However, land use variables are measured at both origin and destination. With alternative specific variables, discrete choice models include 26 model parameters in the expanded form. Final model variables are explained in Table 3.3. The difference between model variables and their formulations for aggregate and disaggregate models are summarized in Table 3.4 and Table 3.5.

Table 3.3. Variables used in the empirical models

Variable Set	Labels	Empirical Models
A. Socieconomic Variables:		
Household Income	(HHINC)	both
Number of Cars in Household	(NCAR)	both
Number of Company cars in Household	(CCAR)	disaggregate
House Ownership	(HOWNR)	aggregate
Household Size	(HHSIZE)	aggregate
The Zonal Average of Worker	(WRKR)	aggregate
Driver's License	(DRL)	disaggregate
The Presence of Akbil Card (unlimited or not)	(AKBIL) (SAKBIL)	disaggregate
B. Travel Time and Cost (Generic Variables):		
Travel Time (in minutes)	(TT)	disaggregate
Travel Costs (as monetary)	(TC)	disaggregate
C. Land Use Variables:		
C.1. Density		
Employment / Population Density	(EPDENS)	aggregate
Population Density	(PDENS)	both
Industrial Employment Density	(IEDENS)	aggregate
Commercial Employment Density	(CEDENS)	aggregate
Commercial and Industrial Area Density	(CIDENS)	aggregate
C.2. Diversity		
Jobs - Housing Balance[12]	(JHB and EWDENS)	both
Land Use Mix (Dissimilarity Index)	(LUMIX)	aggregate
C.3. Accessibility		
Transit Accessibility	(TRACC)	both
Other Land Use Variables[13]:		
Intra-Zonal Travel	(INTRA)	disaggregate
Zonal Area	(AREA)	aggregate

[12] In the content of the study, two different formulations are used to estimate JHB ratio in order to make comparisons with the existing literature.

[13] Since these variables cannot be categorized under three-category for land use, the variables are tested independently in the models.

The variables characterizing the socioeconomic characteristics of individuals or travel zones differ according to the aggregate (zonal) and disaggregate data. At aggregate (zonal) level, the variables include zonal averages while the disaggregate model variables represent individual characteristics. Socioeconomic variables are income (**hhinc**), the presence of driver license (**drl**), the presence of akbil card type (**sakbil** and **akbil**), car ownership (**ncar**), company car (**ccar**), household size (**hhsize**), house ownership (**hownr**), zonal average of worker (**wrkr**). Among the variables, the presence of driver license (**drl**), the presence of akbil card and unlimited akbil card (**akbil** and **sakbil**), and transit accessibility (**tracc**) are used as dummy variables (1 or 0). The magnitudes and the signs of these variables depend on the choice of travel mode. For example, it is expected that three dummy variables (akbil, sakbil, and tracc) are positively correlate with transit mode. In addition to these variables, a variable (**intra**) is created for disaggregate OD - based models that measure trips which begin and end in the same travel zone. It is measured as dummy variable for walk mode.

A total of two generic variables characterizing the attributes of alternatives (mode related variables) is used: travel time (**tt**) and travel cost (**tc**). Travel time is measured in minutes whereas travel cost is measured in Turkish Lira (TL). Theoretically, as travel time and cost increases, travelers prefer alternatives with lower time and cost. Therefore, the expected sign of the generic variables are negative, indicating a disutility.

The variables at aggregate level are household income (**hhinc**), house ownership (**hownr**), car ownership (**ncar**), worker rate (**wrkr**), household size (**hhsize**), the size of zonal area (**area**), employment / population density (**epdens**), industrial employment density (**iedens**), population density (**pdens**), commercial employment density (**cedens**), commercial & industrial area density (**cidens**), jobs - housing balance (**jhb**), land use mix (**lusemix**), and transit accessibility (**tracc**). There is no correlation problem among the variables.

The variables at disaggregate level are household income (**hhinc**), the number of cars in household (**ncar**), the number of company cars in household (**ccar**), the presence of driver license (**drl**), the presence of akbil card (**sakbil - akbil**) used in public transport in Istanbul, travel time (**tt**) and travel cost (**tc**) for each mode, and land use variables. Land use variables characterizing the origin and destination are employed in the empirical analysis. These variables are population density (**pdens**), employment /

worker ratio (**ewdens**) as an indicator of jobs - housing balance, and the presence of transit access (sea or metro) in relative travel zone (**tracc**). In addition, there is a variable (**intra**) implying the trips which begin and end in the same traffic zone as dummy variable.

Regarding the land use variables, all of the land use variables are estimated at zonal level. The variables characterizing density are population density (**pdens**), employment per population density (**epdens**), industrial employment density (**iedens**), commercial employment density (**cedens**), and commercial & industrial area density (**cidens**). Employment / population density (**epdens**) is estimated by dividing total employment by population in that zone. Industrial employment density (**iedens**) presents the ratio of industrial employment to the size of each travel zone as hectare. Commercial employment density (**cedens**) is estimated by dividing total commercial employment by the total size of each zone. On the other hand, commercial and industrial area density (**cidens**) is found by dividing the total of commercial and employment areas to zonal area in that zone.

Density variables include population and employment densities for each zone. Population density is generally defined as the number of individuals per given unit of zonal area (person/hectare or person/square kilometers). Employment density is measured by total area of any sectoral employment per hectare such as the size of industrial employment in that zone. Also, employment density can be estimated by dividing total sectoral employment by total zonal area such as workers per hectare. Several empirical studies in Europe and USA (Newman and Kenworthy 1989, Schwanen, et al. 2004) suggested that higher population densities negatively correlated with the use of private car trips. It is positively correlated with public and walking trips. According to the report of National Academy of Sciences (2009), "*increasing population and employment density in metropolitan areas could reduce vehicle travel, energy use, and CO2 emissions from less than 1 percent up to 11 percent by 2050 compared to a base case for household vehicle usage"*. In the case of Istanbul, there is no evidence. In the content of the study, several density measures related to density are tested.

Another important land use dimension is land use diversity indicating the degree of land use composition. Two indexes are generally used in empirical studies: land use mix and land use balance. In this study, a land use mix diversity index is used (**lusemix**) similar to Rajamani et al. (2002), Bhat and Gossen (2004), and Bhat and Guo (2007) as seen in the Equation 2.3. Land use mix index indicates proportion of dissimilar land use types in that zone (percentages of zonal area in residential, commercial, industrial, and other land uses).

Another diversity measure is jobs - hosing balance indicating that imbalance between workplace and residential areas increases traffic congestion. As mentioned before, the jobs - housing balance ratio (JHB) can be measured in a number of different ways. In the content of the study, two different formulations are developed. The first one is based on the formulation developed by Cervero (1989, 1996). The formulation is written as:

$$JHB = \frac{workers}{employedresidents} \tag{3.30}$$

This ratio express quantitatively the relationship between number of workers in a city and number of residents in a city who are employed. In the case, employment / worker ratio (**ewdens**) is estimated by dividing total employment to the total number of workers in that zone. The second JHB formulation is written as follows:

$$1 - \left| \frac{(E_i - c \times P_i)}{(E_i + c \times P_i)} \right| \tag{3.31}$$

Where *E* presents employment size and *P* is the population size at each relative zone. c presents activity rate measured at the zonal level. It is the ratio of people who is capable of work in the relative zone to zonal population. The value of this variable ranged from 0 to 1. 0 represents a pure residential area or a non-residential area while 1 represents a balance between employment and population. Theoretically, if jobs - housing balance occurs, people want to live and work in the same area. It can be expected that long trips would be avoided (Cervero 1989, Sultana 2002, Wang and Chai 2009). In other words, good jobs - housing balance means that there may be short work commute and more non - motorized trips.

The third category of land use is accessibility index. The variable for transit accessibility variable (**tracc**) is created for measuring residential sorting (or self-selection) effects as dummy variable. In this case, transit accessibility represents the presence of transit (rail or sea) in relative TAZ. It is expected that the presence of transit accessibility positively correlated with the choice of transit mode.

The following hypotheses associated with land use variables are derived from the relationship between mode choice and land use in Istanbul.

H.1. Population density is positively correlated with walking and transit mode choice.

H.2. Employment densities are positively correlated with motorized trips.

H.3. Diversity positively correlates with walk and transit mode choice.

H.4. Transit access increases the choice of transit mode.

H.5. Commuters whose trip origin and destination point is in the same zone are more likely to choose non-motorized alternatives.

Table 3.4. Aggregate model variables

Variables	Description
A. Household Socioeconomic Variables	
Zonal Area (area)	Zonal area as hectare (hectare / 100).
Household Income (hhinc)	Zonal average of household income (income (T.L.) / 1000).
Household Size (hhsize)	Zonal average of household size.
House Ownership (hownr)	Zonal average of household ownerships.
Car Ownership (ncar)	The number of car per 1000 people.
Employed (wrkr)	Zonal average of worker.
B. Land Use Variables	
1. Density	
Employment / Population Density (epdens)	The ratio of total employment to total zonal population within each zone.
Population Density (pdens)	Population per zonal area (person/hectare).
Industrial Employment Density (iedens)	Number of industrial employment per zonal area.
Commercial Employment Density (cedens)	Number of commercial employment per zonal area.
Commercial and Ind. Area Density (cidens)	The ratio of total size of commercial and industrial area within each zone to total zonal area.
2. Accessibility	
Transit Accessibility (tracc)	The presence of transit access in each zone.
3. Diversity	
Job - Housing Balance (jhb)	The degree of land use balance between jobs and residents at the zonal level.
Land Use Mix (lumix)	The degree to which land uses are mixed within each zone.

Table 3.5. Disaggregate model variables

Variables	Description
A.Socioeconomic Variables of Trip Maker	
Individual Income (hhinc)	Individual income as monthly.
Driver License (drl)	The presence of driver license as dummy variable.
Unlimited Akbil Card (sakbil)	The presence of unlimited akbil card for public transport vehicles as dummy variable.
Akbil Card (akbil)	The presence of akbil card for public transport vehicles as dummy variable.
Car Owner (ncar)	The number of auto in household.
Company Car (ccar)	The number of company car in household.
B. Travel (Generic) Variables	
Travel Time (tt)	Travel time by each mode.
Travel Cost (tc)	Travel cost by each mode.
C. Land Use Variables	
1. Density	
Population Density (pdens)	Population per zonal area (person/hectare).
2. Accessibility	
Transit Accessibility (tracc)	The presence of transit access in each zone as dummy variable.
3. Diversity	
Employment / Worker Ratio (ewdens)	The ratio of total employment to total number of worker within each zone.
Intrazonal (intra)	Origin and destination of hbw trip is in the same zone or not as dummy variable

3.4.3. Model Structure and Formulations

3.4.3.1. Baseline Category Logit

Logit models provide an efficient model to analyze travel demand. When response variable is binary (0 or 1), binary logistic regression models the probability of an event that is occurred or not. In this case, since choice set includes four modes (or four categories), this type of logit models is called as “multicategory logit”. Multinomial responses can be divided into two categories: nominal (unordered categories) and ordinal responses. If response categories are ordered, cumulative logit models are preferred. Mode choice problem in the case is an example of nominal (unordered) response. Baseline category logit model is used for nominal responses. Baseline category logit model compares each group with a reference group simultaneously. In this choice analysis, baseline category logit model compares walk as an unmotorized mode with car, service, and transit modes sequentially. In other words, walk mode is used as the baseline category. Baseline category logit model only selects the set of the variables as the best subset of variables using different selection methods such as stepwise, forward, and backward. In the content of the study, stepwise selection method is used. The three logit equations described the log odds that people who live in the zones in Istanbul select other primary travel modes instead of walk. According to the formulation of baseline category (or generalized logit) logit model, the probabilities for each mode can be written as:

π_1=probability of walk mode,
π_2=probability of car mode,
π_3=probability of service mode,
π_4=probability of transit mode.

The logit equation for car mode is as below:

$$\log\left(\frac{\pi_2}{\pi_1}\right) = \log\left(\frac{\pi_{car}}{\pi_{walk}}\right) = \beta_0 + \beta_1 X_1 + \ldots\ldots\ldots + \beta_k X_k \tag{3.32}$$

Baseline category logit models are estimated for only socioeconomic data and then the whole data set. Therefore, the effects of land use characteristics are analyzed separately from other factors affecting mode choice. For testing the goodness of fit in this model, several test statistics are used. Deviance and Pearson Chi-Square test statistics are used for testing the goodness of fit of the models. In other words, these test statistics provide important measures for model check. Maximum likelihood analysis of variance is used for detecting statistically significant variables. In the content of the study, the hypothesis is tested that expanded model outperforms the base model with only socioeconomic variables. The null hypothesis (H_0) for the case is as follows:

H_0 = The base model with socioeconomic characteristics fits the data.

H_1 = The expanded model fits better.

Also, the parameter estimate is estimated by maximum likelihood estimator instead of weighted least square estimator. Finding the best set of variables for the models, stepwise method is used as a variable selection method. The significance level for selecting variables is performed at the 0.01 % level of significance. Most of the variables obtained from household travel survey are eliminated from the models because of low significance levels and multicollinearity problems. The remaining variables are entered into the models. The model variables are presented in Table 3.3. PROC LOGISTIC and CATMOD statement in SAS is used for binary and also nominal response outcomes. Baseline category logit model is estimated in SAS software using both PROC LOGISTIC and PROC CATMOD. Mode choice data are arranged in the frequency format instead of individual data. Since only SAS Proc Logistic (or Proc Catmod) procedure allows the input of binary response data that are grouped. Proc Logistic procedure in SAS software is used for mode choice data that are grouped[14].

[14] Further detailed desciptions of logit models are contained in Allison (1999), Agresti (2002), and Hosmer and Lemeshow (2000).

3.4.3.2. Multinomial Logit Models

In the content of the study, one of the empirical mode choice methods at disaggregate level in Istanbul includes discrete choice models developed from consumer choice theory (Ben-Akiva and Lerman 1985, Domencich and McFadden 1975). The modeling framework used for estimating the probability a commuter opted for a particular mode in Istanbul is expressed in terms of multinomial logit model as:

$$P_{niod} = \frac{e^{V_{niod}}}{\sum_{j \in C_n} e^{V_{njod}}} \tag{3.33}$$

Where P_{niod} is the probability of an individual *n* choosing mode *i* for home based work travel between origin (o) and destination (d). C_n represents the choice set and V_{niod} is the utility function. V_{niod}, deterministic (systematic) component of utility function, includes alternative specific constant (ASC), travel attributes or generic variables (TT and TC), socioeconomic variables (SE), and land use variables (LU). In order to measure the effect of land use, a series of logit model is estimated in Nlogit software. Four different multinomial logit model specifications for home - based work trips in Istanbul are developed. The MNL models and their forms are as follow:

$$\begin{aligned} U_{walk} &= V_{walk} + \varepsilon_{walk} \\ U_{transit} &= V_{transit} + \varepsilon_{transit} \\ U_{auto} &= V_{auto} + \varepsilon_{auto} \\ U_{service} &= V_{service} + \varepsilon_{service} \end{aligned} \tag{3.34}$$

The probability that a choice response is observed is written as a function of a set of explanatory variables as follows:

A. The Base Model: Only alternative specific constants (ASC). The model took the form as follows:

$$P_{niod} = \exp(V_{niod} = f(ASC) / \sum_{j \in Cnod} \exp(Vnjod = f(ASC) \tag{3.35}$$

B. Model 1: Only alternative specific constants and travel attributes (time and cost). The model took the form:

$$P_{niod} = \exp(V_{niod} = f(ASC, T_{iod})) / \sum_{j \in Cnod} \exp(Vnjod = f(ASC, T_{jod})) \quad (3.36)$$

C. Model 2: Adding socioeconomic variables to the Model 1. Model 2 took the form:

$$P_{niod} = \exp(V_{niod} = f(ASC, T_{iod}, SE_n)) / \sum_{j \in Cnod} \exp(Vnjod = f(ASC, T_{jod}, SE_n)) \quad (3.37)$$

D. Model 3 (Expanded Model): Adding land use variables to Model 2. Model 3 took the form:

$$P_{niod} = \exp(V_{niod} = f(ASC, T_{iod}, SE_n, LU_{od})) / \sum_{j \in Cnod} \exp(Vnjod = f(ASC, T_{jod}, SE_n, LU_{od})) \quad (3.38)$$

Firstly, the base model is estimated. After Model 1 and Model 2, expanded model finally is estimated. The expanded model includes alternative specific constants, travel time and cost (generic variables), and land use attributes. Land use attributes both at trip origin and destination are entered into the utility functions. Comparisons of the equation as explained above allow marginal effects of adding socioeconomic and land use variables to mode choice utility function to be measured. It is expected that expanded model statistically improves models' explanatory level. This hypothesis is tested in terms of different goodness of fit criteria such as changes in the log likelihood function and pseudo R^2. In the models, estimated parameters represents the impact of the explanatory variables used in the models on the utility of the alternatives. The utility specification of the expanded model is given in Table 3.6. The utility functions according to the models are as follow:

Base Model: Only ASC Variables

$$\begin{aligned} V_{walk} &= \beta_1 \\ V_{transit} &= \beta_2 \\ V_{auto} &= \beta_3 \end{aligned} \tag{3.39}$$

Model 1: ASC + GENERIC Variables

$$\begin{aligned} V_{walk} &= \beta_1 + \beta_4 tt \\ V_{transit} &= \beta_2 + \beta_4 tt + \beta_5 tc \\ V_{auto} &= \beta_3 + \beta_4 tt + \beta_5 tc \\ V_{service} &= \beta_4 tt + \beta_5 tc \end{aligned} \tag{3.40}$$

Model 2: ASC + Generic + Socioeconomic Variables

$$\begin{aligned} V_{walk} &= \beta_1 + \beta_4 tt \\ V_{transit} &= \beta_2 + \beta_4 tt + \beta_5 tc + \beta_7 sakbil + \beta_8 akbil \\ V_{auto} &= \beta_3 + \beta_4 tt + \beta_5 tc + \beta_6 drl + \beta_9 hhinc + \beta_{10} ncar + \beta_{11} ccar \\ V_{service} &= \beta_4 tt + \beta_5 tc \end{aligned} \tag{3.41}$$

Model 3: (Expanded Model): FULL DATA

$$\begin{aligned}
& V_{walk} = \beta_1 + \beta_4 tt + \beta_{12}\,\mathrm{int}\,ra + \beta_{13} ewdens(origin) + \beta_{16} ewdens(destination) \\
& \beta_{19}\, pdens(origin) + \beta_{22}\, pdens(destination) \\[1em]
& V_{transit} = \beta_2 + \beta_4 tt + \beta_5 tc + \beta_7 sakbil + \beta_8 akbil + \beta_{14} ewdens(origin) + \\
& \beta_{17} ewdens(destination) + \beta_{20}\, pdens(origin) + \beta_{23}\, pdens(destination) \\
& + \beta_{25} tracc(origin) + \beta_{26} tracc(destination) \\[1em]
& V_{auto} = \beta_3 + \beta_4 tt + \beta_5 tc + \beta_6 drl + \beta_9 hhinc + \beta_{10} ncar + \beta_{11} ccar + \\
& \beta_{15} ewdens(origin) + \beta_{18} ewdens(destination) \\
& \beta_{21}\, pdens(origin) + \beta_{24}\, pdens(destination) \\[1em]
& V_{service} = \beta_4 tt + \beta_5 tc
\end{aligned} \tag{3.42}$$

Table 3.6. The systematic utility function of disaggregate OD – based multinomial logit model

	ASC	ASC	ASC	TRAVEL TIME	TRAVEL COST	DR. LICENSE	TRSAKBIL	TRAKBIL	INCOME	HH AUTO OWNERSHIP	COMPANY AUTO
	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}	β_{11}
WALK	**ONE**	**0**	**0**	**tt**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
TRANSİT	**0**	**ONE**	**0**	**tt**	**tc**	**0**	**sakbil**	**akbil**	**0**	**0**	**0**
AUTO	**0**	**0**	**ONE**	**tt**	**tc**	**drl**	**0**	**0**	**hhinc**	**ncar**	**ccar**
SERVICE	**0**	**0**	**0**	**tt**	**tc**	**0**	**0**	**0**	**0**	**0**	**0**

	Intra Zonal Travel	Emp. / Worker (JHB)	Population Density	Transit Access
	β_{12}	β_{13}-β_{18}	β_{19}-β_{24}	β_{25}- β_{26}
WALK	**intra**	**ewdens**	**pdens**	**0**
TRANSİT	**0**	**ewdens**	**pdens**	**tracc**
AUTO	**0**	**ewdens**	**pdens**	**0**
SERVICE	**0**	**0**	**0**	**0**

3.4.3.3. Bayesian Belief Networks

Bayesian belief networks (BBNs) as a graphical model provide guidance about causal relationships between variables. In other words, BBNs represent dependencies and independencies among variables using directed acyclic graph. This is a part of qualitative (structural) of the network. On the other hand, quantitative (probabilistic) part is represented by conditional probability tables for each node in the network. The most important difference between conventional models (regression and logit models) and soft computing methods (artificial neural network, fuzzy logic, etc) is that conventional models provide information about the signs of the model parameters and statistical significance while soft computing methods may not. However, BBNs provide a graphical model that shows the direction and strength of the relationships among the variables while other methods maynot. Although this model has gained popularity in environmental sciences, decision support systems, healthcare management, medical diagnostic problems, and risk assessment in recent years, the application in transportation modeling is rather limited. In the content of the study, BBNs are developed to investigate the causal relationships among the variables. Also, the purpose of the models is to predict the choice probabilities of travel modes. BBNs are estimated at both aggregate and disaggregate level. The process to develop a Bayesian network in this case is summarized as follows:

1. Deciding what variables and their states should be included into the models.
2. Discretization of the model variables.
3. Building a Bayesian network structure with train set using expert knowledge (domain knowledge) and learning algorithms.
4. Creating conditional probability tables for each node in the network using expert knowledge and learning algorithms.
5. Compiling the network and inferences in Hugin software.
6. Sensitivity analysis and performance measures of BBNs models using test set.

The most disadvantages in studying a BBNs is computation time of learning algorithms from data. The size of conditional probability table (CPT) expands as model variables increases. The conditional probability tables (CPTs) of the Bayesian network becomes too large. The size of a CPT grows exponentially with the number of parents. To reduce the number of parents, the significance level should be adjusted. However, this adjustment was inadequate for performing algorithms in Hugin. Therefore, only the variables that are statistically significant according to the result of logit models are selected. In other words, the variables that did not contribute the explanatory power of logit models were eliminated from the networks. For example, the variables except jobs - housing balance at both origin and destination are entered into the network at disaggregate level. HUGIN runs out of memory due to the large size of CPTs. Because of this, BN PowerConstructor software is preferred. After deciding model variables, the most of the softwares developed for Bayesian networks needs to discretize continuous variables for applying learning algorithms. For example, Hugin includes equal distribution and equi-distance methods. Since the existing literature cannot provide enough guidance, discretization process is performed according to expert knowledge. At both levels, data structure differs from each other. Therefore, the states of the models (nodes) can vary as seen in Table 3.7 and Table 3.8.

In BBNs, for building network structure (or developing model), 80% of the case file randomly selected from empirical data is used in learning or training. The learned BBNs are tested on the random subset of 20% of the case file that was not used in the learning process. There are two alternatives in BBNs to construct a network: expert knowledge and learning algorithms from data. Learning algorithms in BBNs can be divided into two groups: structural and parametrical learning from data. As mentioned before, parametric learning determines CPTs of each node of a network while structural learning determines the causality among the variables in a network. In the literature related to BBNs, there are many learning algorithms. Different softwares may include different learning algorithms such as search and score and dependency analysis methods. These methods are expected to find the correct structure. BN PowerConstructor is applied to construct bayesian belief networks and estimate CPTs. The method used in BN PowerConstructor for structural learning from data is based on dependency analysis. The method requires conditional independence (CI) tests. Since the algorithms cannot detect exact relationships among the variables, some relationships from the existing literature were constructed as manual. Therefore, both approaches (manual and automatically) are applied in structural learning.

Table 3.7. The model variables used in BBNs at disaggregate level

Variables (Nodes)	Label	States
Walking Time	WTIME	8 states
Transit Time	TRTIME	8 states
Auto Time	ATIME	10 states
Service Time	STIME	6 states
Transit Cost	TCOST	8 states
Auto Cost	ACOST	5 states
Service Cost	SCOST	6 states
Driving License	DRL	2 states
Income	INCOME	6 states
Akbil Card Usage	AKBIL	2 states
Unlimited Akbil Card Usage	SAKBIL	2 states
The Number of Car in HH	NCAR	3 states
The Number of Company Car in HH	CCAR	2 states
Intra Zonal Travel for Walk Mode	INTRA	2 states
Emp. / Worker Density at origins	ORATIO3	10 states
Emp. / Worker Density at destinations	DRATIO3	11 states
Pop. Density at origins	OPDENS	3 states
Pop. Density at destinations	DPDENS	4 states
The Presence of Transit Access at origins	OTRACC	2 states
The Presence of Transit Access at destinations	DTRACC	2 states
Mode Choice (The Query Node)	MODECHOICE	4 States: Walk (1), Transit (2), Car (3), Service (4).

The network structures are detected from mode choice data in BN PowerConstructor and then some relationships are derived manually (semi-automatic). After building bayesian network structure, conditional probability tables are derived from empirical data with parametric learning algorithm in BN PowerConstructor. The network is built up in the software program BN PowerConstructor. After that, Hugin compiles the network. The last step is to test how well the predictions of the network match the actual cases and make sensitivity analysis. In order to do this, data accuracy pane using testing data in Hugin is used to calculate different scores and generate an analysis report.

Table 3.8. The model variables used in BBNs at aggregate level

Variables (Nodes)	**Label**	**States**
Household Income	HHINC	3 states
Household Size	HSIZE	3 states
The Number of Car per 1000 People	NCAR	3 states
House Ownership	HOWNR	2 states
Working	WRKR	2 states
The Size of Zonal Area	AREA	3 states
Employment / Population Density	EWDENS	3 states
Pop. Density	PDENS	3 states
Job – Housing Balance	JHB	4 states
Land Use Mix	LUSEMIX	3 states
Industrial Employment Density	IEDENS	4 states
Commercial Employment Density	CEDENS	3 states
Com. & Ind. Emp. Area Density	CIDENS	2 states
The Presence of Transit Access	TRACC	yes, no
Mode Choice (The Query Node)	MODE CHOICE	4 states: Walk(1), Car (2), Service (3), Transit (4).

The softwares used in BBNs such as Hugin and Netica provide an analysis report that is used to model assessment. This analysis report includes some scoring rules, error rate, a confusion matrix, and ROC curve. Quadratic loss and spherical payoff are the most used scoring rules. Quadratic loss ranges from 0 to 2, with 0 being. The formulation of spherical payoff is written as follows (Marcot, et al. 2006):

$$MOAC = \frac{P_c}{\sqrt{\sum_{j=1}^{n} P_j^{\,2}}} \tag{3.43}$$

Where MOAC is the mean probability value of a given state and n is the number of states in bayes network. P_c is the probability predicted for the correct state while P_j represents the probability predicted for state j. Spherical payoff ranges from 0 to 1. 1 represents the best model performance. Hugin Researcher Version 7.1 provides its own predictive accuracy scoring measures: Euclidian distance and Kulbach - Leibler divergence[15]. In the content of the study, the scoring rules, Euclidian distance and Kulbach - Leibler divergence, are used. These scoring rules show similarities quadratic loss and spherical payoff that are derived from other softwares used in BBNs. The classification of a BBN model including binary output can be tested with a receiver operating characteristic (ROC) curve (Marcot, et al. 2006, Dlamini 2009). Since the mode choice problem includes multinomial response, ROC curve is not used. Confusion matrix (or crosstab), scoring rules, and error rate are used for model performance of BBNs at both levels.

[15] The formulation for the scoring rules are discussed in www.norsys.com and www.hugin.com.

CHAPTER 4

DATA SOURCES AND PROCESSING

This chapter provides information about the study area, data sources, and processing. Firstly, the case study is described. After that, the household survey data used in empirical models is described. In the content of the study, 2006 Household Travel Survey prepared for 2007 Istanbul Transportation Master Plan is used. Data represents the most recent travel information in Istanbul. This section includes descriptive statistics of empirical data used in aggregate and disaggregates models.

4.1. Description of The Case Study and Istanbul Household Survey Data

The boundaries of Istanbul province are selected as this study. Istanbul is situated on both sides of the Bosporus Strait. The Bosporus Strait divides Istanbul into two parts: the European side and the Asian side. Istanbul is surrounded by the province of Kocaeli in the east, by Marmara Sea in the south, by the Black Sea in the north, and by the province of Tekirdağ in the west as seen in the Figure 4.1. Its history has over 2500 years. The city is the largest city in Turkey with a population of around 12.573 million in 2007 according to official census data based on the address based population registration system while it was 1.078 million in 1945. A list of the population of Istanbul by years is given in Table 4.1. According to this table, the increase in population in the last 10 years has been over 2.5 million. The total area of the city boundaries covers 5512 square kilometers. 8.156.867 people live on the European side while 4.416.867 people live on the Asian side (IBB 2010).

Figure 4.1. The Case Study
(Source: Istanbul Metropolitan Municipality (IBB) - Directorate of City Planning, GIS based Land Use Database)

Table 4.1. Population levels in Istanbul
(Source: Governorship of Istanbul 2010)

Year	Population (person)
1927	806.863
1945	1.078.399
1960	1.882.092
1975	3.904.588
1990	7.195.773
1997	9.198.809
2000	10.018.735
2007	12.573.836
2008	12.697.164

The city includes 32 districts[16] in total. The eleven counties are located on the Asian side while the others are located on the Europeanside. The spatial configuration of land use in Istanbul is displayed in the Figure 4.2. According to the Figure 4.2, green areas (dark and light green) in the fringe show the forests and agricultural areas whereas brown areas show urban areas (residential). The most of the areas for the districts of Catalca, Silivri, Beykoz, and Sile includes non-residential areas. 58.4 % of forest areas is located on European side while 41.6 % of them is located on Asian side. Commercial areas area concentrated on the residential areas. The total industrial area is 10.476 hectare in Istanbul metropolitan area. The industrial firms that needs large industrial areas are located in Maltepe - Kartal districts and Kağıthane in the west side. The industrial firms have been located in Tuzla, and Küçükçekmece associated with developing highways (IBB 2005). Regarding urban transportation in Istanbul, there are two bridges connecting the continents in the city. These bridges carry a heavy load of commuting and intercity traffic. The direction of commuting trips in the morning is toward the CBD whereas this direction in the evenings towards the fringes. In Istanbul, highway is 232 kilometers in length while public road is 324 kilometers in length. On the other hand, the existing rail systems in Istanbul consist of tram, funicular, teleferic, light rail transit (LRT), and metro as seen in Figure 4.3. The properties of the rail systems are summarized in Table 4.2.

[16] During transportation master plan studies, Istanbul included 32 districts. The number of the districts increased from 32 to 39 in 2008. 25 of them are located in Europeanside whereas 14 of them are in Asian side.

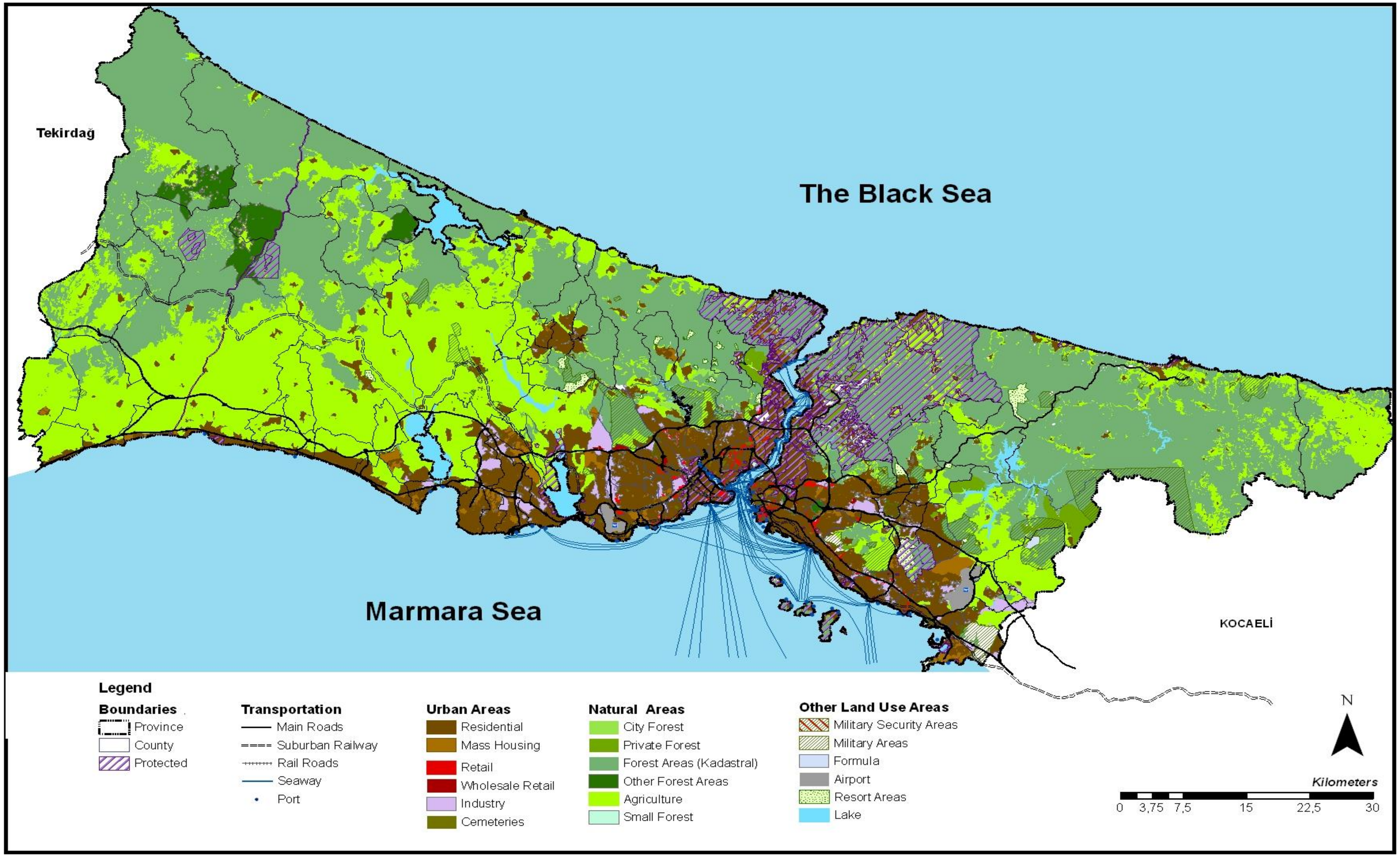


Figure 4.2. Land Use in Istanbul Metropolitan Area
(Source: IBB - Directorate of City Planning, GIS based Land Use Database)

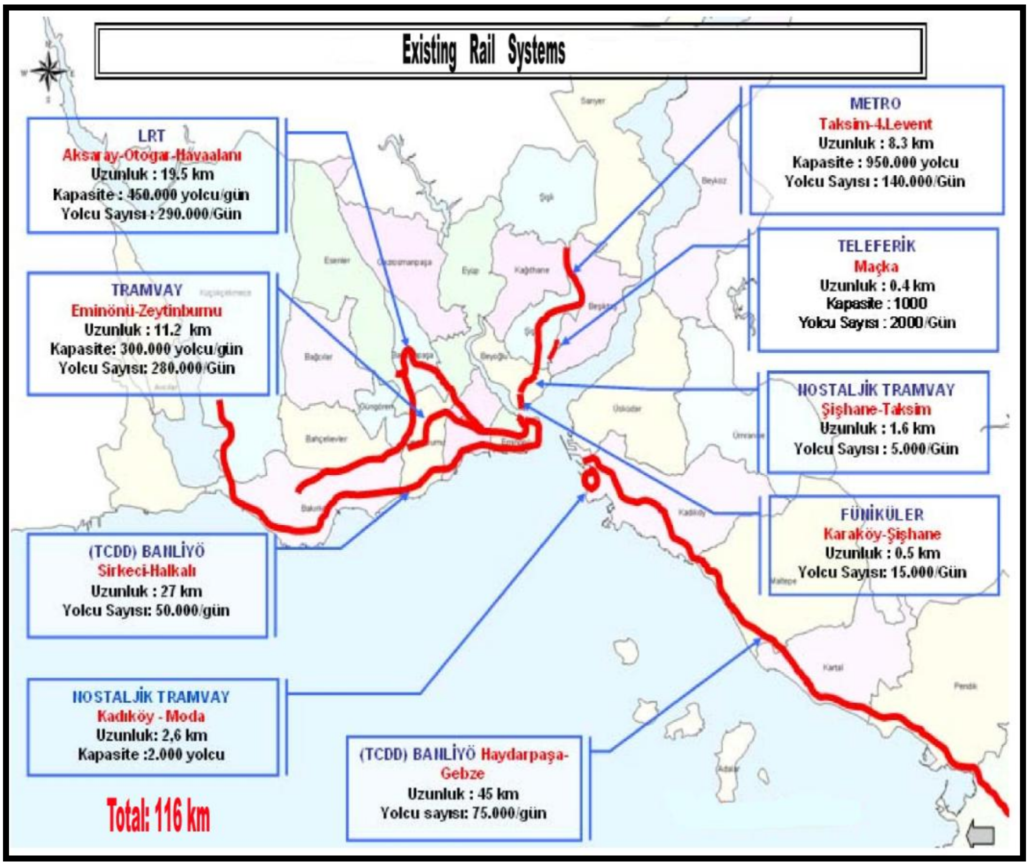


Figure 4.3. Existing rail systems in Istanbul
(Source: adapted from IBB 2005)

Table 4.2. Rail systems and its properties
(Source: IBB 2005)

The Route of Travel	Type	Length (km)	Passenger Carried Daily (person)	Carrying Capacity Daily (person)
Taksim – 4. Levent	Metro	8,3	120.000	950.000
Aksaray-Havaalanı	LRT	19,5	290.000	450.000
Eminönü-Zeytinburnu	Tramvay	11,2	280.000	300.000
İstiklal C. (Tünel-Taksim)	Nostaljik Tramvay	1,6	5.000	6.000
Kadıköy-Moda		2,6	1.700	15.000
Tünel-Karaköy	Füniküler	0,5	13.000	15.000
Sirkeci-Halkalı	Banliyö	27	50.000	250.000
Harem-Gebze	(TCDD)	45	75.000	
Maçka	Teleferik	0,4	1.000	2.000
TOTAL		116,1	835.700	1.988.000

The rail systems in Istanbul are about 116 kilometers in length. In comparison with the cities in Europe, this length is rather limited. For example, Athina has 55 kilometers in length with 51 stations whereas total length of the rail system is 215.5 kilometers in Paris. Madrid metro is one of the longest metro network in Europe with 284 kilometers and 283 stations[17]. Sea transportation is important for inner-city travel. Sea transportation is supported by private and public vehicles. Seabus, fast ferry, and motorboats have been used for sea transportation. Istanbul has a strategic position in air transportation. There are two international airports: Atatürk Airport on European side and Sabiha Gökçen Airport on Asian side.

The data used for the empirical analysis is the 2006 Household Travel Survey conducted by the Transportation Department of the Metropolitan Municipality of Istanbul. 2006 Household survey was used for 2007 Transportation Master Plan in Istanbul. In Istanbul, three transportation master plan and model calibration studies have been prepared up till now for the years 1987, 1997, and 2007. 2007 Istanbul Transportation Master Plan includes the boundaries of the metropolitan area (urban and rural areas). The plan includes 539.000 hectares. Transportation master plan in 2007 has the survey with highest sampling rate. At the beginning of the study, 80% response rate was aimed. 90.000 households were considered for the survey due to budget limitations and previous experiences. In order to make realize, sampling rate was estimated at about 2.2%. At the end of the study, 263.768 people in 70.888 households (as response) participate in this survey, resulting in a database of 356.000 trips in total. 451 travel analysis zones were determined as seen in the Figure 4.6. These travel zones consist of 33 districts (32 districts in Istanbul and 1 district in Gebze, the province of Kocaeli). 2006 Household Travel Survey was randomly made with people who live in 451 travel analysis zones (OD HH 2006). Table 4.3 represents the all trips by different travel modes in Istanbul. 32.3% of the total trips is home - based work (HBW); 21.4% is home - based school (HBS); 37.2% is home - based other (HBO); and 9.1% is non-home based (NHB) trips (Appendix B). The share of private modes is 29% while the share of public transportation is 71% in Istanbul. According to the 2006 Household Travel Survey, the leading transportation mode is walking (49.28%). Private mode usage is only around 14.57%, and public transit is around 35.73%. Service usage is around 11%.

[17] The detailed information for the rail systems in European cities can be found in www.UrbanRail.Net.

The share of rail transit is only about 2.3%. In comparison with European cities, this rate is rather at low levels. The usage of sea transportation is lower than rail systems with 1%. Modal split by travel modes is displayed in Figure 4.4 and Figure 4.5.

Table 4.3. Trip distribution according to trip purpose in 2006
(Source: OD HH 2006)

Travel Modes	**HBW %**	**HBS %**	**HBO %**	**NHB %**	**Total Percent (%)**
Walk	**27,47**	71,09	60,48	31,64	**49,28**
Drive Alone	**11,75**	0,73	4,99	18,58	**7,19**
Shared Ride	**6,00**	2,03	7,27	12,44	**5,76**
Taxi	**1,06**	0,28	2,31	2,67	**1,35**
Service	**19,22**	11,54	1,16	5,22	**10,73**
Dolmuş	**1,39**	0,45	1,04	1,16	**1,03**
Minibus	**10,70**	4,14	9,07	7,81	**8,35**
Public Bus	**14,04**	6,38	8,40	9,52	**10,05**
Private Bus	**2,48**	1,26	2,08	1,89	**2,01**
Motorcycle	**0,27**	0,02	0,11	0,36	**0,16**
Bicycle	**0,08**	0,01	0,07	0,03	**0,05**
Metro (Taksim - 4.Levent)	**0,84**	0,33	0,42	1,08	**0,59**
LRT (Aksaray - Airport)	**0,81**	0,29	0,46	0,68	**0,56**
Tram	**0,99**	0,54	0,56	1,38	**0,76**
Tünel	**0,03**	0,01	0,02	0,08	**0,02**
Ferry	**1,08**	0,46	0,60	1,21	**0,78**
Sea Bus	**0,12**	0,07	0,06	0,13	**0,09**
Sea Motor	**0,18**	0,07	0,11	0,20	**0,13**
Suburb Train	**0,56**	0,18	0,29	0,35	**0,37**
Other	**0,91**	0,13	0,53	3,64	**0,75**
Total	**100**	**100**	**100**	**100**	**100**

Table 4.4 and Table 4.5 represent the rates in modal split and trip purpose throughout the years. According to Table 4.4, the share of private car, service, and rail systems have increased. On the contrary, the usage for taxi, dolmuş, bus and sea transportation has decreased. Regarding modal split by the years in Table 4.5, as mentioned before, the biggest share is home - based other trips. From 1996 to 2007, home-based other trips increased approximately 19%. Home - based work trips were 53% in 1987. Then, hbw trips increased about 2% in 1997. HBW trips decreased to the lowest level (32.3%) in 10 years (OD HH 2006).

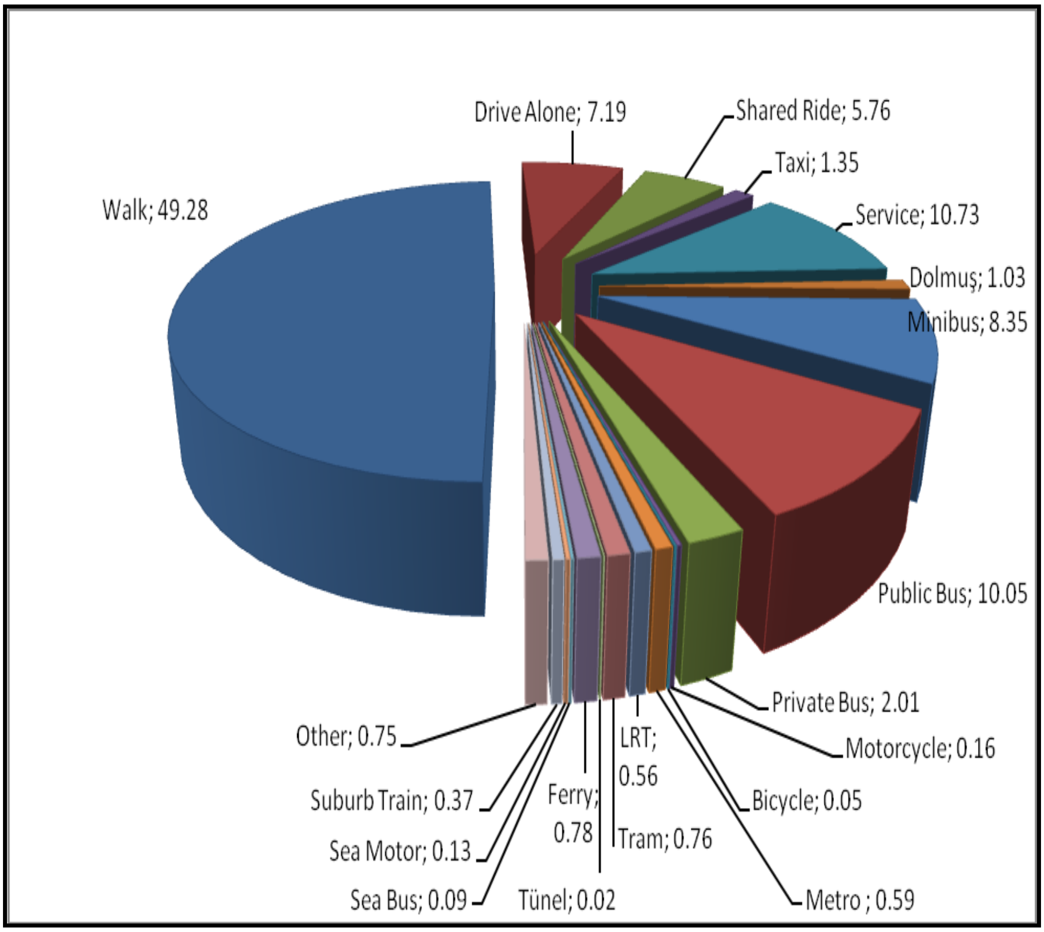


Figure 4.4. Modal split by travel modes in Istanbul

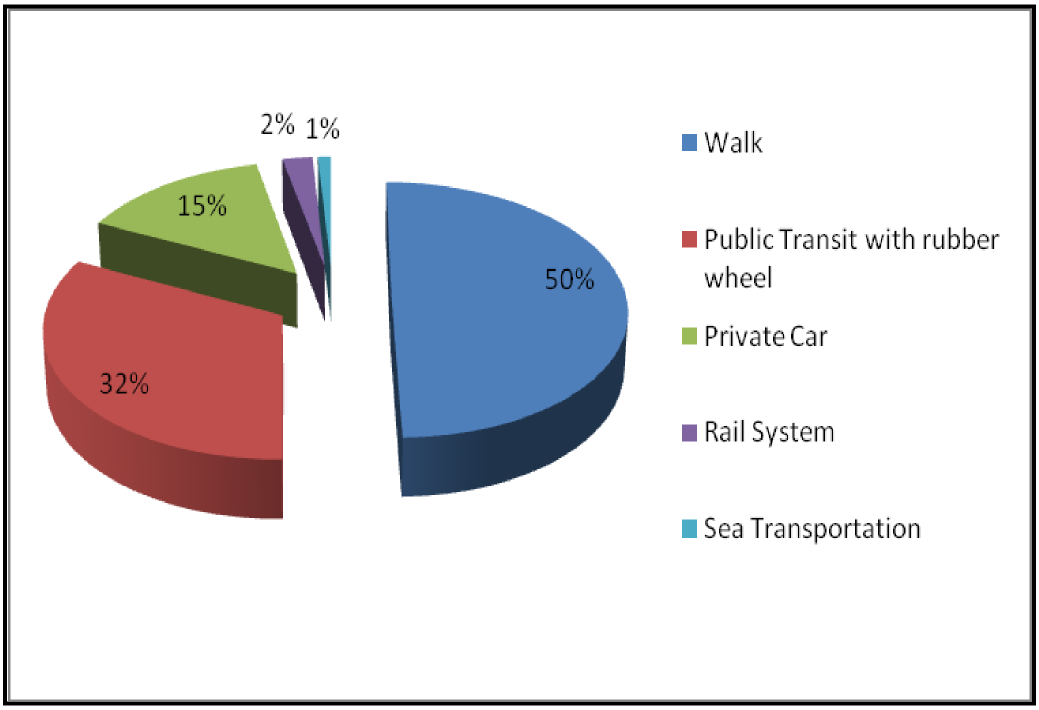


Figure 4.5. Modal split by main travel modes in Istanbul

Table 4.4. Modal split for motorized trips by the years in Istanbul
(Source: OD HH 2006)

Travel Modes	1987 (%)	1996 (%)	2007 (%)
Private Car	19,3	19,2	26,34
Taxi + Dolmuş	10,2	9,4	4,75
Service Vehicles	10,4	11,5	21,48
Bus	35,2	34,1	24,12
Minibus	19	19,6	16,71
Rail Systems	3,8	3,6	4,6
Sea	2,1	2,6	2

Table 4.5. Trip purpose distribution throughout the years in Istanbul
(Source: OD HH 2006)

Trip Purpose	1987 (%)	1997 (%)	2007 (%)
HBW	53	55	32.3
HBS	16	14.5	21.4
HBO	19	18.3	37.2
NHB	12	12.2	9.1
TOTAL	**100**	**100**	**100**

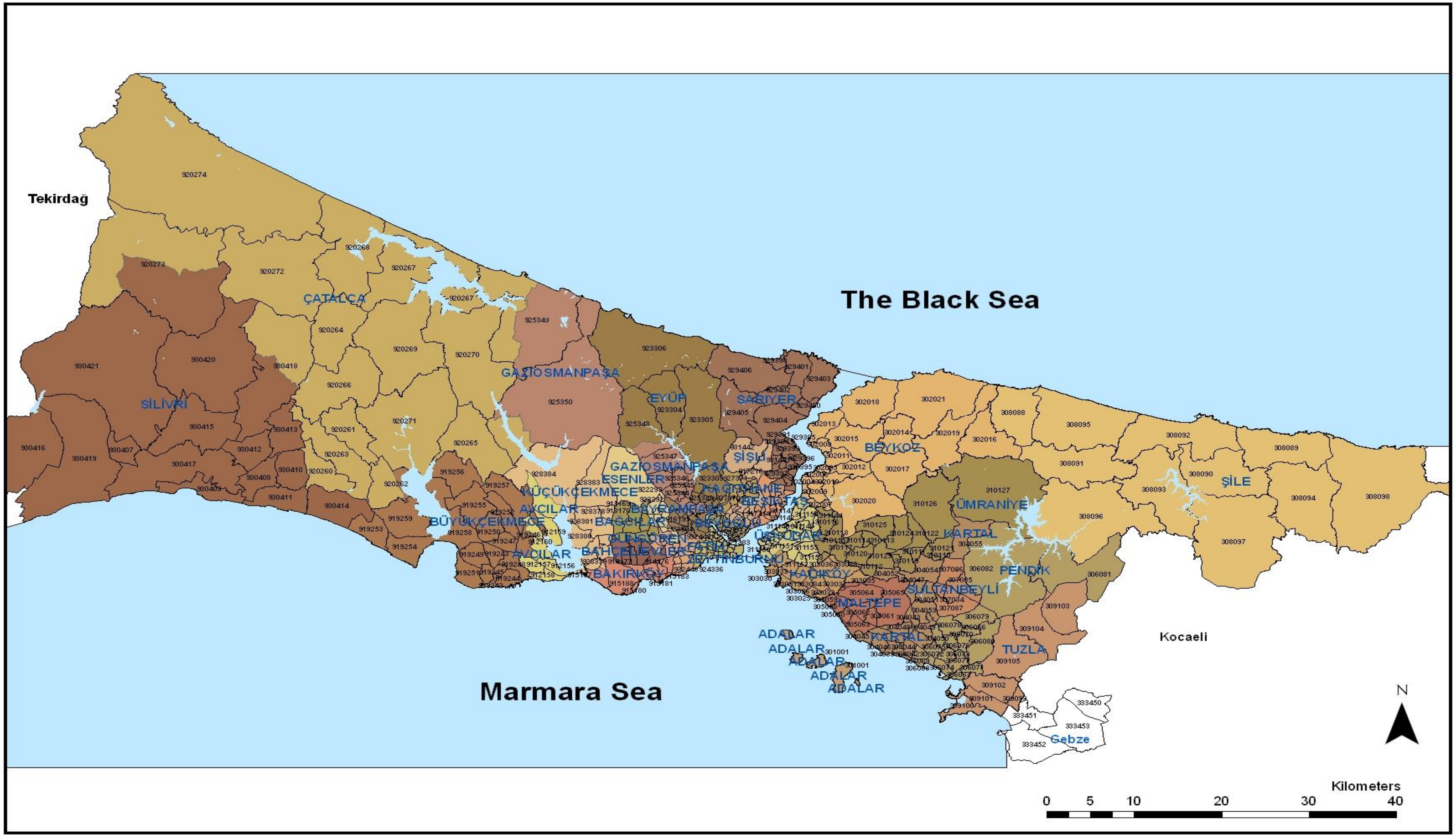


Figure 4.6. Travel Analysis Zones in Istanbul

(Source: IBB - Directorate of City Planning, GIS based Land Use Database)

According to OD HH 2006, average travel time for non - motorized trips (walk and bicycle) is 32 minutes while average travel time for motorized trips is 49 minutes. For HBW trips, average travel times are 41,92 minutes for non-motorized trips and 51,95 minutes for motorized trips. Table 4.6 represents average travel times by the years. In the last two decades, travel times firstly decreased in 1996 and then increased in 2006. The main reason for the reduction in travel time is that a new bosphorus bridge and rail systems were introduced into transportation system after 1987.

Table 4.6. Average travel times by the years (minutes)
(Source: OD HH 2006)

	Motorized Trips			**Non-Motorized Trips**		
Trip Purpose	**1987**	**1996**	**2006**	**1987**	**1996**	**2006**
HBW	55,6	43,0	52,0	45,4	37,9	41,9
HBS	50,9	37,4	48,5	28,8	26,2	23,3
HBO	51,2	41,9	49,8	36,5	34,4	27,8
NHB	44,6	34,0	52,0	35,0	31,3	36,5
Total	**52,8**	**40,7**	**48,9**	**38,0**	**34,3**	**32,2**

Table 4.7. Mobility rates by trip purposes
(Source: OD HH 2006)

Trip Purpose	**Gross Mobility Rates**	**Net Mobility**
HBW Trips	0.56	1.94
HBS Trips	0.37	2.02
HBO Trips	0.58	2.17
NHB Trips	0.12	1.64
Total Trips	**1.64**	**2.40**

For total trips, gross mobility rate is 1,64 while net mobility rate is 2,40. For HBW trips, gross mobility rate is 0,56 while net mobility rate is 1,94 (OD HH 2006)[18]

[18] These rates represents mobility rates that were estimated by the models. Survey rates are close to these rates.

as seen in Table 4.7. As seen in Table 4.7, mobility rates for home - based other (HBO) trips takes the largest percentage (58%) while home - based work and home - based school contributes 56% and 77%, respectively. Table 4.8 and Figure 4.7 represent the distribution of start and end times for HBW trips in Istanbul.

Table 4.8. The Distribution of start and end times for HBW trips
(Source: adapted from OD HH 2006)

Time Period	Start	End
24:00 – 01:00	0,12	0,58
01:00 – 02:00	0,17	0,26
02:00 – 03:00	0,10	0,12
03:00 – 04:00	0,10	0,09
04:00 – 05:00	0,15	0,12
05:00 – 06:00	0,60	0,32
06:00 – 07:00	6,15	1,95
07:00 – 08:00	18,38	10,53
08:00 – 09:00	17,12	23,63
09:00 – 10:00	4,42	8,79
10:00 – 11:00	1,56	2,40
11:00 – 12:00	0,97	1,13
12:00 – 13:00	1,37	1,39
13:00 – 14:00	0,90	0,93
14:00 – 15:00	0,88	0,83
15:00 – 16:00	1,53	1,16
16:00 – 17:00	2,62	1,92
17:00 – 18:00	7,23	4,30
18:00 – 19:00	13,05	9,35
19:00 – 20.00	11,69	13,43
20:00 – 21:00	5,53	8,86
21:00 – 22:00	2,65	4,30
22:00 – 23:00	1,60	2,16
23:00 – 24:00	1,11	1,44

Peak hour for start time of hbw trips is between 07:00 and 08:00 am in the morning and 18:00 and 19:00 pm in the evening. Peak hour for end time of hbw trips is between 08:00 - 09:00 am and 19:00 - 20:00 pm in the evening. According to Figure 4.7, hbw trips have two peaks in the morning and evening.

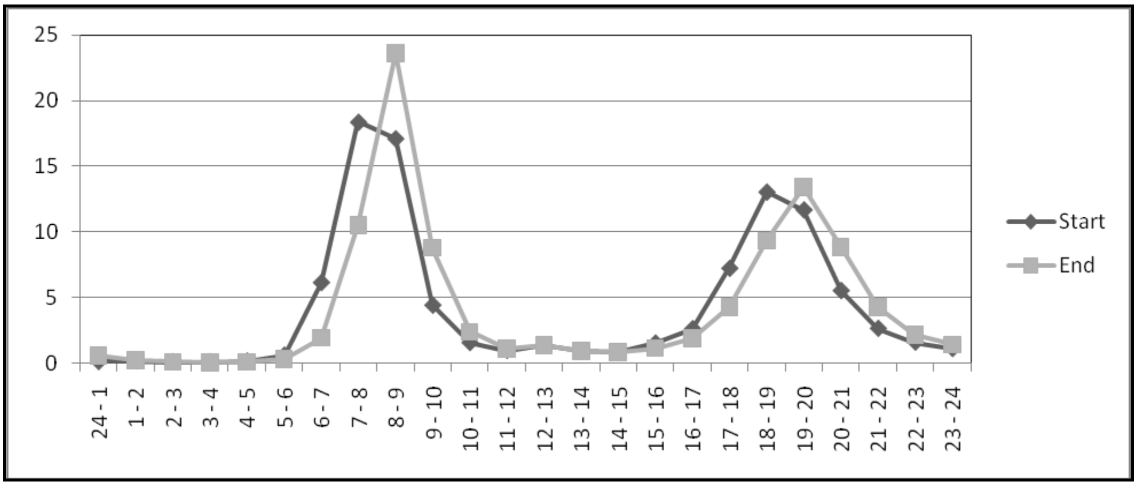


Figure 4.7. Start and end times for HBW trips

In the survey, 451 travel analysis zones are determined as seen in the Figure 4.6. Household travel survey is randomly conducted with the households in which live in 451 travel analysis zones (TAZs) throughout the metropolitan area. The travel data includes information on the socioeconomic, travel, and self - reported distance and time for relative travel. Weekend trips were not included in the data. Some land use variables were obtained using Geographic Information systems. Land use data is at zonal level. Land use data was obtained by Metropolitan Municipality of Istanbul.

In the content of the study, at aggregate level, hbw trips are analyzed for 406 TAZs due to available data. Also, 45 zones are not included into the study contain natural and military areas. Home - based work trips in these zones are low levels. Therefore, 406 of 451 TAZs are selected as the study area. At disaggregate level, the models are calibrated for 451 travel analysis zones (origin - destination pairs). Table 4.9 represents the properties of travel analysis zones in Istanbul. Fatih and Üsküdar are the provinces that include the most travel zones. Adalar only includes one zone. Çatalca is the biggest district in Istanbul. The total population in the case study is 12,006,999 people. The most crowded district is Gaziosmanpaşa whereas the adalar is the smallest district in Istanbul. The lowest population density is in Şile whereas the highest population density is Güngören. The average household size does not vary substantially. The average household size ranges from 2.52 to 4.44. The total number of household varies substantially. Gaziosmanpaşa has the biggest value for the total number of household in Istanbul. The lowest value for total number of household is in Adalar.

Table 4.9. The properties of the districts in 2006 Household Travel Survey (Source: adapted from OD HH 2006)

Counties (Districts)	**The Number of Zones**	**Total Zone Area**	**The Total Number of Household**	**The Average Household Size**	**Total Population**	**Population Density (person/ha.)**
Adalar	1	1098,71	6591	2,52	16592	15,10
Beykoz	20	31444,72	64559	3,56	230628	7,33
Kadıköy	17	3824,41	205543	2,91	597906	156,34
Kartal	17	7767,89	135419	3,59	473429	60,95
Maltepe	10	5186,78	121707	3,45	400851	77,28
Pendik	17	19848,12	142948	3,63	508386	25,61
Sultanbeyli	5	2884,09	53548	4,44	239231	82,95
Şile	11	79037,04	10860	3,29	35180	0,45
Tuzla	7	12442,65	37682	3,65	133733	10,75
Ümraniye	22	21571,73	210470	3,92	800737	37,12
Üsküdar	28	4062,15	180139	3,21	585087	144,03
Avcılar	5	2850,24	80124	3,65	283114	99,33
Bağcılar	10	2175,63	174694	4,10	721073	331,43
Bahçelievler	7	1655,40	159252	3,63	574070	346,79
Bakırköy	11	2983,37	61575	2,86	174658	58,54
Bayrampaşa	8	954,12	70013	3,65	255150	267,42
Beşiktaş	20	1782,15	70979	2,67	179299	100,61
Beyoğlu	24	895,68	64881	3,53	226664	253,06
Büyükçekmece	18	20422,30	158716	3,62	576045	28,30
Çatalca	15	133563,55	22758	3,69	82035	0,61
Eminönü	9	506,55	13279	3,49	45158	89,15
Esenler	10	4382,20	113182	4,03	462306	105,50
Eyüp	13	20352,19	71434	3,72	261203	12,83
Fatih	30	1080,25	115766	3,21	369133	341,71
Gaziosmanpaşa	14	35280,27	250033	3,94	997398	28,27
Güngören	6	720,08	79844	3,78	296145	411,27
Kâğıthane	15	1560,12	105549	3,59	374890	240,30
Küçükçekmece	12	12708,80	200849	3,73	742568	58,43
Sarıyer	22	15137,30	81464	3,45	274742	18,15
Silivri	15	85668,16	37239	3,33	123230	1,44
Şişli	21	3443,13	93514	3,05	277879	80,71
Zeytinburnu	7	1129,41	77463	3,70	287821	254,84
Gebze	4	8333,56	107649	3,89	400658	48,08
Total	**451**	**546752,75**	**3379719**	**3,55**	**12.006.99**	**21,96**

4.2. Descriptive Statistics of Empirical Data

Before estimating empirical models, descriptive statistics are estimated for whole data. The results of descriptive statistics for aggregate and disaggregate data are presented in Table 4.10 and 4.11. Firstly, the result of descriptive statistics at aggregate level is discussed. Then, descriptive statistics for disaggregate level is presented.

Table 4.10. Descriptive statistics for aggregate mode choice data

Variables	Minimum	Maximum	Mean	Std. Deviation	Variance
Zonal Area (area) (ha / 100)	0.1	322.38	10.51	29.43	866.24
Worker (wrkr)	0.2	0.49	0.323	0.04	0.002
Car Ownership per 1000 people (ncar)	11	589	122.04	71.46	5106.42
House Owner (hownr)	0	0.96	0.59	0.11	0.01
Household Size (hhsize)	1.77	4.95	3.474	0.54	0.29
Household Income (TL/1000) (hhinc)	0.41	4.431	1.084	0.42	0.18
Employment / Population (epdens)	0.02	8.62	0.4915	0.82	0.67
Population Density (person / hectare) (pdens)	0.168	868.29	186.50	177.88	31642
Job - Housing Balance (jhb)	0.056	0.99	0.53	0.23	0.05
Land Use Mix (lumix)	0.00006	0.73	0.31	0.14	0.021
Industrial Employment Density (iedens)	0	155.52	13.17	20.84	434.30
Commercial Employment Density (cedens)	0.0026	80.31	8.16	10.89	118.63
Commercial & Ind. Area Density (cidens)	0	0.79	0.14	0.15	0.02
Transit Accessibility (tracc)	0	1	0.27	0.45	0.19

The total number of traffic analysis zones is 406 at aggregate level. The change between minimum statistic and maximum statistic for many variables differ substantially. For example, the size of zonal area ranges from 10.09 ha to 32238 ha. The average size of traffic analysis zones in Istanbul is 10.50 ha. Household income ranges from monthly 410 TL to 4431 TL. The difference in household income among the zones is rather high. Average household size is about 3.5 people. The difference among households in the zones related to house ownership and the number of cars per 1000 person is also high.

Regarding the land use characteristics, population density (person / hectare) ranges from 0.16 to 868.286. Jobs - housing balance as a measure of land use diversity increases about 1 in Istanbul. The lowest level in jobs - housing balance ratio is almost 0.056. A measure of the other diversity index is land use mix diversity index. This index ranges from almost 0 to 0.725. Three employment densities are used: industrial employment density, commercial employment density, and commercial & industrial area density. In comparison to commercial employment density, the change interval for industrial employment density is higher. This suggests that industrial employment in the zones is more dominant than commercial employment. Some zones may not include industrial firms due to the high share of natural and residential areas whereas commercial employment is available in all selected zones. Commercial and industrial area density presents the spatial size in total (commercial as wholesale, retail, and industrial) per zonal area. The zones with the lowest level of this density include greatly natural, forest, green, and military areas. In sum, descriptive statistics suggest that the differences among households as socioeconomic and land use characteristics are attractive at aggregate level.

The empirical application of disaggregate models includes four - alternative mode choice model. The models, MNL and BBNs, aim to predict a commuter choice of travel mode. After the elimination of missing and correlated variables, the empirical data includes a sample of 116992 home - based work trips in total. The sample frequencies of the chosen mode in full data are as follows:

1. Walk Travel: 29.11% (34061),
2. Transit Travel: 30.90% (36156),

3. Car Travel: 19.59% (22924),
4. Service Travel: 20.39% (23851).

Average values of some explanatory variables are as follows:

1. Travel time (minutes):
   A. Walk Travel: 126.19
   B. Transit Travel: 40.85
   C. Car Travel: 18.11
   D. Service Travel: 23.63
2. Travel monetary cost (Turkish Lira):
   A. Transit Travel: 2.15
   B. Car Travel: 6.03
   C. Service Travel: 2.15
3. Household Income: 1250.17 (Turkish Lira)
4. Number of cars available to the household: 0.36
5. Number of company cars available to the household: 0.045
6. Travel distance (kilometers): 8.87

Descriptive statistics for disaggregate data are presented in Table 4.11. The change interval between minimum and maximum is rather high at disaggregate data. Income level ranges from 100 T.L. to 10000 T.L. among the people who live in Istanbul. Average household size is 4,12 people. On average, the number of automobiles is 0,36 whereas the number of company car on average is ,on average, lower than auto ownership. Travel distance ranges from minimum 0,27 km to maximum 135,77 km. Regarding the generic variables, hbw trips undertaken by car have lower values of time than other travel modes. On the other hand, hbw trips by service have lower value of cost than other travel modes[19]. In relation to the land use characteristics of the commuters in the sample, the change intervals vary substantially. For example,

[19] Travel cost for walking mode is not estimated.

population density (population / hectare) at origin ranges from minimum 0,15 to maximum 868,29 while population density at destination ranges from minimum 0,014 to maximum 868,29. The index for jobs - housing balance in general ranges from 0.06 to 256.10. On average, this index is estimated as 5,43 at the origins. Descriptive statistics for training and testing sets at both levels are presented in Table 4.12 - 4.15.

Table 4.11. Descriptive statistics for disaggregate mode choice data

Variables	Minimum	Maximum	Mean	Std. Deviation
Individual Income (TL)	100	10000	1250.17	956.58
Household Size (person)	1	23	4.12	1.75
The Number of Auto in HH	0	3	0.36	0.55
The Number of Company Car in HH	0	2	0.045	0,22
Travel Distance (kilometers)	0.27	135,77	8,87	8,98
Travel Time for Walk (minutes)	4.03	2036.55	126.19	137.26
Travel Time for Auto (minutes)	0.33	197.37	18.11	20.57
Travel Time for Service (minutes)	0.49	254.61	23.62	26.42
Travel Time for Transit (minutes)	5.86	352.46	40.85	32.68
Travel Cost for Service (TL)	1.97	4.75	2.15	0.18
Travel Cost for Transit (TL)	1.00	13.97	2.15	1.51
Travel Cost for Auto (TL)	0.34	82.85	6.03	5.55
Emp. / Worker at Origin (JHB)	0.06	256.10	5.43	28.78
Population Density at Origin	0.15	868.29	229.19	202.07
Emp. / Worker at Destination	0.066	256.10	5.60	29.25
Population Density at Destination	0.014	868.29	227.53	201.44

Table 4.12. Descriptive statistics of disaggregate data in the train set (93594)

VARIABLE	MINIMUM	MAXIMUM	MEAN	STD_DEV
Income	100	10000	1249,80	951,80
Household Size	1	23	4,12	1,75
The number of auto in HH	0	3	0,36	0,55
The number of auto in HH	0	2	0,045	0,22
DIST_KM	0,31	135,77	8,86	8,98
YY_TIME	4,03	2036,55	126,01	137,25
OTO_TIME	0,33	175,28	18,10	20,58
SRVS_TIME	0,49	226,11	23,62	26,43
TRNST_TIME	5,86	352,46	40,83	32,69
PR_SR_C	1,97	4,72	2,15	0,19
FARE	1	13,97	2,14	1,51
OTO_REVISE	0,34	82,85	6,02	5,55
Emp. / Wor. (O)	0,0657	256,1035	5,45	28,88
Popdens (O)	0,1523	868,2864	228,52	201,65
Emp. / Wor. (D)	0,0657	256,1035	5,57	29,15
Popdens (D)	0,0145	868,2864	227,40	201,30

Table 4.13. Descriptive statistics of disaggregate data in the test set (23398)

Variable	MINIMUM	MAXIMUM	MEAN	STD_DEV
Income	100	10000	1251,65	975,47
Household Size	1	20	4,14	1,75
The number of auto in HH	0	3	0,36	0,55
The number of auto in HH	0	2	0,047	0,22
DIST_KM	0,27	114,66	8,92	8,99
YY_TIME	4,03	1719,90	126,96	137,30
OTO_TIME	0,33	197,37	18,15	20,56
SRVS_TIME	0,52	254,61	23,67	26,40
TRNST_TIME	5,86	334,62	40,93	32,64
PR_SR_C	1,97	4,31	2,15	0,18
FARE	1	13,37	2,16	1,52
OTO_REVISE	0,38	76,86	6,07	5,58
Emp. / Wor. (O)	0,066	256,10	5,34	28,36
Popdens (O)	0,1523	868,28	231,84	203,69
Emp. / Wor. (D)	0,0657	256,10	5,73	29,65
Popdens (D)	0,0145	868,29	228,07	201,97

In order to make performance comparisons, aggregate and disaggregate data are divided into two sub - sets. To avoid any empirical bias, descriptive statistics in both data set are kept close to each other as seen in Table 4.12 - 4.15.

The whole empirical data at disaggregate level includes a sample of 116992 home - based work trips. The training data set that are selected randomly from the whole data set includes 93594 records. The remaining 23398 records are used as the testing data set to compare the predictive ability of the empirical models. The whole empirical data at aggregate level includes the socioeconomic and land use characteristics for 406 travel analysis zones. The training data set that is selected randomly from the whole data set includes 325 zonal records. The remaining 81 zonal records are used as the testing data set to compare the predictive ability of the empirical models. The values for training and testing data sets are close to each other at disaggregate level in comparison with aggregate data.

Table 4.14. Descriptive statistics of aggregate data in the train set (325)

VARIABLES	MINIMUM	MAXIMUM	MEAN	STD. DEV.
Zonal Area	0.13	322,38	10,41	31.01
Worker	0.2	0.49	0,32	0.042
Car Ownership	13	589	119,93	73,761
House Owner	0	0.96	0.58	0.11
Household Size	1,8	4,9	3,484	0.54
Household Income	0.41	2,78	1.07	0.39
Employment / Population	0.02	7.91	0.47	0.7
Population Density	0.17	868,29	193,12	179,89
Job - Housing Balance	0.056	0.996	0.531	0.225
Land Use Mix	0	0.725	0.31	0.147
Industrial Employment Density	0	156	14,04	21,753
Commercial Employment Density	0.0025	80,31	8.63	11.5
Commercial & Ind. Area Density	0	0.79	0.14	0.15

Table 4.15. Descriptive statistics of aggregate data in the test set (81)

VARIABLES	MINIMUM	MAXIMUM	MEAN	STD. DEV.
Zonal Area	0.1	111,55	10.9	22.14
Worker	0.23	0.4	0.328	0.04
Car Ownership	11	300	130,48	61,045
House Owner	0.35	0.88	0.6	0.097
Household Size	2,37	4,95	3.44	0.54
Household Income	0.63	4.431	1.16	0.53
Employment / Population	0.075	8.62	0.59	1.18
Population Density	0.481	866,8	1.6	168,03
Jobs - Housing Balance	0.16	1	0.55	0.24
Land Use Mix	0.012	0.592	0.32	0.138
Industrial Employment Density	0	71,99	9.68	16,33
Commercial Employment Density	0.005	31,22	6.3	7.57
Commercial & Ind. Area Density	0	0,53	0.13	0.13

The softwares used in the estimation of discrete choice model require the data to be structured in a way of trip alternative format (Koppelman and Bhat 2006). In this format, each individual are represented by the number of rows that is equal to the number of alternatives within that choice set. For this case, each individual person is represented by four rows of data. The data structure for discrete choice model is shown in Table 4.16.

Table 4.16. General choice data format used in discrete choice models

pid	mode	choice	time	cost	income	age
1	1	0	227.55	0	400	47
1	2	1	67.8	4.76	400	47
1	3	0	58.82	11.04	400	47
1	4	0	75.88	2.28	400	47
2	1	0	152.4	0	1200	33
2	2	1	42.19	3.21	1200	33
2	3	0	53.05	7.66	1200	33
2	4	0	68.43	2.18	1200	33
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.		.
93594	1	0	62.55	0	1500	27
93594	2	0	23.23	1.17	1500	27
93594	3	1	14.34	3.16	1500	27
93594	4	0	18.5	2.05	1500	27

The choice variable as dependent variable, ***choice***, must have one non-zero value for each individual. When every individual have the same choice set, this choice set is a fixed size. In other words, all alternatives are available to all individuals. However, in real situations, a decision maker may not have all alternatives. The number of alternatives can vary across choice set. This choice set is called as *variable number of choices*. In this set, the unavailable alternative is excluded from the choice set. In the case of Istanbul, car and service modes may not be available for all commuters. In order to determine whether an alternative is available or not for each individual, some assumptions should be determined. In the content of the study, choice set is determined as a fixed number of alternatives that all alternatives are available to all individuals.

It is accepted that car and service alternatives may be unavailable to all individuals. In order to determine whether car alternative is available or not for each individual at disaggregate level, data for the choice set with variable number of choices are structured by the assumptions as follows:

1. If there is a car in household, the car is available for the person who is head of household and possesses a driver's license. If not, the lady of the house who possessed of driver license is able to drive. If not, the son of the house possessed of driver's license is able to drive, etc.

2. If there are two cars in household, the cars are available for the head and the lady of the house who possessed of a driver's license.

3. If there are three cars in household, the cars are available for the head, the lady, and the son of the house who possessed of a driver's license.

4. If there is a company car in household, the car is available for the person who is head of household and possesses a driver's license. If not, the lady of the house who possessed of a driver's license is able to drive. If not, the son of the house who possessed of a driver's license is able to drive, etc.

5. If there are two company cars in household, cars are available for the head and the lady of the house who are possessed of a driver's license. If not, the cars are available for the other individuals (the son or the girl of the house) who possessed of a driver's license.

6. If there is no car, car mode is not available for that commuter.

7. Although there is no car in household, the mode choice of commuter may be car mode. Therefore, car mode is available to that commuter.

8. If the mode choice is service, service mode is available to that person. If not, service is not available to commuter.

In addition, if hbw trips by walk mode that exceed 250 minutes, these trips are eliminated from data. It is accepted that hbw trips by walk mode between the two continents (Asia and Europe) are eliminated from data.

CHAPTER 5

EMPIRICAL RESULTS

This chapter focuses on the model results at aggregate and disaggregates levels. The study includes two sets of models for home - based work trips in each level. Each set in logit models includes a base model and an extended model. Significant variables, explanations for the sings, and the findings are presented. Goodness of fit statistics for all models are presented. Next, performance comparisons for logit models and bayesian belief networks at both levels are assessed. Performance comparisons of the models (Logit and BBNs) are made according to the expanded model specification.

5.1. Aggregate Model Results

Baseline category logit model and Bayesian belief networks are estimated at aggregate level. As mentioned previous section, dataset is divided into two subsets: training and testing data. The training data (80% of the case file, n=325 records) is used for estimating model parameters. Then, testing data (20%, n=81 records) is used to compare and test the predictive ability (model performance) for different models (logit and BBNs). Firstly, the results for baseline category logit model are presented. After that, BBNs are introduced. In aggregate level, the models are used to describe commuter’s choices among four alternatives and define the effects of zonal characteristics on mode choice. The choice set is defined as socioeconomic and land use characteristics.

5.1.1. Baseline Category Logit Model Results

In order to test main hypothesis that land use attributes affect mode choice at aggregate level, the analysis is carried out using two different models. Firstly, a base model including only socioeconomic variables is estimated. After that, land use variables are entered into the base model. The expanded model is needed to gauge the marginal influence of land use characteristics. The results of the models are presented in Table 5.1 and Table 5.4.

Table 5.1. The base model for home - based work trips with only SoE variables

Parameter	Mode	Estimate	Std. Error	Chi-Square	Pr > ChiSq
Intercept	car	-3.7098	0.2164	293.8709	<.0001
Intercept	service	-5.9388	0.2234	706.5486	<.0001
Intercept	transit	-1.6921	0.1883	80.7892	<.0001
Area	car	-0.00279	0.000458	37.0342	<.0001
Area	service	-0.00096	0.000383	6.2379	0.0125
Area	transit	-0.00221	0.000372	35.4670	<.0001
House Owner	car	1.5481	0.1452	113.6669	<.0001
House Owner	service	2.3931	0.1424	282.3645	<.0001
House Owner	transit	1.6393	0.1238	175.2523	<.0001
Household Income	car	0.0665	0.0742	0.8040	0.3699
Household Income	service	-0.1254	0.0836	22.470	0.1339
Household Income	transit	0.0152	0.0711	0.0454	0.8313
Car Ownership	car	0.0115	0.000392	857.4960	<.0001
Car Ownership	service	0.00544	0.000435	156.3383	<.0001
Car Ownership	transit	0.00347	0.000368	89.1290	<.0001
Worker	car	-0.0441	0.3879	0.0129	0.9094
Worker	service	3.2361	0.4031	64.4528	<.0001
Worker	transit	1.1372	0.3461	10.7992	0.0010
Household Size	car	0.3170	0.0310	104.2948	<.0001
Household Size	service	0.7256	0.0319	517.8860	<.0001
Household Size	transit	0.0226	0.0269	0.7056	0.4009

In this stage, a base model is calibrated with only socioeconomic data (training data). This data set consists of 325 observations that are selected randomly. Therefore, there are 325 unique combinations of explanatory variables in this data set. Base model model has 21 parameters to be estimated. These combinations present the total number of analysis zones in the case. Walk mode is selected as baseline or reference category. Therefore, the baseline category logit model (or multinomial logit model) with socioeconomic variables becomes $\ln(\pi_{car}/\pi_{walk})$, $\ln(\pi_{service}/\pi_{walk})$, and $\ln(\pi_{transit}/\pi_{walk})$ respectively. Baseline category logit model (base model) investigates the effects of socioeconomic characteristics on mode choice.

The saturated model fits a separate multinomial distribution to each group. In this case, the saturated model has 325 x 3 =975 free parameters. In base model, there are a total of 21 parameters. According to Table 5.2, all variables are significant at the level of 0.001. Likelihood ratio and Pearson chi-square test statistics are used to compare the proposed model which has 21 parameters with the saturated model. These statistics have 975 – 21 = 954 degrees of freedom as shown in Table 5.2 which reveals the proposed model does not fit better than the saturated model.

Table 5.2. Deviance and Pearson Goodness-of-Fit Statistic for the base model

Criterion	Value	DF	Value/DF	Pr>ChiSq
Deviance	7964.4945	954	8.3485	<.0001
Pearson	7890.0667	954	8.2705	<.0001

This model fits the data with a G^2 (Likelihood ratio) = 7964.4945 with a p-value of 0.0001. The Table 5.3 presents Maximum Likelihood (ML) Analysis of Variance derived by Proc Logistic in SAS software. According to the results, variables are highly significant as indicated by p-values in Table 5.3. In other words, there is evidence that model variables affect the choice of travel mode.

Table 5.3. Maximum likelihood analysis of variance for the base model

Source	DF	Chi-Square	Pr>ChiSq
Area	3	54.2227	<.0001
House Owner	3	332.4037	<.0001
Household Income	3	5.7427	0.1248
Car Ownership	3	929.5184	<.0001
Worker	3	79.6221	<.0001
Household Size	3	647.2029	<.0001

When walk mode is selected as the reference category, Table 5.1 presents ML estimates of the parameters. The equations derived from the Table 5.1 determine those for other travel mode comparisons. For instance, the prediction equation for the log odds of selecting car modes instead of transit is written below:

$$\log(\pi_{car} / \pi_{transit}) = \log(\pi_{car} / \pi_{walk}) - \log(\pi_{transit} / \pi_{walk}) \tag{5.1}$$

$$\begin{aligned} &= (-3.7098 - 0.00279 area + 1.5481 hownr + 0.0665 hhinc + 0.0115 ncar \\ &- 0.0441 wrkr + 0.3170 hhsize) \\ &- (-1.6921 - 0.00221 area + 1.6393 hownr + 0.0152 hhinc + 0.00347 ncar \\ &+ 1.1372 wrkr + 0.0226 hhsize) \end{aligned} \tag{5.2}$$

$$\begin{aligned} &= -2.0177 - 0.00058 area - 0.0912 hownr + 0.0513 hhinc + 0.00803 ncar \\ &- 1.1813 wrkr + 0.2944 hhsize \end{aligned} \tag{5.3}$$

Response probabilities for generalized logit models are estimated following expression (Agresti 2002).

$$\pi_j(x) = \frac{\exp(\alpha_j + \beta'_j x)}{1 + \sum_{h=1}^{J-1} \exp(\alpha_h + \beta'_h x)} \tag{5.4}$$

According to the model results as seen in Table 5.1, it may seem that among all socioeconomic predictors, household income for all travel modes, household size for transit mode, and worker for only car mode is not significant. All of the sings of the variables, except working status (worker) for car mode are as expected.

Using Equation 5.4, the estimated response probabilities of the outcomes are presented in the equation as follows:

$$\pi_{car} = \frac{\exp(-3.7098 - 0.00279 + 1.5481 + 0.0665 + 0.0115 - 0.0441 + 0.3170)}{1 + \exp(-3.7098 - .... + 0.3170) + \exp(-5.9388 - ...... + 0.7256) + \exp(-1.6921 - .... + 0.0226)} \tag{5.5}$$

$$\pi_{service} = \frac{\exp(-5.9388 - 0.00096 + 2.3931 - 0.1254 + 0.00544 + 3.2361 + 0.7256)}{1 + \exp(-3.7098 - .... + 0.3170) + \exp(-5.9388 - ...... + 0.7256) + \exp(-1.6921 - .... + 0.0226)} \tag{5.6}$$

$$\pi_{transit} = \frac{\exp(-1.6921 - 0.00221 + 1.6393 + 0.0152 + 0.00347 + 1.1372 + 0.0226)}{1 + \exp(-3.7098 - .... + 0.3170) + \exp(-5.9388 - ...... + 0.7256) + \exp(-1.6921 - .... + 0.0226)} \tag{5.7}$$

$$\pi_{walk} = \frac{1}{1 + \exp(-3.7098 - .... + 0.3170) + \exp(-5.9388 - ...... + 0.7256) + \exp(-1.6921 - .... + 0.0226)} \tag{5.8}$$

The expanded model is estimated for socioeconomic and land use data. As in the previous model, walk mode is selected as a reference category. There are 325 unique combinations of explanatory variables. This number is based on the total number of traffic analysis zones in training data. The results of the expanded model is presented in Table 5.4.

Table 5.4. Expanded model for home - based work trips with full data

Parameter	Mode	Estimate	Std. Error	Chi-Square	Pr > ChiSq
Intercept	car	-2.8718	0.2441	138.4558	<.0001
Intercept	service	-4.9806	0.2539	384.8501	<.0001
Intercept	transit	-0.9447	0.2146	19.3780	<.0001
Area	car	-0.00255	0.000511	24.8153	<.0001
Area	service	-0.00066	0.000431	2.3666	0.1240
Area	transit	-0.00160	0.000415	14.8653	<.0001
House Owner	car	0.9734	0.1497	42.2805	<.0001
House Owner	service	1.6656	0.1479	126.7682	<.0001
House Owner	transit	0.7924	0.1294	37.4857	<.0001
Household Income	car	0.3134	0.0800	15.3326	0.003
Household Income	service	0.2760	0.0884	9.7389	0.0018
Household Income	transit	0.2855	0.0770	213.7454	<.0002
Car Ownership	car	0.0103	0.000425	590.8139	<.0001
Car Ownership	service	0.00332	0.000468	50.3048	<.0001
Car Ownership	transit	0.00248	0.000400	38.3662	<.0001
Worker	car	0.1691	0.4021	0.1768	0.6741
Worker	service	3.1813	0.4177	58.0038	<.0001
Worker	transit	1.6121	0.3606	19.9846	<.0001
Household Size	car	0.2375	0.0363	42.7378	<.0001
Household Size	service	0.6388	0.0371	295.6672	<.0001
Household Size	transit	0.0144	0.0316	0.2066	0.6494
Employment / Pop.	car	-0.1078	0.0310	12.0681	0.0005
Employment / Pop.	service	-0.1989	0.0356	31.2287	<.0001
Employment / Pop.	transit	-0.1865	0.0293	40.5314	<.0001
Population Density	car	-0.00016	0.000075	4.4011	0.0359
Population Density	service	-0.00029	0.000076	14.5231	0.0001
Population Density	transit	-0.00027	0.000065	17.3316	<.0001
Jobs – Housing Balance	car	-0.5728	0.0641	79.8505	<.0001
Jobs – Housing Balance	service	-0.4701	0.0668	49.5392	<.0001
Jobs – Housing Balance	transit	-0.8367	0.0575	211.9997	<.0001
Land Use Mix	car	0.4800	0.0912	27.7189	<.0001
Land Use Mix	service	0.8099	0.0917	78.0911	<.0001
Land Use Mix	transit	0.6637	0.0803	68.3243	<.0001

(cont. on next page)

Table 5.4. (cont.)

Parameter	Mode	Estimate	Std. Error	Chi-Square	Pr > ChiSq
Industrial Emp. Density	car	-0.00003	0.000752	0.0014	0.9702
Industrial Emp. Density	service	-0.00206	0.000787	6.8179	0.0090
Industrial Emp. Density	transit	-0.00260	0.000664	15.3270	0.0001
Commercial Emp. Density	car	-0.00351	0.00190	3.4132	0.0647
Commercial Emp. Density	service	-0.00637	0.00211	9.1418	0.0025
Commercial Emp. Density	transit	0.00291	0.00166	3.1002	0.0783
Com. & Ind. Area Density	car	-0.9749	0.1036	88.5499	<.0001
Com. & Ind. Area Density	service	-1.0175	0.1077	89.3175	<.0001
Com. & Ind. Area Density	transit	-1.2839	0.0923	193.3214	<.0001
TRACCESS	car	-0.0354	0.0263	1.8129	0.1782
TRACCESS	service	-0.0325	0.0276	1.3839	0.2394
TRACCESS	transit	0.0478	0.0232	4.2627	0.0390

Deviance and Pearson goodness of fit statistics test the fit of the model versus saturated model. The current model has 45 parameters whereas the saturated model 930 free parameters. The overall fit statistics displayed in Table 5.5 have 930 degrees of freedom.

Table 5.5. Deviance and Pearson Goodness-of-Fit Statistics

Criterion	Value	DF	Value/DF	Pr > ChiSq
Deviance	6052.9516	930	6.5086	<.0001
Pearson	5923.3220	930	6.3692	<.0001
Number of unique profiles: 325				

The model with only intercept (null model) has been tested against the current model. The null model has three parameters since there are three logit equations. The comparison has $45 - 3 = 42$ degrees of freedom.

Table 5.6. Maximum likelihood analysis of variance

Source	DF	Chi_Square	Pr > Square
AREA	3	30.5923	<.0001
House Owner	3	132.7054	<.0001
Household Income	3	18.7009	<.0003
Car Owner	3	688.0849	<.0001
Worker Rate	3	72.8827	<.0001
Household Size	3	363.6078	<.0001
Employment / Population	3	53.2807	<.0001
Population Density	3	22.7470	<.0001
Jobs – Housing Balance	3	218.4423	<.0001
Land Use Mix	3	101.7436	<.0001
Industrial Emp. Density	3	20.5312	0.0001
Commercial Emp. Density	3	23.0973	<.0001
Com. & Ind. Area Density	3	219.3769	<.0001
TRACCESS	3	14.1827	0.0027

According to Table 5.6, all of the explanatory variables are influential effects. This test suggests that all of the variables should be entered into the model. The analysis of variance table is displayed in Table 5.6. This model fits the data with a G^2 (likelihood ratio) = 6052.9516 with a p-value of 0.0001.

In logit models, different goodness of fit test statistics are used determine how well estimated model fits the data. Pearson chi-square (X^2) and the deviance (G^2) are the most popular statistics among these. In the content of the study, it is tested that expanded model (M_1) including 45 parameters outperforms the model with only socioeconomic data (M_0). The null hypothesis for this case is as follows:

H_0 = The model with 21 parameters fits the data.

H_1 = The model with 45 parameters fits better.

The likelihood ratio test statistic to test the null hypothesis given above is calculated as follows (Agresti, 2002):

$$G^2(M_0 / M_1) = -2(L_0 - L_1) = G^2(M_0) - G^2(M_1) \tag{5.9}$$

The likelihood ratio statistic equals 7964.4945 - 6052.9516 = 1911.5429 suggesting that H_0 should be rejected which means the expanded model is more adequate. This result provides information about which land use characteristics improved model fit. Response probabilities can be estimated using the formulation 5.4.

In other words, with socioeconomic variables, the baseline category logit model gives G^2 = 7964.4945 with p-value = 0.0001 while the model with the full data gives G^2 = 6052.9516 with p-value = 0.0001. Thus, the baseline category logit model with full data fits better.

This result provides information about which land use characteristics improved model fit. In other words, this result proved sub-hypothesis 1 (SH1) that adding land use variables to the models at aggregate level improves the model explanatory power.

As mentioned before, three logit models of car, service, and transit mode choice to walk mode for home - based work trips are estimated in Istanbul. Three logit equations describe the log odds that people who live in traffic analysis zones select travel modes instead of walk mode. Two different data sets are entered into the models. One for socioeconomic data and the other one for extended data. In extended model, land use variables are entered into the model as seen in Table 5.4.

According to maximum likelihood analysis of variance in Table 5.3 and Table 5.6, all variables are highly significant as indicated by p - values. After the inclusion of land use variables, the signs of all predicting variables from the first model, except income for service mode and working status for car mode, did not change. Also, they retain their statistical significance.

Socioeconomic variables exhibit a statistically significant influence on motorized trips. From the table, two of the socioeconomic variables, house ownership and car ownership, are significant and positive for all modes. As expected, the coefficients of income for car mode is positive, showing that commuters who live in high - income zones are more likely to choose motorized alternatives (car and transit). The variable indicating zonal average of working is positive and significant for service and transit modes. Average household size is positively correlated with private modes. One of the possible explanations for this result is that travelling by car for households can be more comfort with children than transit and walk mode. According to the studies of Collins and Kearns (2001), space-time flexibility, safety, and security promote the use of car.

In baseline category logit model, parameters of the models are interpreted in terms of odds ratio. The intercepts provide information about the estimated log-odds for the reference group. The estimated log - odds of car versus walk mode in this group are -2.8718; the estimated log-odds of service versus walk is -4.9806. Regarding odds ratio estimated for income variables, the estimated odds ratios are 1.368 for car, 1.318 for service, and 1.330 for transit. The presence of transit access is characterized by dummy coefficients. The estimated coefficient for transit mode versus walk mode is 0.0478. This means that people who live in the zones with transit availability are more likely to choose transit versus walk mode. In other words, walk mode appear to be less common in the zones for home - based work trips. The estimated odds ratio of the presence of transit access is about 1.05. Car availability has a strong positive influence on the likelihood of choosing the modes versus walk mode. The all of the socioeconomic variables are positively correlated travel modes versus walk mode, as expected.

Several land use variables have a statistically significant effect on mode choice. Jobs - housing balance is a measure of the mix between employment and dwelling units in a specific area. The sign of the jobs - housing balance is negative. In the world, policies promoting jobs - housing balance attempt to locate housing close to jobs. Behind these policies, planners and policy-makers want to decrease traffic congestion and increase accessibility to jobs and affordable housing. They aim to improve the quality of life and protect the environment. In the case of Istanbul, the sign indicates that people is more likely to use non-motorized trips for home - based work trips. Therefore, home based work trips are mainly intrazonal trips. Mixed land use is expected to shorten travel distance. Therefore, it encourages people to use walk or public transport modes. Positive effect of mixed land use versus walk mode indicates that land use mix encourage motorized trips for work trips. This result is the opposite of the results of North American cities whereas it resembles the findings of Asian cities such as Lin and Yang (2009) who studied urban form impacts on travel demand in Taipei, Taiwan. However, this study found that there is a significant relationship between mixed land use and mode choice at aggregate level in the case of Istanbul, in opposition to the study of Zhang (2004) for Hong Kong at disaggregate level. As seen in Table 5.4, density variables are negatively correlated with motorized trips as expected because higher densities encourages non-motorized or public transit trips. The zones becoming employment zone rather than residential encourage non-motorized trips. It can be accepted that home - based work trips are intrazonal. People want to make a shorter car trips and lower travel distance. In empirical studies at disaggregate level, this results shows similarity with the previous studies of North American studies. At aggregate level, the result is the same as the studies of Lin and Long (2008) and Buchanan et al. (2006) who found a negative effect on private mode split at higher density areas. Cervero and Gorham (1995) found that residential density is positively correlated with transit commuting in auto and transit oriented neighborhoods in Los Angeles County. Bhat and Guo (2007) found that households with low income tend to high employment densities in Alameda County in San Francisco Bay area. Increase in household income tends to use motorized trips. Also, in this case, increase in employment densities tends to choose less motorized trips. The findings resemble findings for the study of Bhat and Guo (2007).

At aggregate level, Newman and Kenworthy (1989) studied the relationship between density, mode shares, and vehicle miles traveled (VMT) in 32 major cities in Europe, North-America, Australia, and Asia. They found that increase in density leads to decrease the share of auto mode. There was an inverse relationship between population density and motorized trips. Coevering and Schwanen (2006) extended the studies of Newman and Kenworthy (1989, 1999). They studied the relationship between travel demand and urban form for 31 cities in Europe, Canada, and the USA. They found that higher population densities tended to decrease the share of car trips and increase the share of walking/bicycling modes. In the case of Istanbul, population density has a negative effect on motorized trips (when comparing walk mode) and has a positive effect on non - motorized trips in Istanbul. This result is consistent with the findings of previous studies as mentioned above. Increase in population density tend to increase the share of non - motorized modes (walk and bicycle).

The size of zonal area is negatively correlated with motorized trips versus walk mode. This suggests that higher zonal areas encourage people to live close to the place of employment. It decreases the choice probability of motorized home - based work trips. On the other hand, commuters whose house ownership is low are more likely to use motorized trips. Increase in household income and car ownership tends to use motorized trips. Household size is positively correlated with motorized trips. One of the possible explanations for this result is that higher household size increases the number of workers in household. Therefore, some of workers may prefer car trips. House owner and car ownership have a strong positive influence. Among the socioeconomic and land use attributes, commercial employment density and the presence of transit access are not strongly related to mode choice.

Regarding the sub-hypothesis related to the model variables, increase in land use mix and the presence of transit accessibility in the zones encourage transit usage while increase in population density tends to decrease transit usage versus walk mode. There is no enough evidence of violation of sub-hypotheses H1 and H3 that people who live in high density, mixed use prefer to travel with transit service and walking mode due to the selection of walk mode as a reference category. However, hypothesis H4 that the presence of transit access increases the choice of transit modes is supported. Hypothesis H2 is not supported. Employment densities are negatively correlated with motorized trips. The size of zonal area is negatively correlated with motorized trips. In sum, at aggregate level, land use has an important factor for home - based work trips. Commercial & industrial area density, employment / worker ratio, jobs - housing balance, and land use mix variables are statistically significant for motorized trips versus unmotorized trips.

Figure 5.1 - 5.5 presents the plots of these predicted probabilities against control variables. The baseline category logit model can be used to predict the probability of home - based work trips. Figure 5.1 represents the plot of the predicted probabilities against household income. Only car mode is positively correlated with household income. Increasing in income leads to decrease the choice probabilities of other travel modes. The mode choice probability for car mode ranges from a 10 percent to a 68 percent. However, the probabilities of other modes decrease with increasing household income. For example, the probability of home - based work trips by walk mode falls from 40 percent to 5 percent. At the same time, the probability of hbw trips by transit 27 percent to a 21 percent with a household income about 2800. Commuters with a monthly household income of $410 have an approximately 10 percent chance of commuting by car mode. Commuters with a monthly household income exceeding $3000 have a 21 percent chance of home - based work trips by transit, and an approximately 6 percent chance of home - based work trips by walk mode.

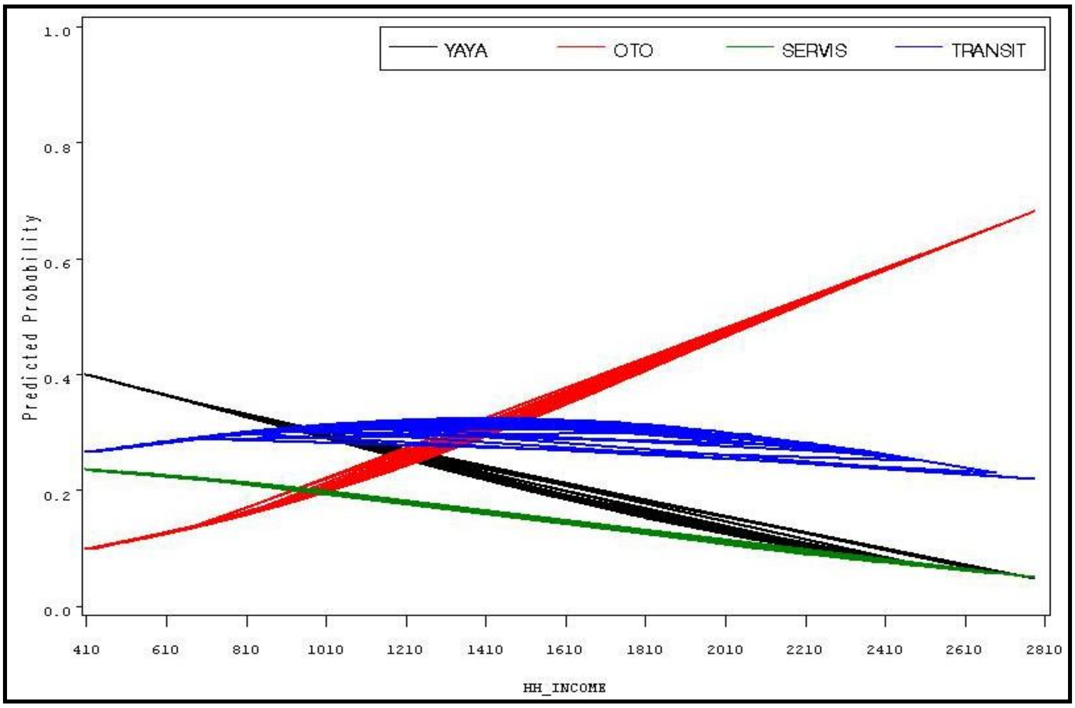


Figure 5.1. Effect of household income on mode choice probability

The Figure 5.2 represents the effects of auto ownership on mode choice probability. Increasing the number of auto ownership in households lead to increase home - based work trips by auto whereas it leads to decrease the choice probability of other travel modes.

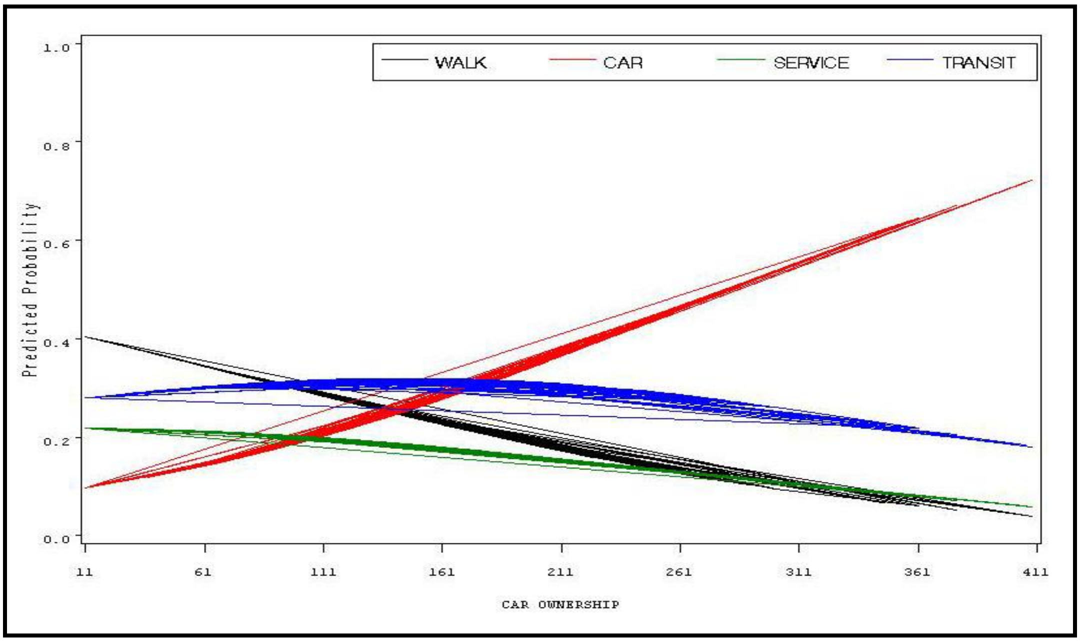


Figure 5.2. Effect of car ownership on mode choice probability

Figure 5.3 represents the change of mode choice probabilities of home - based work trips based the size of zonal area. Increasing the size of zonal area leads to higher demand for unmotorized trips whereas it leads to decrease motorized trips except service mode. The probability of home - based work trips by car falls from 21 percent to 17 percent. The mode choice probability ranges from an approximately 29 percent chance of walk mode to a 33 percent.

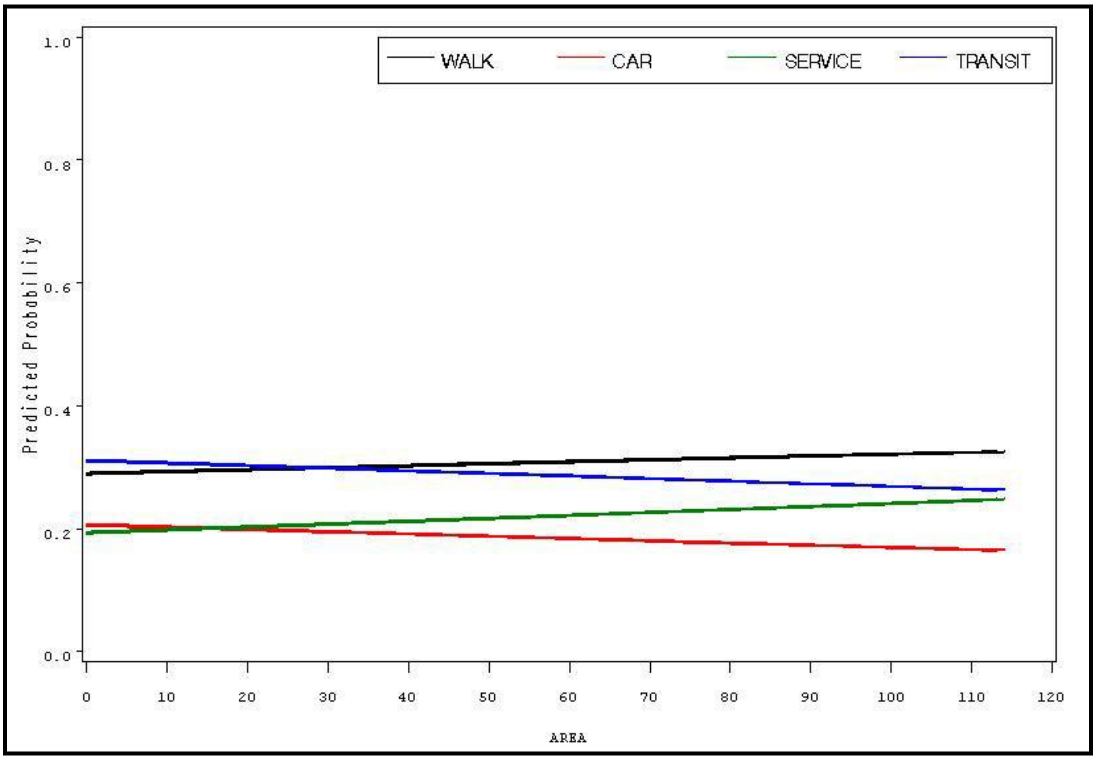


Figure 5.3. Effect of zonal area on mode choice probability

Figure 5.4 represents the change of mode choice probabilities of home - based work trips based population density. Increase in population density leads to higher demand for walk and transit trips. The probability of home - based work trips by car falls from 24 percent to about 22 percent.

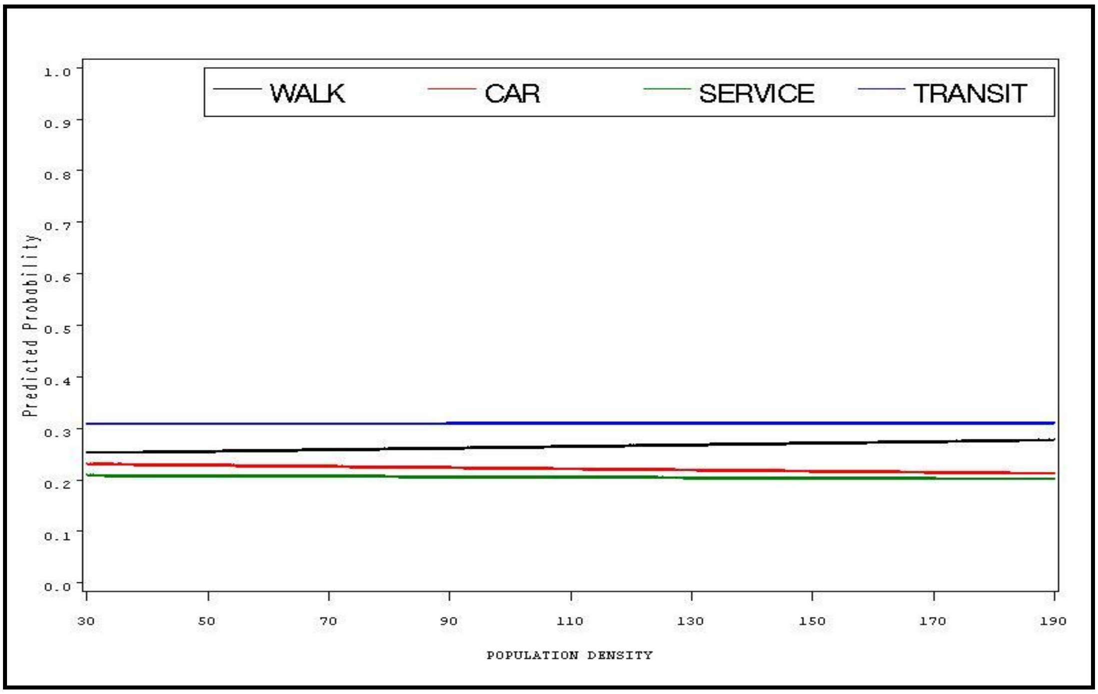


Figure 5.4. Effect of population density on mode choice probability

Figure 5.5 presents the relationship between industrial employment density and mode choice probability. According to this figure, the probability of transit mode falls from 33 percent to about 26 percent whereas the probability of walk mode increases from about 25 percent to 43 percent. At the same time, the probability of home - based work trips by car falls from 23 percent to about 14 percent.

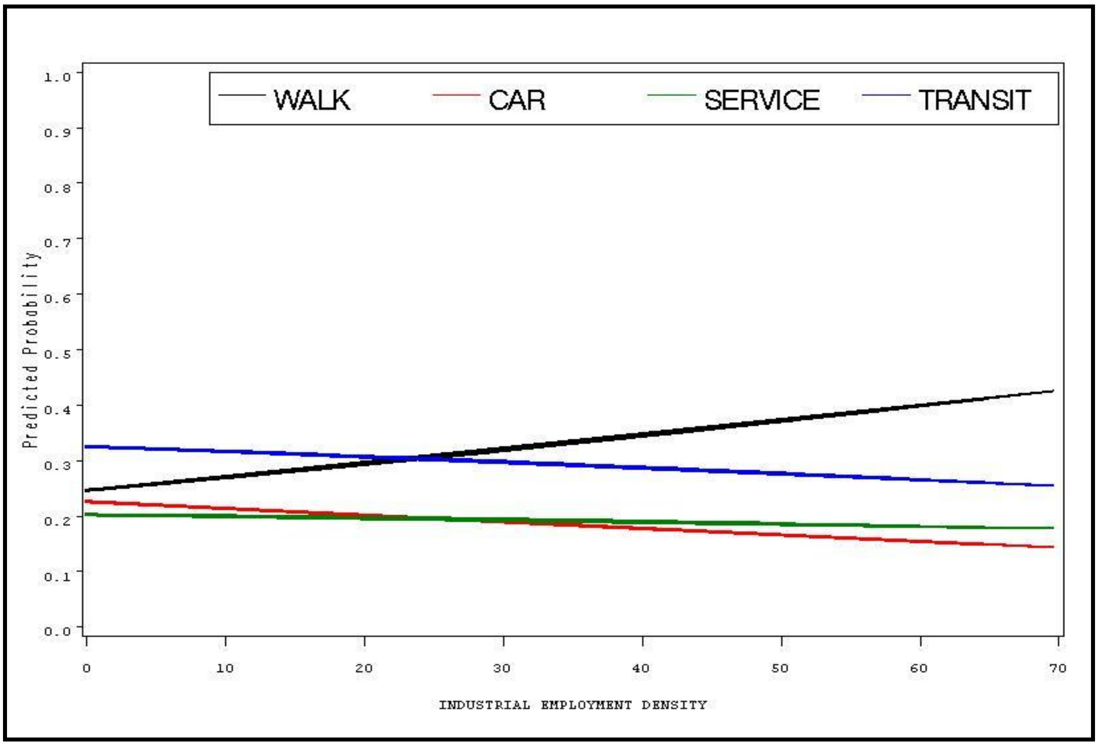


Figure 5.5. Effect of industrial employment density on mode choice probability

5.1.2. Bayesian Belief Network Results at Aggregate Level

Bayesian belief networks at aggregate level are constructed with the variables that are used in baseline category logit model (expanded model). In total, there are 15 nodes, one for query node (Mode Choice). After constructed the structure from aggregate data, structure learning and parametric learning are applied in BN PowerConstructor software. The use of parametric learning from empirical data provide conditional probability tables in the network. The network is compiled in Hugin as seen in Figure 5.6 and 5.7. There are 28 links in the network among the nodes.

The beliefs for each node are shown in Figure 5.6. Sensitivity analysis provide information to determine the impact levels of the nodes on query node. Table 5.7 represents sensitivity of “Mode Choice” due to a finding at another node. The higher mutual info value represents more effective nodes for query node (Mode Choice).

Table 5.7. Sensitivity analysis of mode choice for aggregate analysis

Node	Mutual Info	Variance of Beliefs
NCAR	0.11526	0.0113063
IEDENS	0.07837	0.0130884
HHINC	0.04018	0.0027146
HHSIZE	0.03386	0.0028726
AREA	0.03232	0.0026796
PDENS	0.02075	0.0017143
CIDENS	0.00737	0.0013939
CEDENS	0.00665	0.0012409
JHB	0.0066	0.0011563
LUMIX	0.00651	0.0005384
WRKR	0.00502	0.0003697
EPDENS	0.00139	0.0001511
TRACC	0.00136	0.0000961
HOWNR	0.00042	0.0000488

According to the sensitivity analysis, the number of car in household (ncar) is the most influential node on the query node. Industrial employment density (iedens), average household income (hhinc), household size (hhsize), zonal area (area), and population density (pdens) are the other influential nodes sequentially for mode choice.

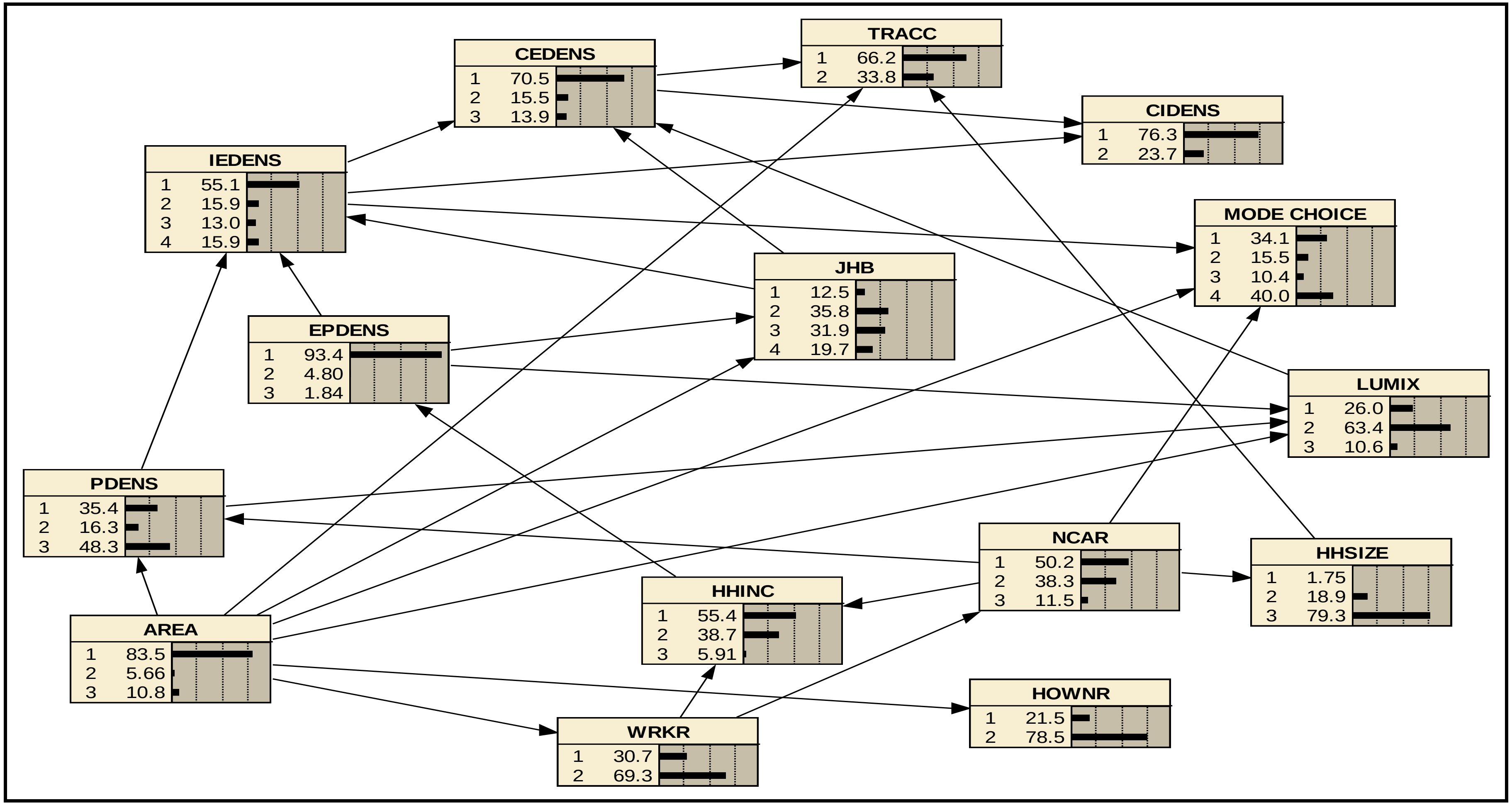


Figure 5.6. The aggregate BBN model of home - based work mode choice in Istanbul

Note: The Node, MODECHOICE, has four states: 1 (WALK), 2 (CAR), 3 (SERVICE), and 4 (TRANSIT).

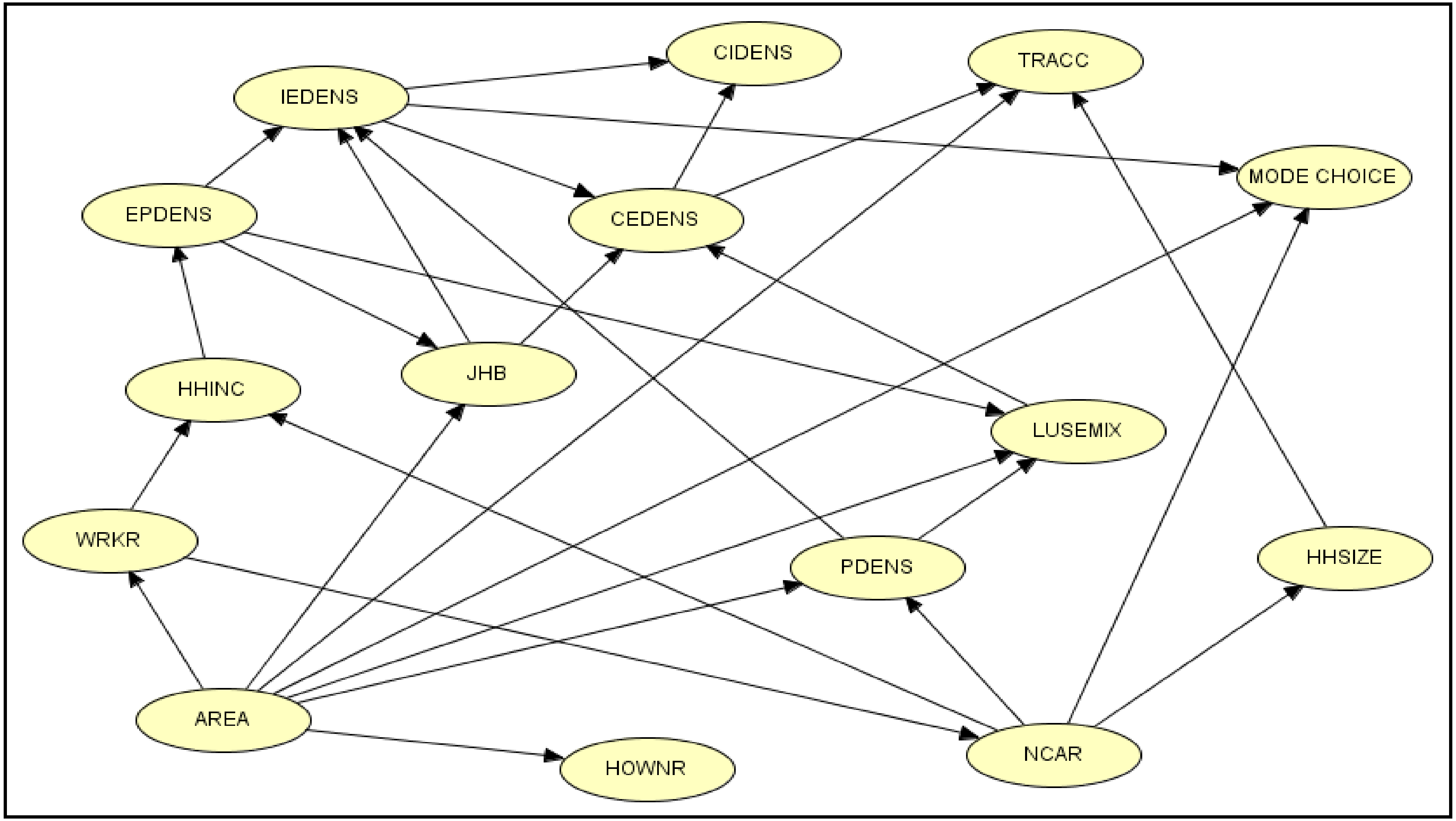


Figure 5.7. Learned BBNs from aggregate data for home - based work trips in Istanbul (Fixed Number of Alternatives)

According to the network, mode choice is affected by the industrial employment density, the size of zonal area, and the number of cars in household in Istanbul. These nodes are parent nodes of mode choice. For compiled network, mode choice is walk with a probability of 34.1%, while its probabilities of being car, service, and transit are 15.5%, 10.4%, and 40%, respectively. These probabilies presents beliefs. The beliefs (probabilities) will be updated as soon as new evidence is entered into the network. For the inference process, Hugin software is used. Table 5.8 presents the revised probabilities at aggregate level that are based on the evidence according to travel modes in Istanbul.

Table 5.8. Inference results on evidence for mode choice variables at aggregate level

Nodes	**States of**	**Mode Choice**		
Evidence for The States	**walk**	**transit**	**car**	**service**
HHINC (1)	63.3	57.9	29.9	58.2
HHINC (2)	32.7	37.6	56.7	35.3
HHINC (3)	4.04	4.47	13.3	6.52
PDENS (1)	30.2	34.9	42.5	44.2
PDENS (2)	12.6	16.9	23.6	15.4
PDENS (3)	57.2	48.3	33.9	40.5
WRKR (1)	33.9	29	24.1	36.3
WRKR (2)	66.1	71	75.9	63.7
NCAR (1)	61.9	53	14.5	54.5
NCAR (2)	33.6	40.5	47.2	32.3
NCAR (3)	4.57	6.49	3.83	13.2
LUMIX (1)	26.0	22.4	29.3	35.1
LUMIX (2)	63.9	67.2	59	53.7
LUMIX (3)	10.1	10.4	11.8	11.3

Increase in income tends to decrease the share of walk, transit, and service modes. On the other hand, the choice probability of car mode in commuters with high income and low income levels is high while choice probability of car mode in commuters with medium income level is higher than other levels. The choice probability for walk and transit is negatively correlated with the number of car in household. Regarding the land use variables, increase in population density tends to increase the choice probability of walk and transit modes while increase in density tends to decrease the share of car trips.

Table 5.9 presents the beliefs for the node “Mode Choice” as a function of entered evidence of the nodes “AREA”. There is a causal relationship between mode choice and the size of zonal area (AREA). Increase in the size of zonal area promote to the choice probability for walk and car mode while it decrease the choice probability for transit mode.

Table 5.9. Inference results based on evidence for zonal area (probabilities)

Node	**Evidence**	**Mode Choice**			
		Walk	**Car**	**Service**	**Transit**
A	**No evidence**	34.1	15.5	10.4	40
R	**1**	32.9	15.1	8.18	43.8
E	**2**	37.8	16.9	22.5	22.7
A	**3**	42.1	17.1	21.4	19.4

There are several scoring measures (rules) for classification success rate. For example, spherical payoff varying in the interval 0 and 1 is the most useful scoring rule in BBNs. In the content of the study, Kulbach - Leibler divergence and Euclidian distance are used. These measures can be estimated by Hugin (or Netica) software. 1 represents the best model performance. Kulbach - Leibler divergence is 0.98324 indicating good model performance. Another score, Euclidian distance, is 0.55455 for aggregate data.

5.2. Disaggregate Model Results

At disaggregate level, the empirical analysis of mode choice for home - based work trips in Istanbul applies multinomial logit model (MNL) as a traditional model and Bayesian belief networks (BBNs) as an alternative method. For both models, the findings of the models are discussed. In MNL model, several models are estimated for the purpose of gauging the marginal influence of land use characteristics on home - based work mode choice in Istanbul. The significant variables in expanded MNL model are used in BBNs. Test statistics for the models are detailed discussed in this section.

5.2.1. Multinomial Logit Model Results

The results of MNL models are presented in Table 5.10 - Table 5.12. Firstly, the results of the base model are discussed. As mentioned before, the base model is a model with only the alternative specific constant (ASC). Base model has only three dummy variables for different travel modes. Each coefficient represents the relative preference for each mode compared with service mode. According to the base model, the estimate for walk and transit mode is positive, reflecting a relative preference for these modes over service mode while auto mode is negative, reflecting a disutility over service mode for hbw trips. In Model 1, travel attributes (generic variables) are entered into the base model. Here time is measured in minutes and cost in T.L. The coefficients of time and cost are the same in the utilities of alternatives. It means that a minute (or a T.L.) has the same marginal utility (or disutility) whether it is incurred on travel modes in Istanbul. The estimate of these variables is negative since, all else being equal; commuters prefer lower time and cost alternatives, as expected. According to this model, each additional minute of travel time reduces the odds of choosing that alternative by 2.95 %. Generic variables have a statistical significance on mode choice in Model 1.

In Model 2 and expanded model, the estimated coefficients on travel time and cost retains their signs, implying that the utility of a travel mode decreases as that mode become more expensive or take up more time. In other words, this situation results in reducing the choice probability of the corresponding mode. In Model 2 and the expanded model, all the coefficient estimates have the expected signs. The estimated coefficients for ASC retain their sign in three out of four MNL models while the estimated coefficients for socioeconomic variables retain their sign and level of significance. Regarding the results of model 2 and the expanded model, car travel is significantly less attractive than service travel. Travel characteristics (time and cost) are in the expected direction with a p-value below 0.05. The expanded model indicates that each additional minute results in a 2% reduction in the odds of choice. The coefficients for the mode dummies are highly significant. Therefore, people prefer to go by walk and transit modes compared with a service. Socioeconomic characteristics have significant effect on mode choice behavior in Istanbul. For example, increasing the number of household auto and company car promotes the probability of choosing auto modes. Each T.L. increase in income increases the probability of choosing car mode over service. Commuters who have akbil cards used in public transportation vehicles tend to travel by transit mode. Akbil possession increases the odds of choosing a transit mode over a service by 69%. With respect to driving license, it was found to have positive parameter on the choice of car mode.

Regarding land use variables, for home - based work trips, increase in population density at both origins and destinations is significantly associated with the choice of walk, auto, and transit modes. In other words, the coefficients for population density represent the relative preferences for each mode compared with service mode. Pinjari et al. (2007) in San Francisco Bay area and Zhang (2004) in Boston found that population density at both origin and destination has a positive effect in the use of walk, bicycle, and transit usage. Also, the relationship is statistically significant. The findings of the thesis for population density are consistent with the previous studies. Also, hypothesis H1 is supported.

Table 5.10. Multinomial logit model results for base model and model 1

Variables	Base Model			Model 1		
	Parameter	t value	p-values	Parameter	t value	p-values
Travel Characteristics						
Travel Cost (TCOST)				**-0.0545**	**-33.17**	**0.0001**
Travel Time (TTIME)				**-0.0295**	**-131.88**	**0.0001**
Mode Constants						
Walk	**0.3542**	**37.55**	**0.0001**	**2.1681**	**145.89**	**0.0001**
Transit	**0.4121**	**44.21**	**0.0001**	**1.0299**	**94.81**	**0.0001**
Car	**-0.0439**	**-4.24**	**0.0001**	**0.0791**	**5.82**	**0.0001**
Service (base)						
Socioeconomic Characteristics						
Driver's License: 0=No, 1=Yes (car)						
Unlimited Akbil Card Usage: 0=No, 1=Yes (transit)						
Akbil Card Usage: 0=No, 1=Yes (transit)						
Income (car)						
Vehicle Ownership: number of automobile in HH (car)						
Company Car Ownership: number of automobile in HH (car)						
Log-Likelihood of Unrestricted Model (LLU)			**259498**			**259498**
Log-Likelihood of Restricted Model (LLR)			**3829.8**			**38907**
Log Likelihood Function			**-127834**			**-110295**
Estrella			**0.0404**			**0.3626**
Adjusted Estrella			**0.0403**			**0.3625**

Table 5.11. Multinomial logit model results for model 2

Variables	Model 2		
	Parameter	t value	p-values
Travel Characteristics			
Travel Cost (TCOST)	**-0.0627**	**-33.80**	**0.0001**
Travel Time (TTIME)	**-0.0301**	**-128.24**	**0.0001**
Mode Constants			
Walk	**2.1433**	**139.99**	**0.0001**
Transit	**0.6954**	**57.50**	**0.0001**
Car	**-1.8857**	**-67.85**	**0.0001**
Service (base)			
Socioeconomic Characteristics			
Driver's License: 0=No, 1=Yes (car)	**1.2883**	**51.38**	**0.0001**
Unlimited Akbil Card Usage: 0=No, 1=Yes (transit)	**1.0736**	**53.36**	**0.0001**
Akbil Card Usage: 0=No, 1=Yes (transit)	**0.8061**	**39.52**	**0.0001**
Income (car)	**0.000123**	**12.74**	**0.0001**
Vehicle Ownership: number of automobile in HH (car)	**1.6148**	**87.96**	**0.0001**
Company Car Ownership: number of automobile in HH (car)	**1.4307**	**38.59**	**0.0001**
Land Use Chatacteristics			
Intrazonal Travel (Trips which begin and end in the same traffic zone)			
Emp./ Worker Ratio at origin (walk)			
Emp./ Worker Ratio at origin (transit)			
Emp./ Worker Ratio at origin (car)			
Emp./ Worker Ratio at destination (walk)			
Emp./ Worker Ratio at destination (transit)			
Emp./ Worker Ratio at destination (car)			
Pop. Density at origin (walk)			
Pop. Density at origin (transit)			
Pop. Density at origin (car)			
Pop. Density at destination (walk)			
Pop. Density at destination (transit)			
Pop. Density at destination (car)			
Transit Accessibility at origin :0=No, 1=Yes (transit)			
Transit Accessibility at destination :0=No, 1=Yes (transit)			
Log-Likelihood of Unrestricted Model (LLU)			**259498**
Log-Likelihood of Restricted Model (LLR)			**65402**
Log Likelihood Function			**-97048**
Estrella			**0.553**
Adjusted Estrella			**0.553**

Table 5.12. Multinomial logit model results for expanded model

Variables	Expanded Model		
	Parameter	t value	p-values
Travel Characteristics			
Travel Cost (TCOST)	-0.0532	-27.71	0.0001
Travel Time (TTIME)	-0.0203	-80.76	0.0001
Mode Constants			
Walk	0.4628	15.21	0.0001
Transit	0.2180	10.78	0.0001
Car	-2.1136	-62.03	0.0001
Service (base)			
Socioeconomic Characteristics			
Driver's License: 0=No, 1=Yes (car)	1.2881	51.02	0.0001
Unlimited Akbil Card Usage: 0=No, 1=Yes (transit)	0.9831	47.73	0.0001
Akbil Card Usage: 0=No, 1=Yes (transit)	0.6925	33.21	0.0001
Income (car)	0.00011	11.39	0.0001
Vehicle Ownership: number of automobile in HH (car)	1.6704	89.86	0.0001
Company Car Ownership: number of automobile in HH (car)	1.4701	39.26	0.0001
Land Use Chatacteristics			
Intrazonal Travel (walk)	2.2118	96.71	0.0001
Emp. / Worker Ratio at origin (walk)	0.0041	6.50	0.0001
Emp. / Worker Ratio at origin (transit)	0.0067	15.70	0.0001
Emp. / Worker Ratio at origin (car)	0.0033	6.50	0.0001
Emp. / Worker Ratio at destination (walk)	-0.00036	-0.56	0.5729
Emp. / Worker Ratio at destination (transit)	0.0027	6.42	0.0001
Emp. / Worker Ratio at destination (car)	0.0030	6.16	0.0001
Pop. Density at origin (walk)	0.00072	10.57	0.0001
Pop. Density at origin (transit)	0.00026	4.75	0.0001
Pop. Density at origin (car)	0.00051	7.84	0.0001
Pop. Density at destination (walk)	0.0010	14.81	0.0001
Pop. Density at destination (transit)	0.0004	7.35	0.0001
Pop. Density at destination (car)	0.0004	6.16	0.0001
TRACC at origin: 0=No, 1=Yes (transit)	0.2322	13.05	0.0001
TRACC at destination: 0=No, 1=Yes (transit)	0.1845	0.0178	0.0001
Log-Likelihood of Unrestricted Model (LLU)			259498
Log-Likelihood of Restricted Model (LLR)			77840
Log Likelihood Function			-90829
Estrella			0.628
Adjusted Estrella			0.627

Increase in the ratio for employment / worker (ewdens) as a measure of jobs - housing balance (jhb) at origins is positively correlated with travel modes (walk, car, and transit modes). However, at destinations, this ratio did not show statistically significance to commuter’s decisions with walk travel. Increase in this ratio at destinations is negatively associated with walk mode. People tend to choose other travel modes rather than unmotorized modes. Jobs - housing balance show statistically significant for all the modes at the origin and at the destination at the 1 % level, except walk mode at destination. The findings are consistent with the research by Zhang (2004) who found that land use balance had no influence on mode choice for commuting by transit and nonmotorized modes (walk and bicycle) at trip origins in Boston. In this study, ewdens ratio only at the trip origins is associated with higher probabilities of commuting by walk mode. Higher balance promotes to use motorized modes (car and transit) at the trip origins and the trip destinations. Jobs - housing balance may matter to home - based work trips in Istanbul. There is not enough evidence for hypothesis H3. Land use balance as a measure of diversity is positively correlated with the choice of walk and transit modes at the trip origins while land use balance is only positively correlated with motorized trips at trip destinations.

In terms of employment density, Vega and Reynolds-Feighan (2008) found that employment density showed small negative effects on the choice probability for car use in Dublin. Zhang (2004) found that employment density (jobs/acre) at both origins and destinations were associated with higher probabilities of commuting by transit and nonmotorized modes (walk and bike) in Boston. In Hong Kong, Zhang (2004) found that employment density showed statistical relevance to commuter’s decision on travel by rail and bus at trip origin and by drive at trip destination. Cervero (2002) suggested that gross density (population+employment / gross square miles) were statistically significant to commuting by drive alone and group ride automobile in Montgomery County, Maryland. Vega and Reynolds-Feighan (2006) suggested that the size of employment area of the job destination have a positive effect in the use of public transport. In terms of employment density, the findings are not consistent with each other in terms of significance. In this study, employment density variable was not entered into the models due to multicorrelation problem with ewdens variable. However, there is enough evidence of supporting for hypothesis H2.

The presence of transit access (sea or rail) to work places promotes transit usage, as expected. It can be considered that travel zones with higher transit access promote people to live near rail stops or ferry stations. It is due to residential sorting effects. Therefore, hypothesis H4 is supported. When studying the impact of land use dummy variable that selects commuters working and living in the same area, the sign of intrazonal travel variable confirmed the hypothesis. People tend to make intrazonal travel for home - based work trips and commuters who live in the zones that include working and living areas tend to choose more walk alternative compared with service. This finding may be due to traffic congestion and parking problems. This finding is consistent with research by Vega and Reynolds-Feighan (2006) that commuters working and living in the same area reduce car travel in Dublin. Also, this finding supports hypothesis H5.

To determine whether the expanded model is statistically significant and overall model significance, several test statistics are used: informal tests, goodness of fit measures, and other statistical tests. These tests provide guidance for evaluating each model and compare different model specifications.

a. *Informal Tests:*

Firstly, some informal tests are applied into the models. Informal tests include examining the signs of the model parameters associated with the existing literature or priori expectations. One of the most used tests is to check the signs of the model parameters. All the coefficient estimates have the expected signs. Travel time and cost variables is negative, as expected, implying that the utility of a travel mode in Istanbul increases as the mode becomes cheaper or slower. The estimate of alternative specific constant variables is positive, except auto, implying that walk and transit modes have a relative preference over service mode. All estimated coefficients of socioeconomic variables have the expected positive sign.

Regarding the land use variables, the signs may not fit the sings of previous studies. In the expanded model, land use variables are entered into the model. Firstly, entering land use variables to the model did not change the signs of other variables. All variables in model 2 retained their significance to predict mode choice in Istanbul. It means that land use has an independent influence on mode choice in Istanbul. The result is a similar with the findings of Cervero (2002), Cervero and Kockelman (1997), and Zhang (2004) studies.

b. Determining Overall Model Significance (Test of Entire Models):

In order to test the hypothesis that expanded model statistically improves upon other models, a likelihood ratio test is used in the same way that the F test is used in multiple regression model. The likelihood ratio test is also used to determine overall model significance in multinomial logit model applications. Likelihood ratio test provide information about the estimated model parameters whether or not they improve the predictive capability of the model. To compare log-likelihood function of any model against the LL of other model, the formulation of LL ratio test statistics is (Ben Akiva and Lerman 1985, Hensher, et al. 2005):

Under the null hypothesis that all the parameters are equal to zero:

$$-2(LL_{base\,\mathrm{mode}l} - LL_{estimated\,\mathrm{mode}l}) \sim \chi^2 \qquad (5.10)$$

$$-2(LL_R - LL_U) \sim \chi^2 \qquad (5.11)$$

It is χ^2 distributed with k degrees of freedom. Where LL_R represents the log-likelihood with base model (restricted model), LL_U represents the log-likelihood for the estimated model such as expanded model. Chi-squared distributions and critical values against different degrees of freedom are presented in Figure 5.8 and Table 5.13. Any model specification with a higher value of log-likelihood function is accepted to be better than other model. This situation can be obtained from several softwares such as SAS, STATA, and Nlogit. Table 5.14 presents the summary measures of goodness of fit obtained from SAS and Nlogit. This test statistic is chi-squared distributed.

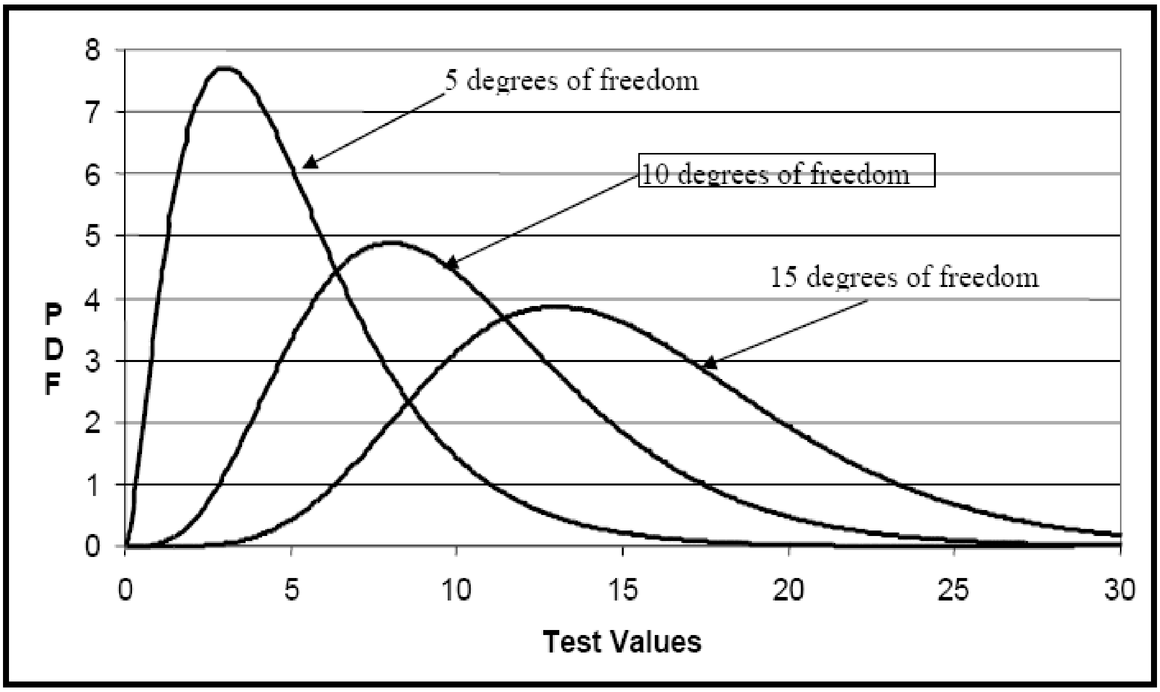


Figure 5.8. Chi squared distributions for different degrees of freedom (Source: Koppelman and Bhat 2006).

Table 5.13. Critical chi-squared (χ^2) values for confidence levels (Source: Evans 2007)

Level of Confidence	Critical Values							
	1	**2**	**3**	**4**	**5**	**10**	**20**	**21**
90%	2.71	4.61	6.25	7.78	9.24	15.99	28.41	29.62
95%	3.84	5.99	7.81	9.49	11.07	18.31	31.41	32.67
99%	6.63	9.21	11.34	13.28	15.09	23.21	37.57	38.93
99.90%	10.83	13.82	16.27	18.47	20.51	29.59	45.32	46.8

To determine whether the expanded model including land use variables is superior to Model 2, the estimated log-likelihood ratio tests (-2LL) is compared to a chi-square statistic with degrees of freedom equal to the difference in the number of parameters estimated for estimated model and Model 2. The expanded model includes 26 parameters while Model 2 includes 11 parameters. Regarding land use variables, main hypothesis of the study is that land use variables characteristics affect mode choice decisions for home - based work trips in Istanbul at both level.

$$H_0 = \beta_{Land\ Use\ Variables} = 0 \tag{5.12}$$

$$\chi^2 = -2\left[-90829 - \left(-97048\right)\right] = 12438 \tag{5.13}$$

The statistical test of the hypothesis that land use has no effect on mode choice with 15 (26-11) degrees of freedom at α=0.05 (95% confidence), the critical $\chi^2_{(21)d.f.}$ is 24,99. The log-likelihood ratio tests are summarized in Table 5.14. The log-likelihood ratio test can be used for comparing different choice model specifications.

Table 5.14. Goodness of fit measures for MNL models

Goodness of Fit Measures	**Base model**	**Model 1**	**Model 2**	**Expanded Model**
Likelihood Ratio (R)	3829.8	38907	65402	77840
Upper Bound of R (U)	259498	259498	259498	259498
Log Likelihood Function at convergence	-127834	-110295	-97048	-90829
Log Likelihood Function at constants	-127834	-127834	-127834	-127834
Test Statistic [-2*(LLR - LLU)]		35078	61572	74010
Degree of Freedom	3	5	11	26
Critical Chi-Squared Value at 99% Confidence		20.515	31.264	54.051
Rejection Confidence	99.9%	99.9%	99.9%	99.9%
Rejection Significance	0.001	0.001	0.001	0.001
Estrella	0.0404	0.3626	0.553	0.628
Adjusted Estrella	0.0403	0.3625	0.5528	0.6277
McFadden's LRI				

When comparing the test statistics of 12438 to the chi-square critical value of 24.99, the test statistic is greater than the critical value. If the -2LL value exceeds the critical chi-square value, the null hypothesis that the specified model is no better than the base comparison model is rejected. It means that analyst is able to reject the hypothesis that expanded model does not statistically improve the LL over the Model 2. The log-likelihood of expanded model is statistically closer to zero than that of Model 2. The null hypothesis is rejected with high confidence. In other words, land use variables should not be excluded from the model. Expanded model including land use variables outperforms Model 2 that includes socioeconomic, ASC, and generic variables.

c. *Overall Goodness of Fit (Determining Model Fit)*

LL ratio test provide guidance to compare different choice model specifications. Other test for determining model fit is called as pseudo R^2. This ratio is suggested by McFadden (1974) and it is called as the likelihood ratio index (rho-squared) that is analogous to the R^2 in linear regression model. A pseudo R^2 or likelihood ratio index (ρ^2) for a choice model is estimated by the following formulation:

$$R^2 = 1 - \frac{LL_{estimated\,\mathrm{model}}}{LL_{base\,\mathrm{model}}} \quad (5.14)$$

$$or \rho^2 = 1 - \frac{\ln L}{\ln L_0} \quad (5.15)$$

$$\rho^2 = 1 - \frac{-97048}{-127834} = 0{,}32 \text{ for Model 2} \quad (5.16)$$

$$\rho^2 = 1 - \frac{-90829}{-127834} = 0{,}41 \text{ for the Expanded Model} \quad (5.17)$$

Where L presents the value of the maximum likelihood function at maximum and L_0 is a likelihood function when regression coefficients are zero. McFadden's likelihood ratio index is ranged from 0 to 1. In this case, a value of 0.41 for pseudo R^2 is not equal to an R^2 of 0.41 for a linear regression model since MNL model is non-linear. The pseudo-R^2 in expanded model is higher than the pseudo R^2 in Model 2. It means that expanded model can explain higher variance than Model 2. A pseudo R^2 of 0.41 represents an R^2 of approximately 0.80 for the equivalent R^2 of a linear regression model. McFadden's likelihood ratio index (LRI) or pseudo R^2 can be obtained from SAS and Nlogit softwares as seen in Table 5.14.

Another goodness of fit measure is Estrella and Adjusted Estrella. Estrella (1998) proposes a goodness-of-fit measure to be desirable in discrete choice modeling: (SAS 2004, 663):

1. The measure must take values in [0; 1], where 0 represents no fit and 1 corresponds to perfect fit.

2. The measure should be directly related to the valid test statistic for the significance of all slope coefficients.
3. The derivative of the measure with respect to the test statistic should comply with corresponding derivatives in a linear regression.

Its formulation is written as (SAS 2004):

$$R_{E1}^2 = 1 - \left(\frac{\ln L}{\ln L_0} \right)^{-\frac{2}{N} \ln L_0} \tag{5.18}$$

Where lnL_0 is estimated with null parameter values and N represents the number of observations. For this case, Estrella and Adjusted Estrella increases from the base model including only ASC variables to expanded model. According to Estrella measure, ***the expanded model is superior to the other models***. Estrella measure confirms sub-hypothesis SH-1 that adding land use variables improve the model explanatory power at disaggregate level.

d. Measuring of Willingness to Pay (Value of Travel Time)

One of the aims to use discrete choice models is to measure willingness to pay (WTP) in order to take advantage from the utility of a travel mode. WTP is estimated as the ratio of two parameters including time and cost. When estimating WTP, the parameters for both time and cost are expected to be statistically significant and one of both parameters at least is measured in monetary values. Therefore, both time and cost should be entered into the utility function (Hensher, et al. 2005, Koppelman and Bhat, 2006). WTP presents ***value of travel time savings (VTTS)***. Travel time savings is a measure in transportation literature for determining road and public transportation pricing because travellers may spend money to save time. The most general application in the world is used for calculating the value for money of spending public funds on transport investments. As such, WTP is calculated as follows:

$$VTTS = \left(\frac{\beta_{time}}{\beta_{\cos t}} \right) \times 60 \tag{5.19}$$

For Model 1 including only ASC and generic variables, this formulation gives the result as follows:

$$VTTS = \left(\frac{-0.0295}{-0.0545} \right) \times 60 = 32.48\,T.L./\ perhour \qquad (5.20)$$

In sum, multinomial logit model in expanded form out-performed the basic one that is similar with Cervero (2002) and Zhang (2004). Based on both the pseudo R^2 and likelihood ratio test, land use variables in Istanbul for home - based work trips contribute significantly in explaining travel mode choice decisions for commuters at both aggregate and disaggregate level.

5.2.2. Bayesian Belief Network Results at Disaggregate Level

Bayesian Belief Networks are constructed with the variables that are statistically significant in multinomial logit model (expanded model). According to this, there are 20 nodes (variables), except query node, implying mode choice. After structure learning from mode choice data, parametric learning is applied in BN PowerConstructor. The compiled network in Hugin is shown in Figure 5.9 and 5.10. In total, BN PowerConstructor produced 43 links in this network learned from training data. The network for disaggregate mode choice includes 2264948 conditional probability in total.

Table 5.15. Sensitivity analysis of mode choice for disaggregate analysis

Node	Mutual Info	Variance of Belief
WTIME	0.18672	0.0231627
TRTIME	0.15036	0.0206352
ATIME	0.12371	0.0176776
INTRA	0.10787	0.0180908
STIME	0.10449	0.0128564
TCOST	0.09699	0.0098920
NCAR	0.06259	0.0056779
ACOST	0.05546	0.0057518
DRL	0.02706	0.0015842
SAKBIL	0.00695	0.0008842
INCOME	0.00336	0.0002542
CCAR	0.00284	0.0002793
DEWDENS	0.00280	0.0002795
OTRACC	0.00259	0.0002753
DPDENS	0.00252	0.0002470
OPDENS	0.00237	0.0002325
SCOST	0.00209	0.0002558
DTRACC	0.00181	0.0001931
OEWDENS	0.00172	0.0001721
AKBIL	0.00061	0.0000662

The beliefs are shown for each node in the Figure 5.9. In order to know how sensitive our belief in query node's value is to the findings of other nodes, a sensitivity analysis is estimated. "MODECHOICE" is selected as the query node in this analysis. The result of sensitivity analysis is presented in Table 5.15.

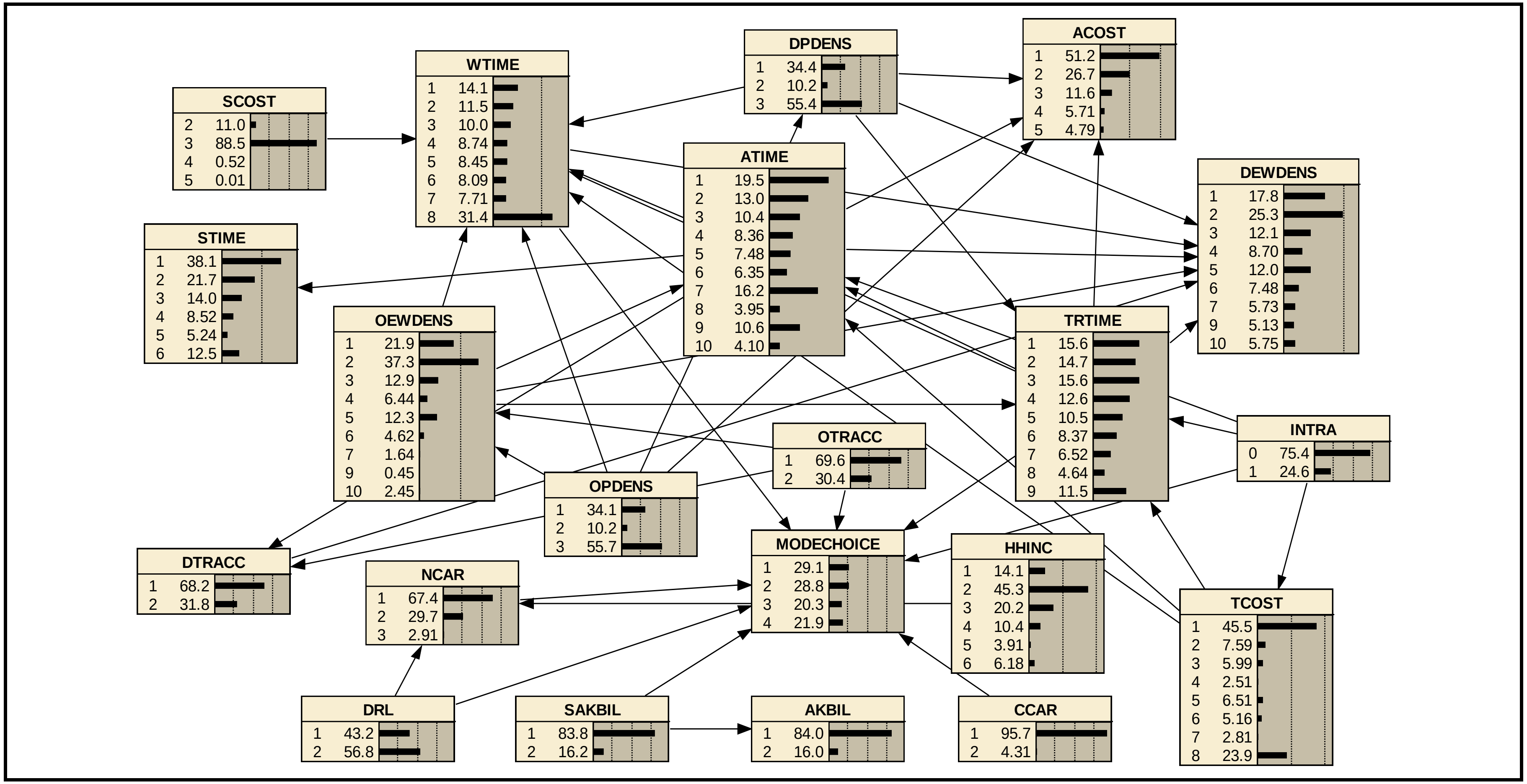


Figure 5.9. The disaggregate BBN model of home - based work mode choice in Istanbul

Note: The Node, MODECHOICE, has four states: 1 (WALK), 2 (TRANSIT), 3 (CAR), and 4 (SERVICE).

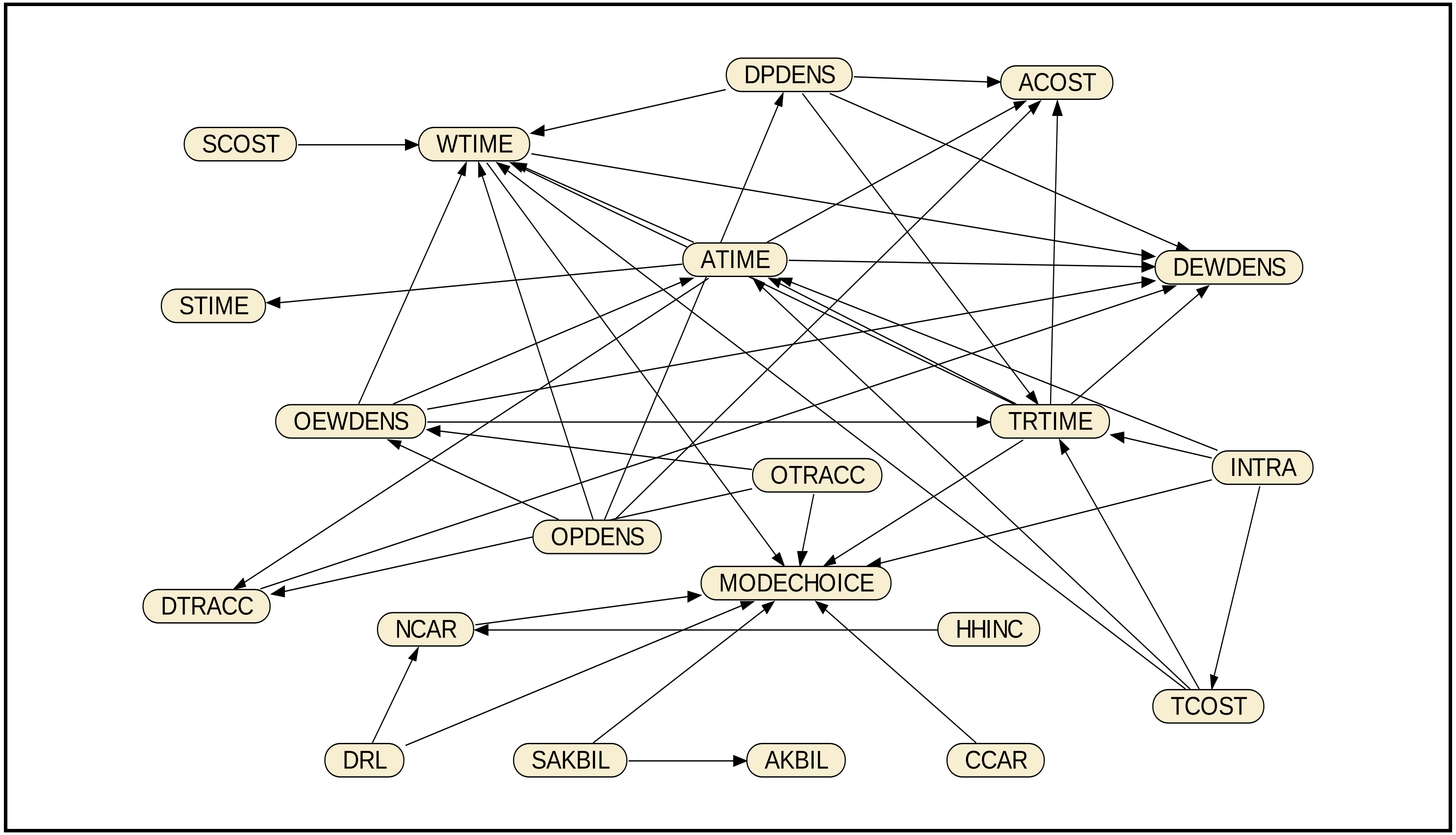


Figure 5.10. Learned BBNs for home - based work trips in Istanbul (Fixed Number of Alternatives)

The table for sensitivity analysis provide guidance about which nodes can most influence the key variable, “Mode Choice”. The degree of influence of the nodes in the network is calculated as a measure of mutual information (or entropy reduction) and quadratic score (or variance of belief). Mutual info provide information about the degree of sensitivity of one node to another in the network while quadratic score represents a measure between query node and other nodes. According to this table, the most influential nodes of "Mode Choice", are: travel time nodes for walking (wtime), transit (trtime), auto (atime), intra travel for walk mode (intra), travel time for service (stime), and travel cost by transit (tcost). Travel time is the most significant factor causing the largest entropy reduction in travel mode choice in Istanbul at disaggregate level. For the inference process, Hugin software is used. Table 5.16 presents the revised probabilities (beliefs) that are based on the evidence according to different travel modes.

Table 5.16. Inference results based on evidence for mode choice variables

Nodes	**States of**	**Mode Choice**		
Evidence for The States	**walk**	**transit**	**car**	**service**
akbil card (1)	29.2	28.2	20.5	22.1
akbil card (2)	28.3	31.8	19.2	20.7
sakbil (1)	29.6	26.8	20.9	22.6
sakbil (2)	26.3	39	16.7	18
driver license (1)	32.4	31.6	11.6	24.5
driver license (2)	26.6	26.7	26.9	19.8
ncar (1)	31.8	32.6	11.9	23.8
ncar (2)	23.8	21.4	36.9	18
ccar (1)	29.3	29.1	19.7	21.8
ccar (2)	23.6	22.1	32.2	22.1
wtime (1)	66.1	10.7	13.7	9.44
wtime (2)	53.4	15.1	17.2	14.3
wtime (3)	41.4	20.8	20.7	17.1
wtime (8)	5.12	42.1	21.7	31
trtime (1)	63	11.4	14.3	11.3
trtime (2)	51.3	15.4	17.2	16.1
trtime (3)	30.9	29.4	19.9	19.8
trtime (9)	7.96	34.9	22.5	34.6

A commuter having an akbil card and unlimited akbil card (sakbil) used in public transportation in Istanbul tends to choose transit mode increasingly. The commuters who have a drive license may tent to go to works by auto, as expected. Increase in car and company car in household increase the choice probability of car mode. Increase in walking travel time cause to decrease the choice probability for walk mode while increase in travel time increases the choice probability for transit mode.

Table 5.17 presents the beliefs for the node “Mode Choice” as a function of entered evidence of the nodes “INCOME”. There is no direct link between mode choice and income. According to Table 5.17, increase in income promote to the choice probability for car mode while it decrease the choice probability for transit mode, as expected.

Table 5.17. Inference results based on evidence for income node (probabilities)

		Mode Choice			
		Walk	**Transit**	**Car**	**Service**
Node	**No evidence**	29.1	28.8	20.3	21.9
I	**0-1**	30.1	30.4	16.9	22.6
N	**1-2**	29.5	29.5	18.8	22.2
C	**2-3**	28.8	28.4	21.1	21.7
O	**3-4**	28.1	27.2	23.6	21.1
M	**4-5**	27.9	26.9	24.1	21
E	**5-6**	26.8	25.3	27.5	20.4

The complexity of the relationship between land use and travel mode choice, the choice of the states and discretisation of the variables in BBNs may influence the accuracy of the network. Kulbach - Leibler Divergence is 0.932 indicating good model performance. Euclidian distance is 0.51429.

5.3. Performance Comparisons of Mode Choice Models

Section 2.4 presents performance comparison methods. There is no standard approach on how to compare the results of BBNs with classical approaches. In order to evaluate the performance of conventional models (logit models) and alternative method (BBNs), confusion matrices are estimated for performance comparisons of aggregate and disaggregate models. Confusion matrix provides a table comparing predicted with actual outcomes. In this study, testing data is used to test the accuracy of the models while training data set is used for estimating model parameters in the models at both aggregate and disaggregate levels.

This matrix provides information about overall accuracy of the models. Error rate estimated from this matrix can be used for performance comparison of the models. The overall accuracy in conjunction with different confusion matrices is estimated using by the formulation as:

$$\text{Overall Accuracy}=\frac{\sum n}{N}\times 100 \tag{5.21}$$

Where n is the total number of pixels that actually belong to that class and N represents the total number of observations in confusion matrix. This formulation can be rewritten as follows:

(Number of Correct Predictions) / (Total Number of Observations) (5.22)

Estimated confusion matrices including proportions of correct predictions (presented in parantheses) for aggregate level are represented in Table 5.18 and Table 5.19. According to this result, baseline category logit model correctly predicted the mode chosen 48 (14+21+12+1) times out of the total of 81 choices made whereas BBNs correctly predicted 48 (16+22+10+0) times out of the total of 81 choices made. These choice models correctly predicted the actual choice outcome for almost 59 percent of the total number of cases in test set.

Table 5.18. Confusion matrix for baseline category logit model

Actual	Predicted: Walk	Transit	Car	Service	Total
Walk	**14 (0,583)**	9 (0,375)	1 (0,042)	0 (0)	**24**
Transit	8 (0,242)	**21 (0,636)**	3 (0,09)	1 (0,03)	**33**
Car	1 (0,053)	6 (0,316)	**12 (0,632)**	0 (0)	**19**
Service	1 (0,2)	3 (0,6)	0 (0)	**1 (0,2)**	**5**
Total	***24***	***39***	***16***	***2***	**81**

Table 5.19. Confusion matrix for BBNs at aggregate level

Actual	Predicted: Walk	Transit	Car	Service	Total
Walk	**16 (0,552)**	9 (0,31)	2 (0,069)	2 (0,069)	**29**
Transit	8 (0,2)	**22 (0,55)**	7 (0,175)	3 (0,075)	**40**
Car	0 (0)	2 (0,167)	**10 (0,833)**	0 (0)	**12**
Service	0	0	0	**0**	**0**
Total	***24***	***33***	***19***	***5***	**81**

According to the tables in the above, Baseline category logit model predicts the walk mode correctly 58 percent of the time, the transit alternative correctly about 64 percent of the time, and the car, and the service alternatives correctly about 63 and 20 percent of the time, respectively. BBNs predicts the walk alternative correctly 55 percent of the time, the transit alternative correctly 55 percent of the time, and the car and the service modes correctly 83 and 0 percent of the time, respectively. At aggregate level, BBNs failed to predict the service mode more than baseline category logit model. In order to increase the model performance, new estimation algorithms or the states of the nodes in BBNs should be adjusted again. Both models can predict the car and transit alternatives correctly more than other alternatives.

The overall model error rate in BBNs is 40.74%, implying that the model had the majority of its predictions correct for mode choice observations while the overall model rate is 40.74% in baseline category logit model. According to this result, aggregate models have the same error rate. In this situation, there is no superiority of BBNs over baseline category logit model.

For disaggregate level (as seen in Table 5.20 and Table 5.21), multinomial logit model correctly predicted the mode chosen 13975 (5430 + 4362 + 2807 + 1376) times out of the total of 23398 choices made whereas BBNs correctly predicted 14363 (5825 + 4661 + 2874 + 1003) times out of the total of 23398 choices made. For test set, MNL model correctly predicted the actual choice outcome for almost 60 percent of the total number of cases in test set whereas BBNs correctly predicted the actual choice outcome for almost 61 percent of the total number of cases in test set.

Table 5.20. Confusion matrix for MNL at disaggregate level

Actual	**Predicted**				**Total**
	Walk	**Transit**	**Car**	**Service**	
Walk	**5430 (0,722)**	748 (0,099)	609 (0,081)	726 (0,096)	**7513**
Transit	877 (0,112)	**4362 (0,557)**	835 (0,106)	1755 (0,224)	**7829**
Car	298 (0,061)	911 (0,186)	**2807 (0,575)**	862 (0,176)	**4878**
Service	193 (0,060)	1247 (0,392)	362 (0,114)	**1376 (0,432)**	**3178**
Total	***6798***	***7268***	***4613***	***4719***	**23398**

Table 5.21. Confusion matrix for BBNs at disaggregate level

Actual	Predicted				Total
	Walk	Transit	Car	Service	
Walk	**5825 (0.724)**	836 (0.103)	623 (0.077)	758 (0.094)	**8042**
Transit	529 (0.065)	**4661 (0.570)**	901 (0.110)	2081 (0.255)	**8172**
Car	290 (0.058)	920 (0.185)	**2874 (0.579)**	877 (0.177)	**4961**
Service	154 (0.069)	851 (0.383)	215 (0.097)	**1003 (0.451)**	**2223**
Total	***6798***	***7268***	***4613***	***4719***	**23398**

According to the tables for disaggregate models, MNL model predicts the walk mode correctly 72 percent of the time, the transit alternative correctly about 56 percent of the time, and the car, and the service alternatives correctly about 58 and 43 percent of the time, respectively. BBNs predicts the walk alternative correctly 72 percent of the time, the transit alternative correctly 57 percent of the time, and the car and the service modes correctly 58 and 45 percent of the time, respectively. In opposition to the aggregate models, disaggregate BBNs succeed more than aggregate BBNs in terms of model performance. At aggregate level, baseline category logit model predicts transit mode correctly while Bayesian belief networks correctly predict car mode more than other alternatives. At disaggregate level, multinomial logit model and BBNs are more successful for the correct prediction of walk mode.

The overall model error rate in BBNs is 38.61%, implying that the model had the majority of its predictions correct for mode choice observations while the overall model rate is almost 40% in MNL. According to this result, BBNs as an alternative method in disaggregate mode choice modeling is superior upon the MNL.

According to the results, hypohesis (SH-2) that alternative method is superior to conventional models is confirmed for only disaggregate level. At aggregate level, superiority is not clear. As mentioned in Chapter 4, there are two different data structures (data setup) for discrete choice model estimation. When every individual can select all alternatives, the choice set size is a fixed size. However, in reality, all alternatives (all choices) are not available to all individuals such as limited income and lack of driving license. In this situation, the number of alternatives vary across individuals. According to choice analysis, the data should be structured in one of two formats. At disaggregate level, the study analyzed model results and model performances for different data formats. The model results that are estimated in Nlogit program, for choice set with fixed size and varying size for full data are given in Table 5.22.

Table 5.22. MNL model results for different choice data

Variables	**Fixed Number of Choices**			**Variable Number of Choices**		
	Parameter	**t value**	**P values**	**Parameter**	**t value**	**P values**
Travel Characteristics						
Travel Time (TTIME)	-0.0098	-97.69	0.0001	-0.0320	-141.843	0.0001
Travel Cost (TCOST)	-0.8166	-63.54	0.0001	-0.0741	-38.724	0.0001
Number of Observation	116992			116992		
Log Likelihood Function	-154688			-74604		

As expected, the estimate of generic variables (time and cost) is negative for both different structure types. Lower time and cost alternatives in home - based work trips are preferred in the case of Istanbul. Generic variables have statistically significant effect on mode choice. Model performances that are evaluated in a contingency table (confusion matrix) are shown in Table 5.23 and 5.24. Mode choice model used for fixed numbers of choices predicts walk alternative correctly 24 percent of the time, transit alternative correctly 27 percent of the time, and car and service alternatives correctly 24 and 36 percent of the time, respectively. On the other hand, mode choice model used for variable numbers of choices predicts walk alternative correctly 39 percent of the time, transit alternative correctly 79 percent of the time, and car and service alternatives correctly 59 and 44 percent of the time, respectively.

Table 5.23. Confusion matrix at disaggregate level for fixed numbers of choices

Actual	**Predicted** Walk	Transit	Car	Service	**Total**
Walk	**8168 (0.24)**	8676 (0.25)	8464 (0.25)	8753 (0.26)	***34061***
Transit	5158 (0.14)	**9828 (0.27)**	8728 (0.24)	12442 (0.35)	***36156***
Car	3462 (0.15)	6124 (0.27)	**5587 (0.24)**	7751 (0.34)	***22924***
Service	3196 (0.13)	6416 (0.27)	5632 (0.24)	**8607 (0.36)**	***23851***
Total	***19984***	***31044***	***28411***	***37553***	**116992**

Table 5.24. Confusion matrix at disaggregate level for variable numbers of choices

Actual	**Predicted** Walk	Transit	Car	Service	**Total**
Walk	**13139 (0.39)**	19034 (0.56)	1888 (0.05)	0	***34061***
Transit	4847 (0.13)	**28582 (0.79)**	2726 (0.08)	0	***36156***
Car	3407 (0.08)	14000 (0.33)	**25554 (0.59)**	0	***42961***
Service	128 (0.03)	774 (0.20)	1242 (0.33)	**1670 (0.44)**	***3814***
Total	***21521***	***62390***	***31410***	***1670***	**116992**

According to the results, the model performance for variable numbers of choices is superior upon the model including the fixed numbers of choices. In the content of the study, the expanded model specification including variable numbers of choices are estimated using Nlogit. However, the model give insignificant t values and unexpected signs of the coefficients. The model produce an error that Nlogit is unable to estimate standard errors for utility function of the expanded form. The pattern of missingness of the alternatives may cause this situation that the parameters are not identified.

CHAPTER 6

CONCLUSION

The purpose of this research is to expand the understanding derived from the previous empirical research on the effects of land use on mode choice behavior for home - based work trips (HBW) by accounting for conventional (logit models) and an alternative approach (BBNs). In existing literature, urban settings mostly took place in North-American and European cities. From developing part of the world, there have been lack of empirical evidence to support the relationship between land use and mode choice. The study introduced several socieconomic and land use characteristics relevant to the topic under discussion. In order to achieve a better understanding of the relationship between land use and travel demand, comparing and analyzing the results of aggregate and disaggregate models together needed to be develop. While previous studies has tried to analyze the effects at either aggregate level or disaggregate level, this study has analyzed the effects of land use on mode choice at both levels. Therefore, this approach has provided detailed information about the effects for comparing the results with different cases in the literature. The empirical analysis in this study is based on 2006 Household Travel Survey prepared for 2007 Istanbul Transportation Master Plan. The model specifications tested several variables describing zonal - individual socioeconomic characteristics, travel characteristics for disaggregate analysis, and land use characteristics. In the content of the study, land use characteristics have been approximated mainly by density (population and employment), diversity (land use mix and jobs - housing balance), and accessibility.

In the last 15 years, soft computing methods, especially neural networks, fuzzy logic, and hybrid approaches (neuro-fuzzy modeling) become more attractive than conventional models. BBNs are new models and more flexible than logit models in mode choice modeling. An important contribution of the study is that bayesian belief networks (BBNs) that have been rarely used in mode choice studies have been proposed to analyze complex and probabilistic relationships among the variables. In opposition to the previous studies such as Scuderi and Clifton (2005), BBNs are used to inference and forecast in the content of the study.

The overall results of the study are summarized as follows: (1) land use characteristics have an independent influence on mode choice for home - based work trips in Istanbul. Many variables retained their signs and statistical significance after the inclusion of land use characteristics. (2) There is evidence to support hypothesis SH-1 that adding land use variables to the models at aggregate and disaggregate levels improves the model's explanatory power. The result is consistent with the findings of the previous studies (e.g. Cervero and Kockelman (1997), Zhang (2004), and Cervero (2002)). (3) The hypothesis (SH-2) that soft computing methods (BBNs) are superior to conventional models (logit models) in mode choice modeling at both levels is supported only at disaggregate level. For this case, there is no evidence to support this hypothesis (SH-2) at aggregate level.

The empirical results show that, as in the case of aggregate level data, many socieconomic factors significantly affects travel mode choice for HBW trips in Istanbul. The variables, zonal average of working (**wrkr**), for car mode and household size (**hhsize**) have not statistically significant effect on mode choice. The signs of socieconomic characteristics are as expected. The variables, household income (**hhinc**), household size (**hhsize**), house ownership (**hownr**), car ownership (**ncar**), zonal average of working (**wrkr**) are positively correlated with motorized trips. Regarding the land use variables, it is found that many land use variables have a statistically significant effect on mode choice. A negative correlation is found between density (population and employment) and motorized trips. This result does not support hypothesis H2 that employment densities positively correlated with motorized trips while the result supports hypothesis H1 for only walking mode. Transit choice at aggregate level is negatively correlated with population density. According the empirical studies in both the USA and Europe (Schwanen, et al., 2004, Frank and Pivo 1994, Newman and Kenworthy 1989, Coevering and Schwanen 2006), traveling by car for home - based work trips is negatively correlated with population density. The result of the study is consistent with the literature at aggregate level. At aggregate level, it is found that land use mix diversity index is positively correlated with motorized trips. This result is not consistent with the previous studies for the cities in developed countries while it is consistent with the previous studies for the cities in Asia. Therefore, there is no enough evidence to support the hypothesis H3.

At aggregate level, hypothesis H4 that the presence of transit access in the zones increases the choice of transit mode, is supported. One important finding is that the size of zonal area is negatively correlated with motorized trips. The size of area has a negative influence on motorized trips at aggregate level. This finding about the size of the zonal area (**area**) suggest that commuters who live in higher zonal area tends to live close to employment areas. Sensitivity analysis in BBNs suggests that the number of car in household (**ncar**), industrial employment density (**iedens**), household income (**hhinc**), household size (**hhsize**), the size of zonal area (**area**), and population density (**pdens**) are the most influential nodes on mode choice. For aggregate analysis, the findings associated with land use characteristics show similarities with the findings of the empirical studies in Asian cities.

The empirical analysis at disaggregate level, carried out using multinomial logit model (MNL) and bayesian belief networks (BBNs), reveals that all socieconomic factors significantly affects travel mode choice for HBW trips in Istanbul. The coefficients for socioeconomic variables are positive, as expected. Surprisingly, the impact of household income on commuter's utility is significant but marginal in comparison with other variables. The coefficients of the type of akbil cards used in public transportation in Istanbul for the mode dummies represents the relative preferences for transit modes even after the inclusion of the land use variables. The coefficients of travel attributes (travel time and travel cost) are negative, as expected. The coefficients of both variables are treated as generic. Negative signs indicates that commuters prefer lower time and cost alternatives for home - based work trips in Istanbul. According to the expanded model, both variables reach statistical significance. Each additional minute in travel time reduces the odds of choosing that alternative by 2% while each additional T.L. in travel cost reduces the odds of choosing that alternative by 5.3%. In the expanded model, the estimate of alternative specific constant (ASC) for walk and transit is positive while the ASC coefficient for car mode is negative. Car travel is significantly less atractive than service travel. Many travel characteristics, socieconomic characteristics, and alternative specific coefficients retain their significance and signs after the inclusion of the land use variables.

Regarding the land use variables for disaggregate analysis, many land use variables at both orgins and destinations are statisticaly significant for mode choice in Istanbul. Main hypothesis of the study is supported. Land use variables in the expanded model improved overall predictability (model's explanatory power), as in the case of the aggregate analysis. Statistical tests (LL ratio and pseudo R^2) confirmed sub-hypothesis of the study, SH-1, that adding land use variables to the models at disaggregate levels improves the model explanatory power. This result for disaggregate data is consistent with the earlier studies done in other countries (Cervero 2002, Zhang 2004) and also in aggregate data. According to overall accuracy for MNL and BBNs, the proportion of explained variation in BBNs is more than MNL models. Therefore, hypothesis SH-2 that soft computing methods (BBNs) are superior to conventional models (logit models) in mode choice modeling, is supported for disaggregate analysis.

A positive correlation is found between population density and relative modes (walk, transit, and car). This result supports hypothesis H1. However, this is not consistent with the result of the expanded model at aggregate level. One of the explanations for this is that walk mode was selected as the referent mode in baseline category logit model while service was treated as the referent mode, meaning coefficients on the utility function was interpreted with reference to the service mode. However, the result supports the findings of Pinjari et al. (2007) in San Francisco, Zhang (2004) in Boston, and Coevering and Schwanen (2006) in the cities of Europe, Canada, and US. Employment / Worker ratio (**oewdens** and **dewdens**) as a measure of employment density is positively correlated with travel modes, except walk mode at destination. It was found enough evidence to support hypothesis H2. Diversity positively influences with walk mode and transit only at the origins. Therefore, hypothesis H3 is supported for walk mode and transit mode.

At disaggregate level, the presence of transit access in the zones is positively correlated with the choice of transit mode, as expected. Hypothesis H4 is supported. Commuters working and living in the same zones (**intra**) tend to use walk mode more than service mode. Hypothesis (H5) that commuters whose trip origin and destination point is in the same zone are more likely to choose non-motorized alternatives, is supported at disaggregate level. When studying the effects of the model variables with BBNs for disaggregate data, sensitivity analysis suggests that travel time for walking (**wtime**), transit (**trtime**), car (**atime**), service (**stime**), travel cost for transit (**tcost**), and intra travel for walk mode (**intra**) are the most influential variables for mode choice.

The results of the study found enough evidence for the relationship between mode choice and land use for home - based work trips in Istanbul. It appears that there are a number of travel and socioeconomic characteristics that may explain the variation in choice behavior. For example, socioeconomic characteristics can explain more of the variation in choice behavior than land use characteristics do. In terms of the distance travelled, land use characteristics may be more important than socioeconomic characteristics.

Further empirical evidence from elsewhere in Turkey is needed to verify the external validity of the effects on mode choice. Some of the variables used in this study require further examination and revision. If GIS data for urban form characteristics are available at neighborhood level, model results may assist in developing sufficient land use policies. Bayesian belief networks may provide more flexible structure of error terms while multinomial logit may not. In BBNs example for aggregate and disaggregate data, the BBNs models do not provide highly accurate in analyzing mode choice when considering high error rates that are about 40%. However, the application of performance test using scoring rules and error rates are one of the first applications for mode choice analysis in travel demand modeling.

Even though the numerous studies have focused on the effects of land use (or urban form) on travel behavior, the debate about the significance, magnitude, and which aspects of land use continues. From the perspective of physical planning and urban policy, the study suggests that land use characteristics are not exogenous in the modeling of mode choice behavior made by individuals, as well as zonal. It is found that the commuters working and living in the same zonal area tend to travel by non-motorized modes for home - based work trips in Istanbul. Also, the presence of transit access in the zones promotes the use of transit. Individuals are consistent with their lifestyle values (preferences). Therefore, it indicates that there is evidence for residential sorting effects in Istanbul. In other words, main reasons to travel by car trips are longer travel, waiting, access, and egress times for public transportation, especially for bus travel.

Existing rail systems should be extended in and around the big industrial areas in Istanbul. New residential developments should be concentrated around rail systems. Physical planning should allow to the commuters with more opportunities for switching travel modes. For example, park and ride systems should be extended to serve commuters. Pricing strategy that provides commuters some opportunities to promote the use of transit modes should be developed. In the city centers of Istanbul, high cost in parking may discourage the use of private automobiles. The estimates of the disaggregate models indicates that individuals are willingness to pay for reduced travel time. It means that a program aimed at reducing traffic congestion and increasing comfort in public transportation needs to be developed. Public transportation provides opportunities for the disadvantages groups such as poor and elderly people. Auto - dependent cities generate more air pollution. On the contrary, transit - oriented cities generate less pollution and more energy savings, especially home - based work trips. For decades, policies aimed at encouraging clustered development, higher densities, and improve level of service (LOS) for transit has been implemented. Also, in Europe and Canada, there are some restrictions on auto use while there are some facilities on transit usage. Some policies such as right of way to buses and auto - free zones may provide opportunities for transit to become safer and more attractive for commuters. Bus rapid transit systems may be a good alternative to save travel cost and travel time. Making rail system projects realize will provide safer, cheaper, and faster opportunities to the commuters in Istanbul. As known that, policentric urban structure and decentralization of land use will cause to more use of the private modes for all trip purposes. On the other hand, it leads to less use of public transportation. Spatial mismatch (jobs - housing imbalanced areas) leads to observe longer commutes. As Istanbul has expanded, the effects of land use on mode choice may become more important. Physical planning may assist in reducing the use of private modes. For example, mixed land use development and jobs - housing balanced areas may play an important role in promoting the use of public transportation. Distance travelled to work is highly related to the development of polycentric urban structure. Understanding the relationship between land use and travel behavior contributes to develop urban policies that aim to reduce motorized travel demand. As an urban policy, jobs - housing balance should be achieved at the two continents: Asia and Europe so that traffic load between two continents may decrease. In addition, traffic flow from one to another may decrease depending on jobs - housing balance.

Regarding the further studies, this study can be expanded in a number of ways. First, modeling mode choice and land use can be analyzed with various aspects of travel behavior such as route choice and vehicle miles traveled (VMT). Analyzing and comparing of various travel demand factors may provide the full picture of the relationship between land use and travel demand. Second, in the case of developing countries, the effects of urban form characteristics on travel behavior needs to be expanded with conventional and alternative approaches. In terms of land use characteristics, the study found significant land use variables at both levels in Istanbul. However, new land use (or urban form) variables may be introduced to the models that can be used with different type of measurements. Third, the impact of land use on mode choice should be tested for home - based school (HBS) trips, home - based other (HBO) trips, and non - home - based (NHB) trips. Land use (or urban form) characteristics, their significance and signs may vary across trip purpose. Therefore, analyzing the effects based on different trip purposes may provide useful information for physical planning. Fourth, this study proposed new models for mode choice analysis. Baseline category logit and bayesian belief networks (BBNs) are rarely used models in transportation modeling. Especially, BBNs should be used for other stages in transportation modeling such as trip generation and trip distribution. Main disadvantage for BBNs is the computation time of the learning algorithms that increases when the number of the states of variables increases. Also, the number of the CPTs in BBNs expands when more states and nodes are involved. Fifth, in mode choice analysis, choice set with fixed number of alternatives that all alternatives are available to all individuals, is generally used. The model performances of the further studies using variable number of choice set may be higher than the models including fixed number of alternatives. The study provides information about this situation. Finally, conventional models, especially multinomial logit models (MNL), have been studied for years. In recent years, soft computing methods have become the alternative methods to discrete choice models. Activity-based approaches to travel analysis, bayesian belief networks (BBNs), structural equation modeling, and hybrid models (e.g., neuro - fuzzy and genetic - fuzzy) should be applied in travel demand analysis. In these approaches, the inclusion of land use variables may improve model's explanatory power. Also, different algorithms can be tested in soft computing methods. For example, the application of neural networks in travel demand analysis use generally the feed-forward back propagation algorithm. New algorithms may be developed to better understand mode choice behavior.

REFERENCES

Abane, Albert M. 2010. Travel behaviour in Ghana: empirical observations from four metropolitan areas. *Journal of Transport Geography* (Article in press).

Abdelwahab W. and Sayed T. 1999. Freight mode choice models using artificial neural networks. *Civil Engineering and Environmental Systems* 16(4): 267-286.

Abellan J., Gomez-Olmedo, M. and Moral, S. 2006. Some Variations on the PC Algorithm. *Proceedings of the Third European Workshop on Probabilistic Graphical Models*, Prague, Czech Republic.

Agresti, Alan. 2002. *Categorical data analysis.* John Wiley & Sons, Inc., New Jersey.

Aktaş, E., Ülengin, F., Şahin, Ş. Ö. 2007. A decision support system to improve the efficiency of resource allocation in healthcare management. *Socio - Economic Planning Science* 41(2): 130-146.

Allison, P.D. 1999. *Logistic regression using the SAS system – theory and application*. SAS Institute, Inc. Cary, NC.

Alonso, W. 1964. *Location and land use*. Cambridge: Harvard University Press.

Alpaydın, E. 2004. *Introduction to machine learning.* MIT Press, Cambridge, MA.

Alpizar, F. And Carlsson, F. 2003. Policy implications and analysis of the determinants of travel mode choice: an application of choice experiments to metropolitan Costa Rica. *Environment and Development Economics* 8(4): 603-619.

Alpkokin, P., Hayashi, Y., Black, J., and Gercek, H. 2005. Polycentric employment growth and impacts on urban commuting patterns: Case study of Istanbul. *Journal of Eastern Asia Society for Transportation Studies* 6: 3835-3850.

Andrade, K., Uchida, K., and Kagaya, S. 2006. Development of transport mode choice model by using adaptive neuro - fuzzy inference system. *Journal of the Transportation Research Board TRB* 1977: 8-16.

Avineri, E. 2005. *Soft computing applications in traffic and transport systems: A review.* In: Hoffmann, F., Koppen, M., Klawonn, F. and Roy, R. (Eds.), Soft computing: Methodologies and applications. Springer Series on Studies in Fuzziness and Soft Computing. Springer-Verlag, Germany.

Baker, J. Basu, R., Crapper, M., Lall, S. And Takeuchi, A. 2005. *Urban poverty and transport: the case of Mumbai*. World Bank Policy Research Working Paper 3693.

Ben – Akiva, M. 1973. *Structure of passenger travel demand models*. Ph.D. Dissertation. Department of Civil Engineering, MIT, Cambridge.

Ben – Akiva, M. and Lerman, S.R. 1985. *Discrete choice analysis: Theory and application to travel demand*. Cambridge MA: MIT Press.

Ben-Akiva, M. and Bierlaire, M. 1999. *Discrete choice methods and their applications to short-term travel decisions*. in R. Hall (ed.), Handbook of Transportation Science. International Series in Operations Research and Management Science, 23.

Bhat, C.R. and Gossen, R. 2004. A mixed multinomial logit model analysis of weekend recreational episode type choice. *Transportation Research Part B* 38(9): 767-787.

Bhat, C.R. and J.Y. Guo. 2007. A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. *Transportation Research Part B* 41(5): 506-526.

Boarnet, G. M. and Crane, R. 2001. *Travel by design: The influence of urban form on travel*. New York, Oxford University Press.

Bonnel, P. 2003. The estimation of aggregate modal split models, *European Transport Conference*. Strasbourg, France.

Bouckaert, R. R. 2008. *Bayesian network classifiers in Weka for version 3 – 57.* The University of Waikato.

Bowman, J. and Ben – Akiva, M. 1997. Activity - based travel forecasting. *Activity – Based Travel Forecasting Conference*. In Texas Transportation Institute, Washington D.C.

Bromley, J., Jackson, N.A., Clymer, O.J., Giacomello., A.M., Jensen F.V. 2005. The use of Hugin to develop Bayesian networks as an aid to integrated water resource planning. *Environmental Modelling & Software* 20(2): 231-242.

Buchanan, N., Barnett, R., Kingham, S., and Johnston, D. 2006. The effect of urban growth on commuting patterns in Christchurch. New Zealand. *Journal of Transport Geography* 14(5): 342-354.

Cambridge Systematics. 1994. The effects of land use and travel demand management strategies on commuting behavior. US Department of Transportation, US Environmental Protection Agency, US Department of Energy, Washington, D.C. Report prepared for the Travel Model Improvement Program.

Cantarella, G.E. and De Luca, S. 2005. Multilayer feedforward networks for transportation mode choice analysis: An analysis and a comparison with random utility models. *Transportation Research Part C: Emerging Technologies* 13(2): 121-155.

Celik, H. M. and Yankaya, U. 2006. The impact of rail transit investment on the residential property values in developing countries: The case of Izmir subway, Turkey. *Property Management* 24(4): 369-382.

Celikoglu, H. B. 2006. Application of radial basis function and generalized regression neural networks in non-linear utility function specification for travel mode choice modelling. *Mathematical and Computer Modelling* 44(7-8): 640-658.

Celikoglu, H. B. 2006. Calibration of logit modal split models with feed-forward back-propagation neural networks. *16th Mini EURO Conference: Artificial Intelligence in Transportation*, Poznan University of Technology, Poland.

Cervero, R. 1989. Jobs - housing balancing and regional mobility. *Journal of the American Planning Association* 55(2): 136-150.

Cervero, R. 1991. Jobs - housing balance as public policy. *Urban Land* 50(10): 10-14.

Cervero, R. 1996. Jobs - housing balance revisited: trends and impacts in the San Francisco Bay Area. *Journal of the American Planning Association* 62(4): 492-511.

Cervero, R. 2002. Built environment and mode choice: Toward a normative framework. *Transportation Research Part D* 7: 265-284.

Cervero, R. and Duncan, M. 2002. Residential self selection and rail commuting: A nested logit analysis. Working Paper, University of California Transportation Center, Berkeley California.

Cervero, R. and Gorham, R. 1995. Commuting in transit versus automobile neighborhoods. *Journal of the American Planning Association* 61(2): 210-225.

Cervero, R. and Kockelman, K. 1997. Travel demand and The 3Ds: Density, diversity, and design. *Transportation Research Part D* 2(3): 199-219.

Cervero, R. and Wu, K. 1997. *Influences of land use environments onj commuting choices: An analysis of large U.S. metropolitan areas using the 1985 American Housing Survey.* Working Paper (683), University of California at Berkeley.

Chang, Hsin – Li and Wu, Shun – Cheng. 2008. Exploring the vehicle dependence behind mode choice: Evidence of motorcycle dependence in Taipei. *Transportation Research Part A* 42: 307-320.

Chen C., and McKnight, C. E. 2007. Does the built environment make a difference? Additional evidence from the daily activity and travel behavior of homemakers living in New York City and suburbs. *Journal of Transport Geography* 15(5): 380-395.

Chen, C., Gong, H. and Paoswell, R. 2008. Role of the built environment on mode choice decisions: Additional evidence on the impact of density. *Transportation,* 35: 285 – 299.

Cheng, S. and Long, J. S. 2007. Testing for IIA in the multinomial logit model. *Sociological Methods & Research* 35(4): 583–600.

Coevering, P. Van De and Shwanen, T. 2006. Re-evaluating the impact of urban form on travel patterns in Europe and North America. *Transport Policy* 13(3): 229-239.

Collins, D. and Kearns, R. 2001. The safe journeys of an enterprising school: Negotiating landscapes of opportunity and risk. *Health and Place* 7(4)**:** 293-306.

Commins, N. And Nolan, A. 2010. The determinants of mode of transport to work in the Greater Dublin Area. *Transport Policy* (Article in press).

Cooper, G. F., Herskovitz, E. 1992. A bayesian method for the induction of probabilistic networks from data. *Machine Learning* 9(4): 309-347.

Cornalba, C. and Giudici, P. 2004. Statistical models for operational risk management. *Physica A* 338: 166–172.

Cramer, J. S. 2003. *Logit models from economics and other fields*. Cambridge University Press.

Crane, R. 2000. The influence of urban form on travel: on interpretive review. *Journal of Planning Literature* 15(1): 3-23.

Crane, R. and Crepeau, R. 1998. Does neighborhood design influence travel? A behavioral analysis of travel diary and GIS data. *Transportation Research Part D* 3(4): 225-238.

Dargay, J. and Gately, D. 1999. Income's effect on car and vehicle ownership, worldwide: 1960 – 2015. *Transportation Research Part A* 33: 101-138.

De Palma, A. and Rochat, D. 2000. Mode choice for trips to work in Geneva: An empirical analysis. *Journal of Transport Geography* 8(1): 43-51.

Demir, Y. K. and Gerçek, H. 2006. Ulaştırma türü seçiminde esnek hesaplama yöntemleri. *İTÜ Dergisi D* 5(6): 61-73.

Dirican, A. 2001. Evaluation of the diagnostic test's performance and their comparisons. *Cerrahpaşa Journal of Medicine* 32 (1): 25-30.

Dlamini, W. D. 2010. A bayesian belief network analysis of factors influencing wildfire occurrence in Swaziland. *Environmental Modeling & Software* 25(2): 199-208.

Domencich, T.A. and McFadden D. 1975. *Urban travel demand: A behavioural analysis*. North Holland Publishing Company.

Estrella, A. 1998. A new measure of fit for equations with dichotomous dependent variables. *Journal of Business and Economic Statistics* 16: 198-205.

Evans, James R. 2007. *Statistics, data analysis & decision modeling*. 3rd edition, Pearson Prentice Hall New Jersey.

Ewing, R., Deanna, M. and Li, S. C. 1996. Land use impacts on trip generation rates. *Transportation Research Record* 1518: 1-6.

Ewing, R., Schroeer, W., and Greene, W. 2004. School location and travel analysis of factors affecting mode choice. *Transportation Research Record: Journal of the Transportation Research Board* 1895: 55-63.

Ewing, Reid and Cervero, R. 2001. Travel and the built environment: A synthesis. *Transportation Research Record* 1780: 87-114.

Fischer, M. M. and Gopal, S. 1994. Artificial neural networks: A new approach to modeling interregional telecommunication flows. *Journal of Regional Science* 34(4): 503-527.

Fotheringham, A. S. And Knudsen, D. C. 1987. *Goodness-of-fit statistics.* Geo Books, Norwich.

Frank L. and G. Pivo. 1994. Impacts of mixed use and density on utilization of three modes of travel: single-occupant vehicle, transit, and walking. *Transportation Research Record* 1466: 44–52.

Frank, L., Bradley, M., Kavage, S., Chapman, J., and L., T. K. 2007. Urban form, travel time, and cost relationships with our complexity and mode choice. *Transportation* 35(1): 37-54.

Gakenheimer, R. 1999. Urban mobility in the developing world. *Transportation Research Part A* 33: 671-689.

Gakenheimer, R. and Zegras, C. 2004. Drivers of travel demand in cities of the development world. *Appendix in Mobility 2030: Meeting The Challenges to Sustainability*, World Business Council for Sustainable Development, The Sustainable Mobility Project.

Gebeyehu, M. and Takano, S. 2007. Diagnostic evaluation of public transportation mode choice in Addis Ababa. *Journal of Public Transportation* 10(4): 27-50.

Gomez – Ibanez, J. A. 1991. A global view of automobile dependence. *Journal of American Planning Association* 57: 369-376.

Governorship of Istanbul. 2010. http://www.istanbul.gov.tr/Default.aspx?pid=33&cat=1 (accessed May 28, 2010).

Greene, W. H. 2003. *Econometric analysis.* 5th edition, Prentice Hall, New Jersey.

Greenwald, M. J. 2006. The relationship between land use and intrazonal trip making behaviors: Evidence and implications. *Transportation Research Part D* 11: 432 – 446.

Gujarati, D. N. 1995. *Basic econometrics.* 3rd edition, McGraw-Hill, Inc.

Gwilliam, K. 2003. Urban transport in developing countries. *Transport Reviews* 23(2): 197-216.

Hamed, M. H. And Olaywah, H. H. 2000. Travel - related decisions by bus, servis taxi, and private car commuters in the city of Amman, Jordon. *Cities* 17(1): 63-71.

Hanson, S. 2004. *The context of urban travel: Concepts and recent trends*. in The Geography of Urban Transportation, 3rd edition. S. Hanson and G. Giuliano (eds). The Guilford Press, New York.

Hausman, J. and McFadden, D. 1984. Specification tests for the multinomial logit model. *Econometrica* 52: 1219-1240.

Heckerman, D. 1996. *A tutorial on learning with bayesian networks*. Technical Report, Microsoft Research.

Hensher D. A. and Button, K. 2000. *Handbook of transport modelling*. First edition, Handbooks in Transport Volume 1, Elsevier Science, New York.

Hensher D. A. and Button, K., Haynes, K. E., Stopher, P. 2004. *Handbook of transport geography and spatial systems*. Handbooks in Transport Volume 5, Elsevier Science, New York.

Hensher, D.A. and Ton, T. T. 2000. A comparison of the predictive potential of artificial neural networks and nested logit models for commuter mode choice. *Transportation Research Part E: Logistics and Transportation Review* 36(3): 155-172.

Hensher, D.A., Rose, J.M., and Grene, W.H. 2005. *Applied choice analysis: A primer*. Cambridge, Cambridge University Press.

Hess. D. 2001. Effect of free parking on commuter mode choice-evidence from travel diary data. *Transportation Research Record* 1753: 35-42.

Hosmer, D. W. and S. Lemeshow. 2000. *Applied logistic regression*. 2nd edition, Wiley.

Hugin GUI Help. 2010. The Hugin Graphical User Interface Help Pages, www.hugin.com (accessed May 28, 2010).

Hunt, J. D., Kriger, D. S., and Miller, E. J. 2005. Current operational urban land use – transport modelling frameworks: A review. *Transport Reviews* 25(3): 329-376.

Hyodo, T., Montalbo, C. M., Fujiwara, A. and Soedhodho, S. 2005. Urban travel behavior characteristics of 13 cities based on household interview survey data. *Journal of The Eastern Asia Society for Transportation Studies* 6: 23-28.

IBB. 2005. 1/25000 Ölçekli nazın imar planı analitik etüd ihalesi – arazi kullanım raporu. İstanbul Büyükşehir Belediyesi - Şehir Planlama Müdürlüğü.

IBB. 2010. Istanbul Metropolitan Municipality, www.ibb.gov.tr (accessed May 28, 2010).

Jaimes, F., Farbiarz, J., Alvarez, D., and Martinez, C. 2005. Comparison between logistic regression and neural networks to predict death in patients with suspected Sepsis in the emergency room. *Critical Care* 9(2): R150 – R156.

Janssesns, D., Wets G., Brijs, T., Vanhoof, K., Arentze, T., Timmermans H. 2006. Integrating bayesian networks and decision trees in a sequential rule-based transportation model. *European Journal of Operational Research* 175(1): 16-34.

Jensen, F. V. 2001. *Bayesian networks and decision graphs*. Springer.

Jou, Rong-Chang, Hensher, D. A., and Liu, Yu-Hsin 2010. Urban commuters' mode-switching behaviour in Taipai, with an application of the bounded rationality principle. *Urban Studies* 47(3): 650-665.

Kahn, C.E., Roberts, L.M., Shaffer, K.A. and Haddawy, P. 1997. Construction of a bayesian network for mammographic diagnosis of breast cancer. *Computers in Biology and Medicine* 27(1): 19-29.

Kanafani, A. 1983. *Transportation demand analysis*. McGraw-Hill.

Kennedy, P. 1981. *A guide to econometrics*. The MIT Press, Cambridge, Massachusetts.

Kitamura, R., Mokhtarian, P. C., and Laidet, L. 1997. A micro analysis of land use and travel in five neighborhoods in the San Francisco Bay Area. *Transportation* 24(2): 125-158.

Kjaerulff, U., Madsen, A.L. 2008. *Bayesian networks and influence diagrams: A guide to construction analysis*. Springer.

Kockelman, K. M. 1997. Travel behavior as a function of accessibility, land use mixing, and land use balance: Evidences from the San Francisco Bay Area. Master Thesis, University of California, Berkeley.

Koppelman, F. S. and Bhat, C. 2006. A self instructing course in mode choice modeling: multinomial and nested logit models. U.S. Department of Transportation Federal Transit Administration.

Kutzbach, Mark J. 2009. Motorization in developing countries: Causes, consequences, and effectiveness of policy options. *Journal of Urban Economics* 65: 154-166.

Lai, Wen-Tai and Lu, Jin-Long. 2007. Modeling the working mode choice, ownership and usage of car and motorcycle in Taiwan. *Journal of The Eastern Asia Society for Transportation Studies* 7: 869-885.

Larranaga, P., Poza, M., Yurramendi, Y., Murga R.H., Kuijpers, C.M.H. 1996. Structure learning of bayesian networks by genetic algorithms: A performance analysis of control parameters. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18(9).

Lau, J. C. Y. and Chiu, C. C. H. 2004. Accessibility of workers in a compact city: the case of Hong Kong. *Habitat International* 28: 89-102.

Lauritzen, S. L. 1995. The EM algorithm for graphical association models with missing data. *Computational Statistics & Data Analysis* 19(2): 191-201.

Lee, Sun Mi and Abbott, P.A. 2003. Bayesian networks for knowledge discovery in large data sets: basics for nurse researchers. *Journal of Biomedical Informatics* (36): 389-399.

Levinson, D. and Kumar, A. 1995. A multimodal trip distribution model: Structure and application. *Transportation Research Record* 1466: 124-131.

Limtanakool, N., Dijst, M., and Schwanen, T. 2006. The Influence of socioeconomic characteristics, land use and travel time considerations on mode choice for medium- and longer - distance trips. *Journal of Transport Geography* 14(5): 327-341.

Lin, J. and Long, L. 2008. What neighborhood are you in? empirical findings of relationships between household travel and neighborhood characteristics. *Transportation* 35(6): 739-758.

Lin, Jen-Jia and Yang, An-Tsei 2009. Structural analysis of how urban form impacts travel demand: evidence from Taipei. *Urban Studies* 46 (9): 1951-1967.

Liu, G. 2006. *A behavioral model of work – trip mode choice in Shanghai*. Discussion Papers. No: 444, Statistics Norway, Research Departmant.

Lowry, I. S. 1964. *A model of a metropolis*. Technical Memorandum, RM – 4035 RC, The Rand Corporation, California.

Luce, R. 1959. *Individual choice behavior: A theoretical analysis*. Wiley, New York.

Lynch, S. M. 2007. *Introduction to applied bayesian statistics and estimation for social scientists*. Springer.

Maat, K., Van Wee B., and Stead, D. 2005. Land Use and travel behaviour: expected effects from the perspective of utility theory and activity-based theories. *Environment and Planning B: Planning and Design* 32(1): 33-46.

Madsen, Anders L., Lang, Michael, Kjærulff, Uffe B., Jensen, Frank. 2003. *The hugin tool for Learning bayesian networks*. Springer.

Manski. C. 1973. *The analysis of qualitative choice*. Ph.D. Dissertation, Department of Economics, MIT, Cambridge.

Marcot, B.G., Steventon, J.D., Sutherland, G.D., McCann, R.K. 2006. Guidelines for developing and updating bayesian belief networks applied to ecological modeling and conservation. *Canadian Journal of Forest Research* 36(12): 3063–3074.

Marschak, J. 1960. *Binary choice constraints on random utility indicators*, In Stanford Symposium on Mathematical Methods in the Social Sciences, Stanford University Press, Stanford, Calif.

McArthur, D. P., Kleppe, G., Thorsen, I., and Uboe, J. 2010. The spatial transferability of parameters in a gravity model of commuting flows. *Journal of Transport Geography* (Article in press).

McFadden, D. 1974. *Conditional Logit Analysis of Qualitative Choice Behavior*. In Frontiers in Econometrics, Academic Press, New York.

Meyer, M.D. and Miller, E. J. 2001. *Urban transportation planning: A decision-oriented approach*. McGraw-Hill, A.B.D.

Mills, E. S. 1972. *Studies in the structure of the urban economy*. Baltimore: Johns – Hopkins.

Muth, R. F. 1969. *Cities and housing*. Chicago: Chicago Univesity Press.

Nadkarni, S., Shenoy, P. P. 2001. A bayesian network approach to making inferences in causal maps. *European Journal of Operational Research* 128(3): 479-498.

National Academy of Sciences. 2009. http://www.nationalacademies.org/annualreport/Report_to_Congress_2009.pdf (accessed May 28, 2010).

Neapolitan, R.E. 1990. *Probabilistic reasoning in expert systems: Theory and algorithms*. John Wiley and Sons, New York.

Neapolitan, R.E., Jiang, X. 2007. *Probabilistic methods for financial and marketing informatics*. Morgan Kaufman, San Mateo, CA.

Newman P W G, Kenworthy J. R. 1989. Gasoline consumption and cities. A comparison of US cities with a global survey. *Journal of the American Planning Association* 55(1)**:** 24-37.

Newman, P & Kenworthy, J. 1999. *Sustainability and cities: Overcoming automobile dependence*. Island Press, Washington D.C.

Nijkamp P., Reggiani, A. Tritapepe, T. 1996. Modelling inter-urban flows in Italy: A comparison between neural network approach and logit analysis. *Transportation Research Part C: Emerging Technologies* 4(6): 323–338.

Nijkamp, P., Reggiani, A., and Tritapepe, T. 1997. Spatial choice behaviour: Logit models and neural network analysis. *The Annals of Regional Science* 31(4): 411-429.

Nijkamp, P., Reggiani, A., and Trsang, W. F. 2004. Comparative modelling of interregional transport flows: Applications to multimodal european freight transport. *European Journal of Operational Research* 155(3): 584-602.

Norsys Software Corp. 2010. http://norsys.com (accessed May 28, 2010).

Nurdeen, A., Rahmat, R. A. and İsmail, A. 2007. Modeling of transportation behavior for coercive measures for car driving in Kuala Lumpur. *ARPN Journal of Engineering and Applied Sciences* 2(2): 18-23.

OD HH 2006. *Istanbul ulaşım ana planı hanehalkı araştırması 1. aşama analitik etüd ve model kalibrasyonu işi 2006 – 2007*. İstanbul Büyükşehir Belediyesi, Ulaşım Daire Başkanlığı – Ulaşım Planlama Müdürlüğü.

Oppenheim, N. 1995. *Urban travel demand modeling from individual choices to general equilibrium*. John Wiley & Sons, USA.

Ortuzar, J.D. and Willumsen, L.G. 1995. *Modelling transport*. John Wiley & Sons, England.

Pacione, M. 2009. *Urban geography: A global perspective.* 3rd Edition, Routledge, New York.

Pan, H., Shen, Q., and Zhang, M. 2009. Influence of urban form on travel behaviour in four neighborhoods of Shangai. *Urban Studies* 46(2): 275-294.

Papacostas, C.S. and P.D. Prevendouros 1993. *Transportation engineering and planning*. Prentice-Hall, Inc., New Jersey.

Pearl, J. 1988. *Probabilistic reasoning in intelligent systems: networks of plausible inference.* Morgan Kaufmann Publishers.

Peng, Z. 1997. The jobs - housing balance and urban commuting. *Urban Studies* 34(8): 1215-1235.

Phibanchon, S., Abdulkareem, S., Zain, R., and Abidin, B. 2007. An adaptive fuzzy regression model for the prediction of dichotomous response variables. The 2007 International Conference Computational Science and its Applications. Malaysia.

Phibanchon, S., R. M., Karem, S. A., Zain, R., and Abidin, B. 2007. An adaptive fuzzy regression model for the prediction of dichotomous response variables. The 2007 *International Conference Computational Science and its Applications*, Kuala Lumpur Malaysia.

Pickrell, Don. 1999. Transportation and land use. in Gomez-Ibanez, William B. Tye, and Clifford Winston, Editors, *Essays in Transportation Economics and Policy–A Handbook in Honor of John R. Meyer.* Washington, D.C.: Brookings Institution.

Pinjari, A.R., Bhat, C.R., Pendyala, R.M., and Waddell, P.A. 2007. Modeling residential sorting effects to understand the impact of the built environment on commute mode choice. *Transportation* 34(5): 557 - 573.

Plane, David A. 1981. The geography of urban commuting fields some empirical evidence from New England. *The Professional Geographer* 33(2): 182-188.

Pucher, J. 2004. *The geography of urban transportation.* 3rd Edition, The Guilford Press.

Pucher, J. and Renne, J. 2003. Socieconomics of urban travel : Evidence from the 2001 NHTS. *Transportation Quarterly* 57(3): 49-77.

Rajamani J., C. Bhat, and S. Handy, Y. Song. 2003. Assessing the impact of urban form measure on nonwork trip mode choice after controlling for demographic and level-of- service effects. *Transportation Research Record: Journal of the Transportation Research Board* 1831: 158-165.

Ramanathan, Ramu 1998. *Introductory econometrics with applications*. Fourth Edition, The Dryden Press.

Reilly, M. and Landis, J. 2002. The influence of built-form and land use on mode choice: evidence from the 1996 Bay Area Travel Survey. University of California Transportation Center Research Paper, Berkeley.

Robinson, R. W. 1977. *Counting unlabelled acyclic diagraphs*. Lecture Notes in Mathematics: Combinatorial Mathematics V, Springer-erlag.

Rodriguez, D.A., Khattak, A.J., Evenson, K.R. 2006. Can new urbanism encourage physical activity? comparing a new urbanist neighborhood with conventional suburbs. *Journal of the American Planning Association* 72(1): 43-54.

Rouwendal, J. and Nijkamp, P. 2004. Living in two worlds: a review of home – to – work decisons. *Growth and Change* 35(3): 287-303.

Santoso, Djoen San and Tsunokawa, Koji. 2005. Patial transferability and updating analysis of mode choice models in developing countries. *Transportation Planning and Technology* 28(5): 341-358.

SAS 2004. *SAS/ETS 9.1 User's Guide*. SAS Institute Inc., Cary/NC, USA.

SAS User Guide 2008. *SAS / ETS 9.2 User's Guide*. Chapter 15: The MDC Procedure, SAS Publishing.

Sayed, T. and Razavi, A. 2000. Comparison of neural and conventional approaches to mode choice analysis. *Journal of Computing in Civil Engineering* 14(1): 23-30.

SCAG. 2001. The new economy and jobs – housing balance in Southern California, Los Angeles.

Schwanen T. and Mokhtarian P L, 2005. What affects commute mode choice: Neighborhood physical structure or preferences toward neighborhoods? *Journal of Transport Geography* 13(1): 83-99.

Schwanen, T., Dieleman, F.M., Dijst, M. 2001. Travel behaviour in Dutch monocentric and policentric urban systems. *Journal of Transport Geography* 9(3): 173-186.

Schwanen, T., Dijst, M., and Dieleman, F. M. 2004. Policies for urban form and their Impact on travel: The Netherlands experience. *Urban Studies* 41(3): 579-603.

Schwanen, T., Dijst, M., and Dieleman, F. M. 2004. The impacts of metropolitan structure on commute behavior in the Netherlands: A multilevel approach, *Growth and Change* 35(3): 304-333.

Scuderi, Marco G., Clifton. Kelly J. 2005. Bayesian approaches to learning from data: using NHTS for the analysis of land use and travel behavior. *Journal of Transportation and Statistics* 8(3): 25-40.

Senbil, M. Fujiwara, A., Zhang, J. and Asri, D. U. 2005. Development of a choice model for evaluating sustainable urban form in a developing city. *Proceedings of The Eastern Asia Society for Transportation Studies* 5: 2164-2178.

Senbil, M., Kitamura, R., and Mohamad, J. 2009. Residential location, vehicle ownership and travel in Asia: A comparative analysis of Kei-Han-Shin and Kuala Lumpur metropolitan areas. *Transportation* 36(3): 325-350.

Senbil, M., Zhang, J. and Fujiwara, A. 2007. Motorization in Asia: 14 countries and three metropolitan areas. *IATSS Research* 31(1): 46-58.

Senbil, M., Zhang, J., and Fujiwara, A. 2006. *Motocycle ownership and use in Jabotek (Indonesia) metropolitan area*. Discussion Paper Series 2006-3, Hiroshima University.

Shaughnessy, P. and Livingston, G. R. 2005. *Evaluating the Causal Explanatory Value of Bayesian Network Structure Learning Algorithms*. Technical Report 2005-017, Computer Science Department, University of Massachusetts Lowell.

Snellen D, Borgers A, Timmermans H. 2002. Urban form, road network type, and mode choice for frequently conducted activities: a multilevel analysis using quasi-experimental design data. *Environment and Planning A* 34(7): 1207-1220.

Sohn, J. 2005. Are commuting patterns a good indicator of urban spatial structure? *Journal of Transport Geography* 13(4): 306-317.

Spiegelhalter, D.J., Dawid, A.P., Lauritzen, S.L. and Cowell, R.G. 1993. Bayesian analysis in expert systems. *Statistical Science* 8: 219-283.

Spirtes, P. and Glymour, C. 1991. An algorithm for fast recovery of sparse causal graphs. *Social Science Computing Review* 9(1): 62-72.

Srinivasan, K. K., Pradhan, G. N., and Naidu, G. M. 2007. Commute mode choice in a developing country: role of subjective factors, and variations in responsiveness across captive, semicaptive and choice segments in commute mode choice in a developing country. *Transportation Research Record: Journal of the Transportation Research Board* 2038: 53-61.

Srinivasan, S. 2002. Quantifying spatial characteristics of cities. *Urban Studies* 39(11): 2005-2028.

Srinivasan, S. and Rogers, P. 2005. Travel behavior of low – income residents: studying two contrasting locations in the city of Chennai, India. *Journal of Transport Geography* 13: 265-274.

Stead, D. 2001. Relationships between land use, socioeconomic factors, and travel patterns in Britain. *Environment and Planning B: Planning and Design* 28(4): 499-528.

Steck, H. 2001. Constrained - based structural learning in bayesian networks using finite data sets. Doctoral Thesis, Munich, Germany.

Steck, H., and Tresp, V. 1999. *Bayesian belief networks for data mining*. Proceedings of the Second Workshop Data Mining und Data Warehousing als Grundlage moderner enscheidungsunterstuetzender systeme.

Sultana, S. 2002. Job/housing imbalance and commuting time in the Atlanta metropolitan area: Exploration of causes of longer commuting time. *Urban Geography* 23(8): 728-749.

Taaffe, E. J., Gauthier, H.L., and O'Kelly, M.E., 1996. *Geography of transportation.* 2nd edition, Prentice Hall.

Tang, T. C. and Chi, Li C. 2005. Predicting multilateral trade credit risks: comparisons of logit and fuzzy logic models using ROC curve analysis. *Expert Systems wih Applications* 28(3): 547-556.

Timmermans, H. 2003. *The saga of integrated land use – transport modeling. how many more dreams before we make up?* 10 th International Conference on Travel Behaviour Research, Lucerne.

Torres, F. J., Huber, M. 2003. *Learning a Causal Model from Household Survey Data using a Bayesian belief network*. Presented at the annual meetings of the transportation research board, Washington D.C.

Tortum, A., Yayla, N., and Gökdağ, M. 2008. The modeling of mode choices of intercity freight transportation with the artificial neural networks and adaptive neuro-fuzzy inference system. *Expert Systems and Applications* 36(3): 6199-6217.

Train, K. 2003. *Discrete choice methods with simulation*. Massachusetts: Cambridge University Press.

TRB Special Report 282. 2005. Does the built environment influence physical activity? Examining the evidence. *Transportation Research Board*. Institute of Medicine of the National Academics, Washington D.C.

Türe, M., İ. Kurt, E. Yavuz ve T. Kürüm 2005. Hipertansiyonun Tahmini için Çoklu Tahmin Modellerinin Karşılaştırılması (Sinir Ağları, Lojistik Regresyon ve Esnek Ayırma Analizleri). *Anadolu Kardiyoloji Dergisi* 5(1): 24-28.

Urbanrail.net. 2010. http://www.urbanrail.net/index.html (accessed May 28, 2010).

Vanderlaan, L. 1998. Changing urban systems: an empirical analysis at two spatial levels. *Regional Studies* 32(3): 235-247.

Vasconcellos, E. A. 2001. *Urban transport, environment, and equity: the case for developing countries*. Earthscan, London.

Vasconcellos, E. A. 2005. Urban change, mobility and transport in Sao Paulo: three decades, three cities. *Transport Policy* 12: 91-104.

Vega, A. and Reynolds-Feighan, A. 2006. Choice of residential location and travel mode to work in the greater dublin area, presented to UTSG.

Vega, A. and Reynolds-Feighan, A. 2008. Employment sub - centres and travel to work mode choice in the Dublin Region. *Urban Studies* 45(9): 1747-1768.

VTPI. 2010. Victoria Transport Policy Institute. http://www.vtpi.org/tdm/tdm80.htm (accessed September 28, 2010).

Vythoulkas, P.C. and Koutsopoulos, H.N. 2003. Modeling discrete choice behavior using concepts from fuzzy set theory, approximate reasoning and neural networks. *Transportation Research Part C Emerging Technologies* 11(1): 51-73.

Wan, Xia., Chen, J. and Zheng, M. 2009. The impact of land use and traffic resources supply on commute mode choice. 2009 Second International Conference on Intelligent Computation Technology and Automation, Changsha, Hunan.

Wang, D. and Chai, Y. 2009. The jobs - housing relationship and commuting in Beijing, China: The legacy of Danwei. *Journal of Transport Geography* 17(1): 30-38.

Witten, Ian and Frank, Eibe 2005. *Data mining: practical machine learning tools and techniques*. 2nd Edition, Morgan Kaufmann.

World Bank. 2002. *Cities on the move: a world bank urban transport strategy review*. Washington, D.C.

Wright, Paul H. and Ashford, Norman J. 1989. *Transportation engineering: planning and design*. 3rd ed., John Wiley & Sons, New York.

Xie, C., Lu, J., and Parkany, E. 2003. Work travel mode choice modeling with data mining: Decision trees and neural networks. *Transportation Research Recor*d 1854: 50-61.

Yankaya, U. 2004. Modeling the impacts of Izmir subway on the values of residential property using hedonic price model, Unpublished Master of Thesis, Izmir Institute of Technology, Izmir.

Zegras, C. 2010. The built environment and motor vehicle ownership and use: evidence from Santiago de Chile. *Urban Studies* 47(8): 1793-1817.

Zhang M. 2004. The Role of land use in travel mode choice: evidence from Boston and Hong Kong. *Journal of American Planning Association* 70(3): 344-361.

Zhao, P. 2010. Sustainable urban expansion and transportation in a growing megacity: consequences of urban sprawl for mobility on the urban fringe of Beijing. *Habitat International* 34: 236-243.

Zhao, P., Lü, Bin., and Roo, Gert de. 2010. Impact of the jobs – housing balance on urban commuting in Beijing in the transformation era. *Journal of Transport Geography* (Article in press).

APPENDIX A

THE SOFTWARE CODES TO ESTIMATE THE MULTINOMIAL LOGIT MODELS

Nlogit 4.0 program is used to estimate a greater range of models for discrete choice such as multinomial logit, nested logit, and multinomial probit. The codes listed below only include base model and Model 1 that are the part of disaggregate discrete choice models. These models can be estimated by different programs such as SAS and Matlab. The software packages produce essentially the same results but in different formats.

Base Model:

```
NLOGIT
;lhs = choice, cset, altij
;Choices = walk, transit, car
;Model:
U(walk) = ascwalk /
U(transit) = asctr /
U(car) = asccar $
```

or;

```
NLOGIT ; Lhs = mode; Choices = walk, transit, car, service
; Rh2 = one $
```

Model 1:

```
NLOGIT ; Lhs = mode; Choices = walk, transit, car, service
; Rhs = tcost,ttime
; Rh2 = one $
```

APPENDIX B

TRANSPORTATION MASTER PLAN STUDIES

Table B. 1. Transportation Master Plan Studies in Istanbul in The Last Thirty Years (Source: OD HH 2006)

Çalışma	**IRTC**	**İBUNP**	**İUAP**	**İUAP 2007**
Yapan Kuruluş	IRTC Konsorsiyum	Temel Müh. A.Ş.	İTÜ - İstanbul Büyükşehir Belediyesi	İstanbul Büyükşehir Belediyesi
Yapılış Yılı	1985	1987	1997	2007
Model	TRANSPLAN	TRANPLAN/ TRANSPORT	TRANPLAN	TRANSCAD
Çalışma Alanı (Ha)	97.637	86.962	154.733	539.000
Bölge Sayısı	97	108	209	451
Ulaştırma Ağındaki Bağlantı Sayısı				
Karayolu	1.835	2.200	5.323	15.586
Toplu Taşıma	1.544	4.000	6.423	10.338
Çalışma Alanı Nüfusu	5.784.160	5.760.000	9.057.747	12.006.999
Doğu Yakası	1.917.000 **(%32)**	1.850.000 **(%35)**	3.170.211 **(%35)**	4.422.418 **(%37)**
Batı Yakası	3.867.160 **(%68)**	3.910.000 **(%65)**	5.887.536 **(%65)**	7.584.581 **(%63)**
İstihdam	1.924.000	2.035.000	2.532.211	3.957.336
Doğu Yakası	446.800 **(%22.5)**	457.800 **(%27)**	676.738 **(%26)**	1.179.884 **(%30)**
Batı Yakası	1.477.200 **(%77.5)**	1.577.200 **(%73)**	1.885.473 **(%74)**	2.777.452 **(%70)**

(cont. on next page)

Table B. 1. (cont.)

Çalışma	**IRTC**	**İBUNP**	**İUAP**	**İUAP 2007**
Yapan Kuruluş	IRTC Konsorsiyum	Temel Müh. A.Ş.	İTÜ - İstanbul Büyükşehir Belediyesi	İstanbul Büyükşehir Belediyesi
Kişi	4.779	9.456	37.843	263.768
Örneklem Oranı (%)	**0,08**	**0,16**	**0,42**	**2,2**
Gelir Grupları (YTL)				
Düşük (0 -1000)[20]	% 24.4	% 28.5	% 27.1	% 69.0
Orta (1000 -2000)	% 61.5	% 63.4	% 65.4	% 23.5.0
Yüksek (2000+)	% 14.1	% 8.1	% 7.5	% 7.5.0
Trafiğe Kayıtlı Özel Otomobil Sayısı	297.693	375.200	889.342	1.522.521
Otomobil Sahipliği (1/1000 Kişi)				
Düşük	10	4	7	71
Orta	52	73	79	143
Yüksek	125	283	208	277
Ortalama	51	71	76	103 / 111
Kişi Başına Ortalama Hareketlilik				
Motorlu Araçlarla	0,69	0,87	1,00	0,88
Yaya Dahil	1,03	1,44	1,54	1,74
Yaya Yolculuk Oranı (%)	33	40	35	49,3

(cont. on next page)

[20] 2007 classification

Table B. 1. (cont.)

Çalışma	**IRTC**	**İBUNP**	**İUAP**	**İUAP 2007**
Yapan Kuruluş	IRTC Konsorsiyum	Temel Müh. A.Ş.	İTÜ - İstanbul Büyükşehir Belediyesi	İstanbul Büyükşehir Belediyesi
Ortalama Yolculuk Uzunluğu (Dakika)	46	52,8	41	48,8
Ev - İş	48,5	55,6	43	51,9
Ev - Okul	46,3	50,9	37,4	48,5
Ev - Diğer	43,4	51,2	42	49,8
Ev Uçlu Olmayan	36,7	44,6	34	52,0
Yolculuk Amaçları (%)				
Ev - İş	60	53	55,0	32,3
Ev - Okul	9	16	14,5	21,4
Ev - Diğer	20	19	18,3	37,2
Ev Uçlu Olmayan	11	12	12,2	9,1
Toplam	100	100	100	100
Türel Dağılım (%)				
Özel Taşıma	32,5	30	40	29
Toplu Taşıma	67,5	70	60	71

APPENDIX C

LAND USE FORMULATIONS

Ratio 1: Employment / Population Density (epdens)

$$\frac{Employment}{Population}$$

Ratio 2: Worker / Population Density

$$\frac{NumberofEmployees}{Population}$$

Ratio 3: Employment / Worker Density as a jobs – housing balance ratio (ewdens)

$$\frac{Employment}{NumberofEmployees}$$

Ratio 4: Employment Area Density

$$\frac{Wholesaletrade + industry + retail + retail \,\&\, workingareas}{housingareas}$$

Ratio 5: Population Density (pdens)

$$\frac{Population}{ZonalArea}$$

Ratio 6: Built Up Area Population Density

$$\frac{Population}{BuiltUpArea}$$

Built Up Urban Areas includes urban services, transportation, housing, wholesale trade, retail, industry, tourism, working areas, green and sport areas, military, storage, and health.

Ratio 7: Built Up Area Population Density 2

$$\frac{Population}{BuiltUpArea}$$

Urban Areas includes urban services, transportation, housing, wholesale trade, retail, industry, tourism, working areas, military, storage, and health **.**

Ratio 8: Overall Density (Gross Density)

$$\frac{Population + Employment}{ZonalArea}$$

$$\text{or} \quad \frac{residents + jobs}{ZonalArea}$$

Ratio 9: Gross Density 2

$$\frac{Population + Employment}{BuiltUpAreas}$$

Ratio 10: Job Density

$$\frac{Employment}{BuiltUpAreas}$$

Ratio 11: Worker Density

$$\frac{NumberofEmployees}{BuiltUpAreas}$$

Ratio 12: Road Density Index

Linear roads length over sqaure areas. In other words, a ratio of the total road length within each TAZ over the total areas.

Ratio 13: Rail Road Ratio

$$\frac{Railroadlength}{Roadlength}$$

Ratio 14: Transit Access for each zones (1: Available; 0: Not Available)

Ratio 15: Jobs – Housing Balance 1 (jhb)

$$\left|\frac{(E_i - c \times P_i)}{(E_i + c \times P_i)}\right|$$

Ratio 16 Jobs – Housing Balance 2 (jhb)

$$1 - \left|\frac{(E_i - c \times P_i)}{(E_i + c \times P_i)}\right|$$

E_i = Employment Size,

P_i = Population Size,

C = Aktivite Oranı.

Ratio 17: Employed Residents to jobs balance index (jhb)

$$1 - \left\{\frac{|ER - JOBS|}{(ER + JOBS)}\right\}$$

ER = Number of employed residents,

JOBS = Number of workers.

Ratio 18: Land Use Mix Diversity Index (lumix)

$$\textbf{Land Use Mix Diversity} = 1 - \left| \frac{\left|\frac{r}{T} - \frac{1}{4}\right| + \left|\frac{c}{T} - \frac{1}{4}\right| + \left|\frac{i}{T} - \frac{1}{4}\right| + \left|\frac{o}{T} - \frac{1}{4}\right|}{\frac{3}{2}} \right|$$

T = r + c + i + o,

r: residential area,

c: commercial area,

i: industrial area,

o: other land uses.

Ratio 19: Total Employment Density in TAZ

$$\frac{employment}{area}$$

Ratio 20: Service Employment Density in TAZ

$$\frac{service\_employment}{area}$$

Ratio 21: Industrial Employment Density in TAZ (iedens)

$$\frac{industrial\_employment}{area}$$

Ratio 22: Commercial Employment Density in TAZ (cedens)

$$\frac{commercial\_employment}{area}$$

Ratio 23: Commercial Area Ratio (C_AREA_RATIO): The ratio of commercial and industrial area to total area of a zone (cidens).

$$\frac{commercial + industry}{zonalarea}$$

Ratio 24: Commercial Area Ratio 2 (cidens)

$$\frac{commercial + industry + urbanworkingareas}{BUILT\_UP\_AREA}$$

Ratio 25: Commercial Area Ratio 3 (cidens)

$$\frac{Commercial + Industry + Urbanworkingareas}{ZonalArea}$$

Ratio 26: Transit Length Density (in meters) of transit per acre.

$$\frac{Railroadlength}{Area}$$

Ratio 27: Employment Density in TAZ

$$\frac{Commercial + Industry + Urbanworkingareas}{\mathrm{Re}\,sidentialAreas}$$

Ratio 28: Industry Employment Density (1)

$$\frac{Industry}{Population}$$

Ratio 29: Employment Density in TAZ (2)

$$\frac{Industry + Commercial}{Population}$$

Ratio 30: Employment Density in TAZ (3)

$$\frac{Commercial + Industry + Urbanworkingareas}{Population}$$

APPENDIX D

QUESTIONNAIRE FORM USED IN 2006 (O/D BASED) ISTANBUL HOUSEHOLD TRAVEL SURVEY

İSTANBUL O/D ARAŞTIRMASI HANEHALKI SORUKAĞIDI (2. ETAP)

İyi günler efendim. Adım......................... Bildiğiniz gibi ulaşım, İstanbul'un en büyük problemlerinden biridir. İSTANBUL BÜYÜKŞEHİR BELEDİYESİ ulaşım probleminin çözülebilmesi için yeni projeler üretmek istemektedir. Bu amaçla İSTANBUL BÜYÜKŞEHİR BELEDİYESİ İSTANBUL METROPOLİTAN PLANLAMA MERKEZİ olarak ilerideki ulaşım planlarında da kullanılmak üzere "İSTANBUL ULAŞIM ANA PLANI HANEHALKI ARAŞTIRMASI" yapmaktayız. Sizinle tüm hanehalkı ve yolculuk bilgilerini içeren bir anket yapmak istiyoruz. Araştırma sonunda konut adresleri, İSTANBUL METROPOLİTAN PLANLAMA MERKEZİ'ne sizi arayarak anketin yapılıp yapılmadığının kontrol edilmesi amacıyla verilecektir. Bu anket ile derlenecek bilgiler istatistik amaçlı çalışmalar için kullanılacaktır, gizlidir. Anketimiz yaklaşık 20 dk sürecektir. Bu şartla bize yardımcı olursanız seviniriz. Şimdiden teşekkür ederiz.

ADRES VE ÖRNEKLEME BİLGİLERİ

A1. İL		A6. BİNA / DIŞKAPI NO		A11. EVTİPİ	
A2. İLÇE		A7. BİNA İÇ KAPI NO (DAİRE NO)		1. Apartman	
A3. BUCAK (BELDE)		A8. BÖLGE	1 2 3 4 5	2. Müstakil Ev	
A4. KÖY / MAHALLE		A9. KÜME		3. Site	
A5. CADDE/ SOKAK		A10. ÖRNEK HANE NO		4. Diğer (belirtiniz)	

ZİYARET SAYISI	1	2	3	4
ANKETÖRÜN ADI, SOYADI				
KOD NUMARASI				
TARİH	GÜN AY	GÜN AY	GÜN AY	GÜN AY
BAŞLAMA SAATİ (SAATE BAK VE YAZ)	SAAT DAK.	SAAT DAK.	SAAT DAK.	SAAT DAK.
ZİYARET SONUCU	HANEHALKI TAMAMLANDI 1 HANEHALKI YARIM KALDI 2 EVDE KİMSE YOK 3 GÖRÜŞME ERTELENDİ 4 GÖRÜŞME REDDEDİLDİ 5 DİĞER (Belirtin) X	HANEHALKI TAMAMLANDI 1 HANEHALKI YARIM KALDI 2 EVDE KİMSE YOK 3 GÖRÜŞME ERTELENDİ 4 GÖRÜŞME REDDEDİLDİ 5 DİĞER (Belirtin) X	HANEHALKI TAMAMLANDI 1 HANEHALKI YARIM KALDI 2 EVDE KİMSE YOK 3 GÖRÜŞME ERTELENDİ 4 GÖRÜŞME REDDEDİLDİ 5 DİĞER (Belirtin) X	HANEHALKI TAMAMLANDI 1 HANEHALKI YARIM KALDI 2 EVDE KİMSE YOK 3 GÖRÜŞME ERTELENDİ 4 GÖRÜŞME REDDEDİLDİ 5 DİĞER (Belirtin) X
SORU KAĞIDININ DOLDURULMASI TAMAMLANMAMIŞ İSE SONRAKİ GÖRÜŞMENİN RANDEVU GÜNÜ VE SAATİ	GÜN SAAT DAK.	GÜN SAAT DAK.	GÜN SAAT DAK.	GÜN SAAT DAK.

EKİP BAŞI (Alan Editörü)	ADI / SOYADI ____________	KOD	TELEFON KONTROLÜ YAPILDI
DENETÇİ (Telefon Kontrolü)	ADI / SOYADI ____________	KOD	

CEVAPLAYAN KİŞİ	ADI / SOYADI ____________ HH FERT SATIR NO		HANE İLE İLGİLİ BÜTÜN ÇALIŞMA BİTTİĞİNDE DOLDURULACAKTIR. DOLDURULMUŞ ANKET 1 · DİĞER 4 EKSİK ANKET 2 EKSİK AMA GEÇERLİ ANKET 3
EK HANE HALKI SORUKAĞIDI DOLDURULDU MU?	0 HAYIR · 1 EVET	KAÇ ADET?	

BÖLÜM I. HANEHALKI BİLGİLERİ

KİŞİ NO	B.1. Sürekli olarak bu hanede yaşayan kişilerin isim ve soy isimlerini hane reisinden başlayarak söyler misiniz? (ÖNCE İSİMLERİ ALINIZ. DAHA SONRA DİĞER BİLGİLERİ SATIR SATIR SORUNUZ!) (HANEHALKI ÜYELERİNİN İSİM VE SOYİSİMLERİNİ YAŞI BÜYÜK OLANDAN BAŞLAMAK ÜZERE YAZINIZ.)	B.2. Kaç yaşındasınız? (SADECE TAMAMLANMIŞ YAŞLARI KAYDEDİNİZ. 1 YAŞINDAN KÜÇÜK İSE 00 YAZINIZ.)	B.3. Cinsiyet		B.4. Hane halkı reisine yakınlığınız nedir?	EĞİTİM DURUMU (6 YAŞ VE ÜZERİNE SORULACAK) B.5. Okuma ve yazma Biliyor musunuz?		B.5.1. En son hangi okulu bitirdiniz ? (EN SON DİPLOMA ALINAN OKULU İŞARETLEYİNİZ.)										B.5.2. Hala öğrenci misiniz? (6 YAŞ VE ÜZERİ)		B.5.3. Hangi okula devam ediyorsunuz?									
			ERKEK	KADIN	1. Hanehalkı Reisi 2. Eşi 3. Oğlu 4. Kızı 5. Babası 6. Annesi 7. Kardeşi 8. Kayınpederi 9. Kayınvalidesi 10. Gelini 11. Damadı 12. Torunu 13. Yatılı Misafir 14. Diğer (Belirtiniz)	EVET	HAYIR (Geçiniz B.6.)	Bir okul bitirmemiş	İlkokul	İlköğretim	Ortaokul	Ortaokul dengi meslek okulu	Lise	Lise dengi meslek okulu	Yüksekokul	Fakülte	Yüksek Lisans / Doktora	EVET	HAYIR (Geçiniz B.6.)	Bir okul bitirmemiş	İlkokul	İlköğretim	Ortaokul	Ortaokul dengi meslek okulu	Lise	Lise dengi meslek okulu	Yüksekokul	Fakülte	Yüksek Lisans / Doktora
01			1	2		1	0	1	2	3	4	5	6	7	8	9	10	1	0			3	4	5	6	7	8	9	10
02			1	2		1	0	1	2	3	4	5	6	7	8	9	10	1	0			3	4	5	6	7	8	9	10
03			1	2		1	0	1	2	3	4	5	6	7	8	9	10	1	0			3	4	5	6	7	8	9	10
04			1	2		1	0	1	2	3	4	5	6	7	8	9	10	1	0			3	4	5	6	7	8	9	10
05			1	2		1	0	1	2	3	4	5	6	7	8	9	10	1	0			3	4	5	6	7	8	9	10
06			1	2		1	0	1	2	3	4	5	6	7	8	9	10	1	0			3	4	5	6	7	8	9	10
07			1	2		1	0	1	2	3	4	5	6	7	8	9	10	1	0			3	4	5	6	7	8	9	10
08			1	2		1	0	1	2	3	4	5	6	7	8	9	10	1	0			3	4	5	6	7	8	9	10
09			1	2		1	0	1	2	3	4	5	6	7	8	9	10	1	0			3	4	5	6	7	8	9	10
10			1	2		1	0	1	2	3	4	5	6	7	8	9	10	1	0			3	4	5	6	7	8	9	10
11			1	2		1	0	1	2	3	4	5	6	7	8	9	10	1	0			3	4	5	6	7	8	9	10
12			1	2		1	0	1	2	3	4	5	6	7	8	9	10	1	0			3	4	5	6	7	8	9	10
13			1	2		1	0	1	2	3	4	5	6	7	8	9	10	1	0			3	4	5	6	7	8	9	10
14			1	2		1	0	1	2	3	4	5	6	7	8	9	10	1	0			3	4	5	6	7	8	9	10
15			1	2		1	0	1	2	3	4	5	6	7	8	9	10	1	0			3	4	5	6	7	8	9	10
16			1	2		1	0	1	2	3	4	5	6	7	8	9	10	1	0			3	4	5	6	7	8	9	10
17			1	2		1	0	1	2	3	4	5	6	7	8	9	10	1	0			3	4	5	6	7	8	9	10
18			1	2		1	0	1	2	3	4	5	6	7	8	9	10	1	0			3	4	5	6	7	8	9	10
19			1	2		1	0	1	2	3	4	5	6	7	8	9	10	1	0			3	4	5	6	7	8	9	10
20			1	2		1	0	1	2	3	4	5	6	7	8	9	10	1	0			3	4	5	6	7	8	9	10
21			1	2		1	0	1	2	3	4	5	6	7	8	9	10	1	0			3	4	5	6	7	8	9	10
22			1	2		1	0	1	2	3	4	5	6	7	8	9	10	1	0			3	4	5	6	7	8	9	10
23			1	2		1	0	1	2	3	4	5	6	7	8	9	10	1	0			3	4	5	6	7	8	9	10
24			1	2		1	0	1	2	3	4	5	6	7	8	9	10	1	0			3	4	5	6	7	8	9	10
25			1	2		1	0	1	2	3	4	5	6	7	8	9	10	1	0			3	4	5	6	7	8	9	10

(GEREKİRSE EK FORM KULLANINIZ.)

B.11.1.Hanedeki toplam kişi sayısı.......................

B.11.2.Hanedeki 6 yaş ve üzeri kişi sayısı....................... (CEVAPLAYANA ONAYLATINIZ.)

BÖLÜM I. HANEHALKI BİLGİLERİ

ÇALIŞMA DURUMU (12 YAŞ VE ÜZERİNE SORULACAK) B.6. Geçen hafta içinde bir işte çalıştınız mı? (GEÇEN HAFTA İÇİNDE PARA VEYA MAL KARŞILIĞI BIR SAAT DAHİ ÇALIŞANLAR ÇALIŞTI OLARAK ALINACAKTIR. GEÇEN HAFTA İÇİNDE İZİN, HASTALIK, SEYAHAT VB. GİBİ NEDENLERLE VEYA MEVSİM GEREĞİ İŞİNDE ÇALIŞMAYANLAR İÇİN ÇALIŞMADI FAKAT İŞİ İLE İLGİSİ DEVAM EDİYOR İŞARETLENECEKTİR.)			B.6.1. Yaptığınız bu iş hangi sektöre girer? (İRDELEYİNİZ, İŞARETLEYİNİZ VE B.7'YE GEÇİNİZ) (ÇALIŞTIĞI SEKTÖR)										B.6.2. Çalışmama sebebiniz nedir? (ÇALIŞMAMA SEBEBİ)						B.7. Otomobil sürücü belgeniz var mı? (BUGÜN İÇİN GEÇERLİ OTOMOBİL SÜRÜCÜ BELGESİ. 18 YAŞ VE ÜZERİNE SORULACAK.)		B.7.A. Akbiliniz var mı? (HAYIR İSE B.10.'A GEÇİNİZ)		MAVİ KART KULLANIM DURUMU B.8. Şu anda sınırsız (GÜNLÜK, HAFTALIK, AYLIK) akbil kullanıyor musunuz?		B.9. Öğrenci, öğretmen, ve 60 yaş üzeri vb. indirimli kartınız var mı?		B.10. Özürlü, basın, yaşlılık maaşı alanların kullandığı vb. ücretsiz kartınız var mı?		B.11.KİŞİSEL SEYAHAT FORMU STATÜSÜ				
Çalıştı	Çalışmadı fakat işi ile ilgisi devam ediyor	Çalışmadı (Geçiniz B.6.2.)	Ziraat, Avcılık, Ormancılık ve Balıkçılık	Madencilik, Taş Ocakçılığı	İmalat Sanayi	Elektrik, Gaz, Su	İnşaat	Toptan ve perakende ticaret, Lokanta, Otel, Eğlence	Ulaştırma, Haberleşme, Depolama	Mali, Sigorta, Emlak	Toplum Hizmeti Sosyal, Resmi	İyi tanımlanmamış Faliyetler	Öğrenci	Hasta / Engelli	Emekli	İş arıyor	Ev hanımı	Diğer (Belirtiniz)	YOK	VAR	HAYIR	EVET	HAYIR	EVET	YOK	VAR	YOK	VAR	Dolduruldu	Eksik kaldı	Reddetti	Ulaşılamadı	Diğer (Belirtiniz)
1	2	3	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6 ____	0	1	0	1	0	1	0	1	0	1	1	2	3	4	5 ____
1	2	3	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6 ____	0	1	0	1	0	1	0	1	0	1	1	2	3	4	5 ____
1	2	3	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6 ____	0	1	0	1	0	1	0	1	0	1	1	2	3	4	5 ____
1	2	3	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6 ____	0	1	0	1	0	1	0	1	0	1	1	2	3	4	5 ____
1	2	3	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6 ____	0	1	0	1	0	1	0	1	0	1	1	2	3	4	5 ____
1	2	3	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6 ____	0	1	0	1	0	1	0	1	0	1	1	2	3	4	5 ____
1	2	3	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6 ____	0	1	0	1	0	1	0	1	0	1	1	2	3	4	5 ____
1	2	3	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6 ____	0	1	0	1	0	1	0	1	0	1	1	2	3	4	5 ____
1	2	3	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6 ____	0	1	0	1	0	1	0	1	0	1	1	2	3	4	5 ____
1	2	3	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6 ____	0	1	0	1	0	1	0	1	0	1	1	2	3	4	5 ____
1	2	3	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6 ____	0	1	0	1	0	1	0	1	0	1	1	2	3	4	5 ____
1	2	3	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6 ____	0	1	0	1	0	1	0	1	0	1	1	2	3	4	5 ____
1	2	3	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6 ____	0	1	0	1	0	1	0	1	0	1	1	2	3	4	5 ____
1	2	3	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6 ____	0	1	0	1	0	1	0	1	0	1	1	2	3	4	5 ____
1	2	3	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6 ____	0	1	0	1	0	1	0	1	0	1	1	2	3	4	5 ____
1	2	3	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6 ____	0	1	0	1	0	1	0	1	0	1	1	2	3	4	5 ____
1	2	3	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6 ____	0	1	0	1	0	1	0	1	0	1	1	2	3	4	5 ____
1	2	3	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6 ____	0	1	0	1	0	1	0	1	0	1	1	2	3	4	5 ____
1	2	3	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6 ____	0	1	0	1	0	1	0	1	0	1	1	2	3	4	5 ____
1	2	3	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6 ____	0	1	0	1	0	1	0	1	0	1	1	2	3	4	5 ____
1	2	3	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6 ____	0	1	0	1	0	1	0	1	0	1	1	2	3	4	5 ____
1	2	3	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6 ____	0	1	0	1	0	1	0	1	0	1	1	2	3	4	5 ____
1	2	3	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6 ____	0	1	0	1	0	1	0	1	0	1	1	2	3	4	5 ____
1	2	3	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6 ____	0	1	0	1	0	1	0	1	0	1	1	2	3	4	5 ____
1	2	3	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6 ____	0	1	0	1	0	1	0	1	0	1	1	2	3	4	5 ____

B.11.3.Kişisel yolculuk formu doldurulan kişi sayısı........................

BÖLÜM II. 24 SAATLİK SEYAHAT BİLGİLERİ

ANKET NO	Küme / Hane / Kişi	**Görüşme Tarihi**	Gün / Ay / Yıl: 2 0 0 6
ANKETÖR NO		**Hane Listesindeki Kişi No**	**Ziyaret sayısı**

C.0. 15 dakika içinde başladığı noktaya dönmeyen her harekete seyahat diyoruz.
Şimdi size dün sabah 06:00'dan bu sabah 06:00'ya kadar yaptığınız tüm; işe, eve, parka, arkadaş ziyaretine, kadınlar gününe, bakkala, yemeğe, sinemaya, fatura yatırmaya, düğüne, şehir dışına gibi, yaya ya da araçla yaptığınız yolculukların detaylı bilgilerini soracağım.
Dün sabah 06.00'dan bu sabah 06.00'ya kadar belirli bir amaç için evden dışarı çıktınız mı?

→

SEYAHAT ZİNCİRİ

C.0.1. Dün sabah 06,00 da nerdeydiniz?	C.0.2. Ordan nereye gittiniz?	C.0.3. Ordan nereye gittiniz?	C.0.4. Ordan nereye gittiniz?	C.0.5. Ordan nereye gittiniz?	C.0.6. Ordan nereye gittiniz?	C.0.7. Ordan nereye gittiniz?	C.0.8. Ordan nereye gittiniz?
1 Ev	1 Ev	1 Ev	1 Ev	1 Ev	1 Ev	1 Ev	1 Ev
2 İş	2 İş	2 İş	2 İş	2 İş	2 İş	2 İş	2 İş
3 Okul	3 Okul	3 Okul	3 Okul	3 Okul	3 Okul	3 Okul	3 Okul
4 İş takibi	4 İş takibi	4 İş takibi	4 İş takibi	4 İş takibi	4 İş takibi	4 İş takibi	4 İş takibi
5 Alışveriş	5 Alışveriş	5 Alışveriş	5 Alışveriş	5 Alışveriş	5 Alışveriş	5 Alışveriş	5 Alışveriş
6 Sosyal, spor ve eğlence	6 Sosyal, spor ve eğlence	6 Sosyal, spor ve eğlence	6 Sosyal, spor ve eğlence	6 Sosyal, spor ve eğlence	6 Sosyal, spor ve eğlence	6 Sosyal, spor ve eğlence	6 Sosyal, spor ve eğlence
7 Hastane, Sağlık Merkezi	7 Hastane, Sağlık Merkezi	7 Hastane, Sağlık Merkezi	7 Hastane, Sağlık Merkezi	7 Hastane, Sağlık Merkezi	7 Hastane, Sağlık Merkezi	7 Hastane, Sağlık Merkezi	7 Hastane, Sağlık Merkezi
8 Şehir dışı	8 Şehir dışı	8 Şehir dışı	8 Şehir dışı	8 Şehir dışı	8 Şehir dışı	8 Şehir dışı	8 Şehir dışı
9 Arkadaş Ziyareti	9 Arkadaş Ziyareti	9 Arkadaş Ziyareti	9 Arkadaş Ziyareti	9 Arkadaş Ziyareti	9 Arkadaş Ziyareti	9 Arkadaş Ziyareti	9 Arkadaş Ziyareti
10 Diğer	10 Diğer	10 Diğer	10 Diğer	10 Diğer	10 Diğer	10 Diğer	10 Diğer

DENEĞİN YAPTIĞI BÜTÜN SEYAHATLERİ YAZDIĞINIZDAN EMİN OLUNCAYA KADAR İRDELEYİNİZ. GEREKİRSE HİÇ BELİRTİLMEMİŞ SEYAHAT

→

Yukarıda belirttiğimiz seyahatlerinizin detaylı bilgilerini alacağım. Lütfen hatırlayabileceğiniz şekilde cevaplayınız.

C.1. SEYAHAT NO	BAŞLANGIÇ				VARIŞ
	C.2. SEYAHAT KODLARI C.0.'DA YER ALAN SIRAYI TAKİP EDEREK YAZILACAKTIR 1.Ev 2.İş 3.Okul 4.İş takibi 5.Alışveriş 6.Sosyal, spor ve eğlence 7.Sağlık 8.Şehir dışı 9.Arkadaş Ziyareti 10. Diğer (Belirtiniz) (Diğer yolculuklar için; peki sonra)	C.3. Seyahatinizin başlangıç noktasının açık adresini veya ilçe ve mahallesini söyler misiniz? (MAHALLENİN BİLİNMEDİĞİ DURUMLARDA BİLİNEN EN YAKIN BİR NOKTA (KAVŞAK, OKUL, CAMİ, OTOBÜS DURAĞI VEYA BİLİNEN BİR BİNA) İÇİN REFERANS BİLGİLER EDİNİLECEKTİR)	C.3.1. Seyahatin başladığı mahalle kodu	C.4. Seyahatinizin başlangıç saatini tam olarak söyleyebilir misiniz?	C.5. Hangi amaçla / nereye gittiniz? 1.Ev 2.İş 3.Okul 4.İş takibi 5.Alışveriş 6.Sosyal, spor ve eğlence 7.Sağlık 8.Şehir dışı 9.Arkadaş Ziyareti 10.Diğer (Belirtiniz)
1		Adres: Mahalle: İlçe:		Saat Dakika	
2		Adres: Mahalle: İlçe:		Saat Dakika	
3		Adres: Mahalle: İlçe:		Saat Dakika	

→

NOT

1. 6 YAŞ VE ÜZERİNDEKİ HER KİŞİ İÇİN AYRI BİR YOLCULUK FORMU DOLDURULACAKTIR.
2. 6 YAŞ VE ÜZERİNDEKİ KİŞİLERİN 24 SAATTE YAPTIKLARI TÜM YOLCULUKLAR KAYDEDİLECEKTİR.
3. 6 YAŞ VE ÜZERİNDE OLUP YOLCULUK YAPMAYAN KİŞİLERİN FORMUNA "Hayır çıkmadım" VE "Çıkmama nedeniniz" İŞARETLENECEKTİR.

4

BÖLÜM II. 24 SAATLİK SEYAHAT BİLGİLERİ

Görüşmenin başlangıç saati ve dakikası ☐☐ ☐☐ Saat Dakika	Görüşmenin bitiş saati ve dakikası ☐☐ ☐☐ Saat Dakika
Adı Soyadı ______	Seyahat bilgisinin ait olduğu gün 1 Ptesi 2 Salı 3 Çarş 4 Perş 5 Cuma 6 Ctesi 7 Pazar

Hayır, çıkmadım. 0 → **Çıkmama Nedeniniz**

Evet, çıktım. 1

1 Hastaydım	3 Yapılacak herhangi bir işim yoktu	5 Diğer ______
2 Hava koşulları uygun değildi	4 Seyahat yapacak maddi imkanım yoktu	______

BÖLÜM III'e GEÇİNİZ

C.0.9. Ordan nereye gittiniz?	C.0.10. Ordan nereye gittiniz?	C.0.11. Ordan nereye gittiniz?	C.0.12. Ordan nereye gittiniz?	C.0.13. Ordan nereye gittiniz?	C.0.14. Ordan nereye gittiniz?	C.0.15. Ordan nereye gittiniz?	C.0.16. Ordan nereye gittiniz?
1 Ev	1 Ev	1 Ev	1 Ev	1 Ev	1 Ev	1 Ev	1 Ev
2 İş	2 İş	2 İş	2 İş	2 İş	2 İş	2 İş	2 İş
3 Okul	3 Okul	3 Okul	3 Okul	3 Okul	3 Okul	3 Okul	3 Okul
4 İş takibi	4 İş takibi	4 İş takibi	4 İş takibi	4 İş takibi	4 İş takibi	4 İş takibi	4 İş takibi
5 Alışveriş	5 Alışveriş	5 Alışveriş	5 Alışveriş	5 Alışveriş	5 Alışveriş	5 Alışveriş	5 Alışveriş
6 Sosyal, spor ve eğlence	6 Sosyal, spor ve eğlence	6 Sosyal, spor ve eğlence	6 Sosyal, spor ve eğlence	6 Sosyal, spor ve eğlence	6 Sosyal, spor ve eğlence	6 Sosyal, spor ve eğlence	6 Sosyal, spor ve eğlence
7 Hastane, Sağlık Merkezi	7 Hastane, Sağlık Merkezi	7 Hastane, Sağlık Merkezi	7 Hastane, Sağlık Merkezi	7 Hastane, Sağlık Merkezi	7 Hastane, Sağlık Merkezi	7 Hastane, Sağlık Merkezi	7 Hastane, Sağlık Merkezi
8 Şehir dışı	8 Şehir dışı	8 Şehir dışı	8 Şehir dışı	8 Şehir dışı	8 Şehir dışı	8 Şehir dışı	8 Şehir dışı
9 Arkadaş Ziyareti	9 Arkadaş Ziyareti	9 Arkadaş Ziyareti	9 Arkadaş Ziyareti	9 Arkadaş Ziyareti	9 Arkadaş Ziyareti	9 Arkadaş Ziyareti	9 Arkadaş Ziyareti
10 Diğer	10 Diğer	10 Diğer	10 Diğer	10 Diğer	10 Diğer	10 Diğer	10 Diğer

TÜRLERİNİ HATIRLATARAK BU TÜR SEYAHATLERDEN YAPIP YAPMADIĞINI SORUNUZ

TOPLAM YOLCULUK ADEDİNİ DENEĞE ONAYLATARAK YAZINIZ. ☐☐

C.6. Gittiğiniz yerin açık adresini veya ilçe ve mahallesini söyler misiniz? (AÇIK ADRESİN BİLİNMEDİĞİ DURUMLARDA BİLİNEN EN YAKIN BIR NOKTA (KAVŞAK, OKUL, CAMİ, OTOBÜS DURAĞI VEYA BİLİNEN BİR BİNA) İÇİN REFERANS BİLGİLER EDİNİLECEKTİR)	C.6.1. Seyahatin bittiği mahalle kodu	C.7. Gittiğiniz yere tam olarak saat kaçta vardınız?	KULLANILAN ULAŞIM TÜRLERİ KOMBİNASYONU: C.8. Ulaşım türü sırası	C.9. Oraya giderken sırası ile hangi ulaşım araçlarını / türlerini kullandınız?	C.10. Kendi şahsi imkanlarınızı ve mevcut ulaşım alternatiflerini düşündüğünüzde bu seyahatinizi başka hangi ulaşım türleriyle yapabilirdiniz? (ULAŞIM ARACI KODU GİRİNİZ, BAŞKA ALTERNATİF YOK İSE "YOK" YAZINIZ, EN ÇOK ÜÇ TANE BELİRTİNİZ) 1. ALTERNATİF	2. ALTERNATİF	3. ALTERNATİF
Adres:	☐☐☐	☐☐ Saat ☐☐ Dakika	1				
			2				
Mahalle:			3				
İlçe:			4				
			5				
Adres:	☐☐☐	☐☐ Saat ☐☐ Dakika	1				
			2				
Mahalle:			3				
İlçe:			4				
			5				
Adres:	☐☐☐	☐☐ Saat ☐☐ Dakika	1				
			2				
Mahalle:			3				
İlçe:			4				
			5				

Ulaşım türü kodları

1. Yaya	6. Dolmuş	11. Bisiklet	16. Vapur
2. Özel oto yalnız sürüş	7. Minibüs	12. Metro (Taksim-4.Levent)	17. Deniz otobüsü
3. Özel oto paylaşılan sürüş	8. Belediye otobüsü	13. Hafif Metro (Aksaray-Havaalanı)	18. Deniz motoru
4. Taksi	9. Özel halk otobüsü	14. Tramvay	19. Banliyo
5. Servis aracı	10. Motosiklet	15. Tünel	20. Diğer (Belirtiniz)
			21. Füniküler

5

BÖLÜM II. 24 SAATLİK SEYAHAT BİLGİLERİ

C.1. SEYAHAT NO	BAŞLANGIÇ				VARIŞ
	C.2. SEYAHAT KODLARI C.0.'DA YER ALAN SIRAYI TAKİP EDEREK YAZILACAKTIR 1.Ev 2.İş 3.Okul 4.İş takibi 5.Alışveriş 6.Sosyal, spor ve eğlence 7.Sağlık 8.Şehir dışı 9.Arkadaş Ziyareti 10. Diğer (Belirtiniz) (Diğer yolculuklar için; peki sonra)	C.3. Seyahatinizin başlangıç noktasının açık adresini veya ilçe ve mahallesini söyler misiniz? (MAHALLENİN BİLİNMEDİĞİ DURUMLARDA BİLİNEN EN YAKIN BİR NOKTA (KAVŞAK, OKUL, CAMİ, OTOBÜS DURAĞI VEYA BİLİNEN BİR BİNA) İÇİN REFERANS BİLGİLER EDİNİLECEKTİR)	C.3.1. Seyahatin başladığı mahalle kodu	C.4. Seyahatinizin başlangıç saatini tam olarak söyleyebilir misiniz?	C.5. Hangi amaçla / nereye gittiniz? 1.Ev 2.İş 3.Okul 4.İş takibi 5.Alışveriş 6.Sosyal, spor ve eğlence 7.Sağlık 8.Şehir dışı 9.Arkadaş Ziyareti 10.Diğer (Belirtiniz)
4		Adres: Mahalle: İlçe:		Saat Dakika	
5		Adres: Mahalle: İlçe:		Saat Dakika	
6		Adres: Mahalle: İlçe:		Saat Dakika	
7		Adres: Mahalle: İlçe:		Saat Dakika	
8		Adres: Mahalle: İlçe:		Saat Dakika	
9		Adres: Mahalle: İlçe:		Saat Dakika	
10		Adres: Mahalle: İlçe:		Saat Dakika	
11		Adres: Mahalle: İlçe:		Saat Dakika	

Seyahatlerin hepsi kaydedilmemiş ise, ek yolculuk formu kullanınız.

NOT

1. 6 YAŞ VE ÜZERİNDEKİ HER KİŞİ İÇİN AYRI BİR YOLCULUK FORMU DOLDURULACAKTIR.
2. 6 YAŞ VE ÜZERİNDEKİ KİŞİLERİN 24 SAATTE YAPTIKLARI TÜM YOLCULUKLAR KAYDEDİLECEKTİR.
3. 6 YAŞ VE ÜZERİNDE OLUP YOLCULUK YAPMAYAN KİŞİLERİN FORMUNA "Hayır çıkmadım" VE "Çıkmama nedeniniz" İŞARETLENECEKTİR.

BÖLÜM II. 24 SAATLİK SEYAHAT BİLGİLERİ

C.6. Gittiğiniz yerin açık adresini veya ilçe ve mahallesini söyler misiniz? (AÇIK ADRESİN BİLİNMEDİĞİ DURUMLARDA BİLİNEN EN YAKIN BIR NOKTA (KAVŞAK, OKUL, CAMİ, OTOBÜS DURAĞI VEYA BİLİNEN BİR BİNA) İÇİN REFERANS BİLGİLER EDİNİLECEKTİR)	C.6.1. Seyahatin bittiği mahalle kodu	C.7. Gittiğiniz yere tam olarak saat kaçta vardınız?	KULLANILAN ULAŞIM TÜRLERİ KOMBİNASYONU: C.8. Ulaşım türü sırası	C.9. Oraya giderken sırası ile hangi ulaşım araçlarını / türlerini kullandınız?	C.10. Kendi şahsi imkanlarınızı ve mevcut ulaşım alternatiflerini düşündüğünüzde bu seyahatinizi başka hangi ulaşım türleriyle yapabilirdiniz? (ULAŞIM ARACI KODU GİRİNİZ. BAŞKA ALTERNATİF YOK İSE "YOK" YAZINIZ, EN ÇOK ÜÇ TANE BELİRTİNİZ) 1. ALTERNATİF	2. ALTERNATİF	3. ALTERNATİF
Adres:			1				
		Saat Dakika	2				
Mahalle:			3				
İlçe:			4				
			5				
Adres:			1				
		Saat Dakika	2				
Mahalle:			3				
İlçe:			4				
			5				
Adres:			1				
		Saat Dakika	2				
Mahalle:			3				
İlçe:			4				
			5				
Adres:			1				
		Saat Dakika	2				
Mahalle:			3				
İlçe:			4				
			5				
Adres:			1				
		Saat Dakika	2				
Mahalle:			3				
İlçe:			4				
			5				
Adres:			1				
		Saat Dakika	2				
Mahalle:			3				
İlçe:			4				
			5				
Adres:			1				
		Saat Dakika	2				
Mahalle:			3				
İlçe:			4				
			5				
Adres:			1				
		Saat Dakika	2				
Mahalle:			3				
İlçe:			4				
			5				

1. Yaya	6. Dolmuş	11. Bisiklet	16. Vapur
2. Özel oto yalnız sürüş	7. Minibüs	12. Metro (Taksim-4.Levent)	17. Deniz otobüsü
3. Özel oto paylaşılan sürüş	8. Belediye otobüsü	13. Hafif Metro (Aksaray-Havaalanı)	18. Deniz motoru
4. Taksi	9. Özel halk otobüsü	14. Tramvay	19. Banliyo
5. Servis aracı	10. Motosiklet	15. Tünel	20. Diğer (Belirtiniz)
			21. Funiküler

BÖLÜM III. HANEDEKİ ARAÇ BİLGİLERİ

Şimdi hanedeki araç bilgileriyle ilgili bazı sorular soracağım.

	D.1. Sayacağım ulaşım araçlarınız var mı?		D.2. Aracı akşamları nereye park ediyorsunuz? (BİRDEN ÇOK CEVAP ALABİLİRSİNİZ!)				D.3. ARACA AİT BİLGİLER		
	Yok	Var	1. Cadde / sokak kenarı	2. Kendi garajı (Site / apartman garajı vb.)	3. Ücretli açık otopark	4. Ücretli kapalı / katlı otopark	1. Araç Yapım Yılı	2. Araç Motor Hacmi	3.Motor yakıt tipi
1. Hanehalkına ait özel otomobiliniz var mı?	0 (4'E GEÇİNİZ)	1	1	2	3	4			1. Benzinli 2. Dizel 3. LPG
2. Hanehalkına ait ikinci bir özel otomobiliniz var mı?	0 (4'E GEÇİNİZ)	1	1	2	3	4			1. Benzinli 2. Dizel 3. LPG
3. Hanehalkına ait üçüncü bir özel otomobiliniz var mı?	0 (4'E GEÇİNİZ)	1	1	2	3	4			1. Benzinli 2. Dizel 3. LPG
4. Hanehalkı tarafından kullanılan -ve akşamları kullanan kişide kalan- şirket aracı gibi bir araç var mı?	0 (6'YA GEÇİNİZ)	1	1	2	3	4			1. Benzinli 2. Dizel 3. LPG
5. Hanehalkı tarafından kullanılan -ve akşamları kullanan kişide kalan- şirket aracı gibi ikinci bir araç var mı?	0 (6'YA GEÇİNİZ)	1	1	2	3	4			1. Benzinli 2. Dizel 3. LPG
6. Hanehalkına ait başka motorlu araç var mı?	0 (E.1.'E GEÇİNİZ)	1	1	2	3	4			
6.1. -Varsa- diğer motorlu aracın türü / türleri nedir? (ŞIKLARI OKUYUNUZ VE VARSA BİRDEN ÇOK ŞIK İŞARETLEYİNİZ)			1. Kamyon, 2. Kamyonet, 3. Minibüs, 4. Ticari Taksi, 5. Otobüs, 6. Traktör, 7. Diğer (Belirtiniz)						

BÖLÜM IV. HANENİN MÜLKİYET VE GELİR DURUMUYLA İLGİLİ BİLGİLERİ

E.1. Bu konutta mülkiyet durumunuz nedir?

1	Ev sahibi
2	Kiracı
3	Lojman
4	Ev sahibi değil ama kira ödemiyor
5	Diğer

E.2. Evinizin kirası aylık ne kadardır? (YTL) ____

(DİKKAT! E.2'Yİ KİRACI OLANLARA SORUNUZ)

E.3. Evinizin fiziki durumunu, çevresel şartları, civardaki ortalama kira bedellerini düşündüğünüzde evinizi kiraya verseniz aylık ne kadar gelir getirir? (YTL) ____

(DİKKAT! E.3.'Ü EV SAHİBİ OLANLARA SORUNUZ)

E.4. Hanehalkı *oturduğunuz ev dışında* sayacağım mülklerden hangilerine sahiptir?

1. Apartman dairesi	1
2. Müstakil konut (yazlık vb.)	2
3. Dükkan / Büro	3
4. Arsa	4
5. Tarla / bağ / bahçe	5
6. Diğer (Belirtiniz) ____	X

E.5. Hanenin tüm ev halkına ait ayni ve nakdi kazancı, gayri menkul, faiz gelirleri vb. dahil olmak üzere aylık ortalama geliri sayacağım aralıklardan hangisine girmektedir?

1	250 YTL altı	6	1.251 – 1.500 YTL	11	2.501 – 3.500 YTL
2	251 – 500 YTL	7	1.501 – 1.750 YTL	12	3.501 – 4.999 YTL
3	501 – 750 YTL	8	1.751 – 2.000 YTL	13	5.000 – 7.499 YTL
4	751 – 1000 YTL	9	2.001 – 2.250 YTL	14	7.500 – 9.999 YTL
5	1.001 – 1.250 YTL	10	2.251 – 2.500 YTL	15	10.000 YTL ve üzeri

E.6. Peki hanehalkının aylık gelirini tam, rakamsal olarak söyler misiniz (YTL)? ____

(HANE FERTLERİNİN HEPSİNİN, AYNİ VE NAKDİ BÜTÜN GELİRLERİNİ GÖZ ÖNÜNDE BULUNDURUNUZ.)

BÖLÜM V. TOPLU TAŞIMA SORUNLARI

F. 1. Semtinizde aşağıdaki toplu taşıma araçlarından hangileri hizmet vermektedir?

	EVET	HAYIR		EVET	HAYIR
İETT Otobüs	1	2	Vapur	1	2
Halk / Özel Otobüsü	1	2	Deniz Otobüsü	1	2
Metro	1	2	Minibüs	1	2
Tramvay	1	2	Dolmuş	1	2

F. 2. Semtinizdeki toplu taşıma hatlarında aşağıda sayacağım sorunlardan hangileri yaşanmaktadır?

	EVET	HAYIR	FİKRİM YOK
1. Otobüslerin sefer saatlerine uymaması	1	2	3
2. Sefer sayılarının yetersiz olması	1	2	3
3. Hat sayısının yetersiz olması	1	2	3
4. Yolculuk ücretlerinin pahalı olması	1	2	3
5. Akbil dolum gişelerinin yetersiz olması	1	2	3
6. Araçların kalabalık olması	1	2	3
7. Araçların kirli olması	1	2	3

F. 3. Peki bunların dışında sizce semtinizde ulaşımda yaşanan diğer sorunlar nelerdir? (BİRDEN ÇOK CEVAP ALINABİLİR!)

____

____

____

____

Bizi kontrol etmek için sizi arayabilirler. Sakıncası yoksa telefon numaranızı alabilir miyiz?

TEL NO: ☐☐☐☐ ☐☐☐ ☐☐ ☐☐ TEL YOK: ☐ NO. VERMEK İSTEMİYOR ☐

CEVAPLAYICIYA TEŞEKKÜR EDİNİZ VE GÖRÜŞMEYİ BİTİRİNİZ.

Görüşmenin BİTİŞ saati ve dakikası ☐☐ Saat ☐☐ Dakika

Tarih ☐☐ Gün ☐☐ Ay 2 0 0 6 Yıl

Anketörün Adı Soyadı ______________________

Size göre yanıtların güvenilirlik derecesi nedir?

1. Çok zayıf 1
2. Zayıf 2
3. Orta 3
4. İyi 4
5. Çok iyi 5

ANKETÖR GÖZLEMLERİ

HANE HAKKINDA GÖZLEMLER

BELİRLİ SORULAR HAKKINDA GÖZLEMLER

DİĞER GÖZLEMLER

EKİP BAŞININ (ALAN EDİTÖRÜ) GÖZLEMLERİ

Ekip Başının Adı Soyadı:

DENETÇİNİN GÖZLEMLERİ

Denetçinin Adı Soyadı:

9

VITA

Birthplace / Date: İzmir, July 15, 1978

e-mail: uguryankaya@iyte.edu.tr; uyankaya@gmail.com

Education:

2004 – 2010 **Ph.D.** İzmir Institute of Technology, Faculty of Architecture, Department of City and Regional Planning.

Thesis: Modeling The Effects of Land Use Characteristics on Mode Choice For Home - Based Work Trips: The Case of Istanbul.

2000 – 2004 **M.Sc.** İzmir Institute of Technology, Faculty of Architecture, Department of City and Regional Planning.

Thesis: Modeling The Impacts of Izmir Subway on the Values of Residential Property Using Hedonic Price Model.

1996– 2000 **B.Sc.** Dokuz Eylul University, Faculty of Architecture, Department of City and Regional Planning.

Some Recent Publications:

(1) Yankaya, Uğur and Celik, H. Murat. 2006.*“Modeling the Impacts of Rail Transit Investment on the Values of Residential Property: A Hedonic Price Approach in The Case of Izmir Subway, TURKEY”*, Ecomod: International Conference on Regional and Urban Modeling, 1 - 3 June 2006, Brussels.

(2) Yankaya, Uğur., Celik, H. Murat., Ozdemir Serhan., and Sevil, Hakki Erhan. 2010. *“Chaotic Structure Test and Predictability Analysis on Traffic Time Series in The City of Istanbul”*, İstanbul Kültür University, 3rd International Interdisciplinary Chaos Symposium on Chaos and Complex Systems, 21 – 24 May 2010, Istanbul.

(3) Yankaya, Uğur. and Çelik, H. Murat. 2005.*“İzmir Metrosunun Konut Fiyatları Üzerindeki Etkilerinin Hedonik Fiyat Yöntemi ile Modellenmesi”*, Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi 20(2): 61-79.

(4) Çelik, H. Murat and Yankaya, Uğur. 2006. *“The Impact of Rail Transit Investment on the Residential Property Values in Developing Countries: The Case of Izmir Subway”*, Turkey, Journal of Property Management 24(4): 369-382.

(5) Yankaya, Uğur ve Celik, H. Murat: “*Arazi Kullanım Karakteristiklerinin Tür Seçimi Üzerindeki Etkilerini Modellemek: İstanbul Örneği”*, 1. Ulusal Planlamada Sayısal Modeller Sempozyumu, İstanbul Teknik Üniversitesi, 24 – 26 Kasım 2010, Istanbul.